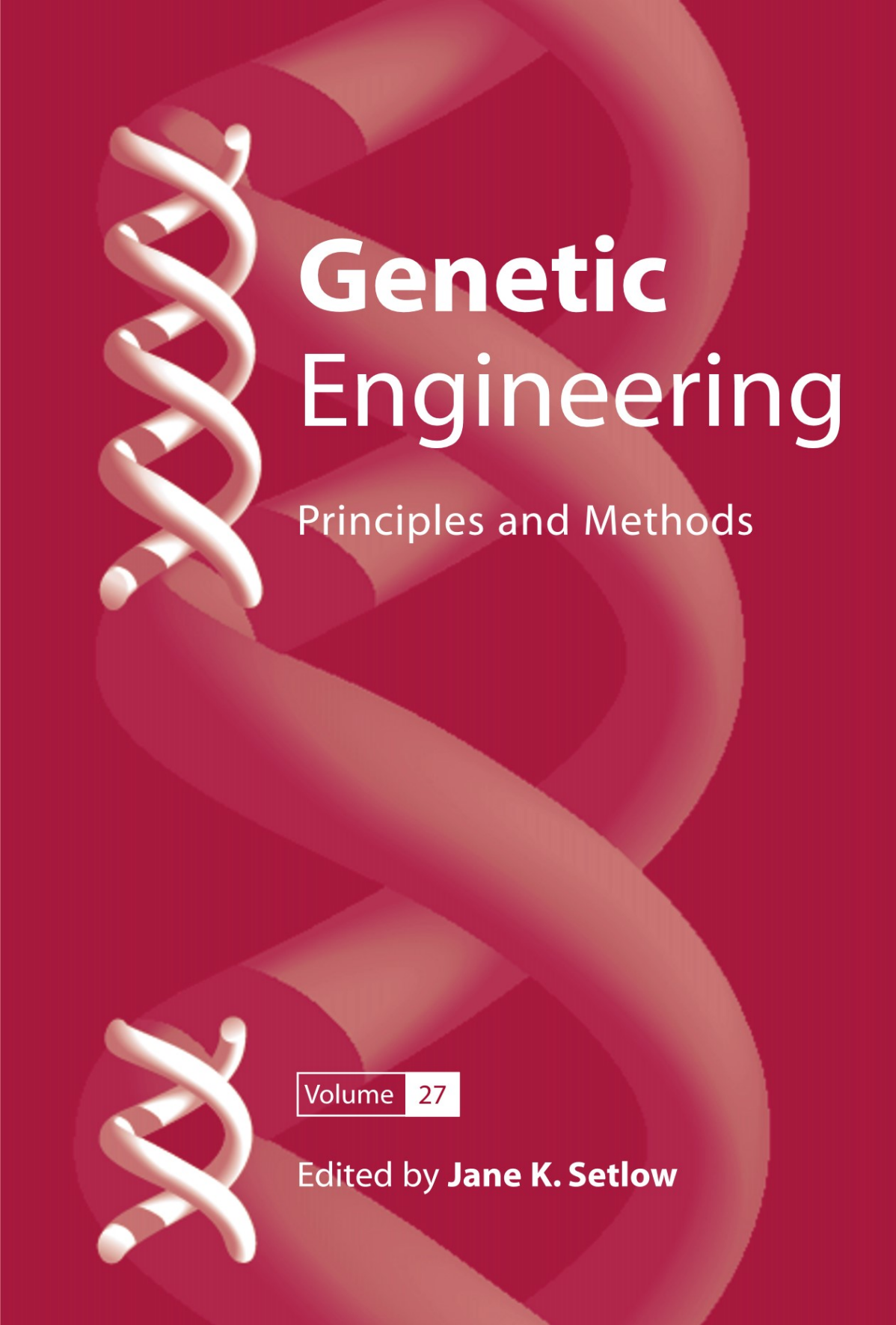

Genetic
Engineering
Principles and Methods
Volume 27
Edited by Jane K. Setlow

Genetic Engineering

Principles and Methods

Volume 27

GENETIC ENGINEERING
Principles and Methods

Genetic Engineering

Principles and Methods

Volume 27

Edited by

Jane K. Setlow
Brookhaven National Laboratory
Upton, New York

Springer

The Library of Congress cataloged the first volume of this title as follows:

Genetic engineering: principles and methods. V. 1–
New York, Plenum Press. (1979–
v. ill. 26 cm.
Editors: J. K. Setlow and A. Hollaender
Key title: Genetic engineering. ISSN 0196-3716
1. Genetic engineering—Collected works. I. Setlow, Jane K. (1979–) II. Hollaender, Alexander, (1979–1986).

QH442.G454 575.1 76-644807
MARC-S

ISBN-10: 0-387-25855-8 e-ISBN 0-387-25856-6
ISBN-13: 978-0387-25855-3

Printed on acid-free paper.

Printed in the United States of America. SPI/EB

9 8 7 6 5 4 3 2 1

springeronline.com

ACKNOWLEDGMENT

The Editor again praises Bonnie McGahern for the competent final processing, including fixing some of the Editor's errors

CONTENTS OF EARLIER VOLUMES

CONTENTS

IDENTIFICATION AND ANALYSIS OF MICRORNAS

Shveta Bagga and Amy E. Pasquinelli

Molecular Biology Section
Division of Biology
University of California, San Diego
La Jolla, CA 92093-0349

INTRODUCTION

The first microRNA (miRNA) gene was uncovered in 1993. After languishing in near obscurity for almost a decade, this gene is now recognized as the founding member of a new class of regulatory RNAs that control gene expression in all multicellular organisms. MicroRNA genes express ~22 nucleotide (nt) RNAs that regulate the expression of protein-coding genes containing sequences of antisense complementarity. The intense interest in understanding the role of miRNAs in regulating gene expression has fueled the development of new methods to study how these tiny RNA genes are expressed and function. In this chapter, we present a brief history outlining the discovery of miRNAs and the current model for their biogenesis and mode of action. We then describe experimental approaches used to analyze miRNA expression patterns and regulatory functions.

Genetic Engineering, Volume 27, Edited by J. K. Setlow

EXPRESSION AND FUNCTION OF miRNAS

Discovery of miRNAs Through Nematode Genetics

Forty years ago, Sydney Brenner proposed adoption of the microscopic nematode *Caenorhabditis elegans* for studying the genetic basis of animal development and behavior. Not only has the worm proven to be a model experimental system for identifying the genes responsible for controlling cell fate and function, but it also enabled the discovery of an entirely unexpected class of genes and a novel regulatory mechanism. The first microRNA gene was uncovered through classical genetic methods to identify a mutation responsible for abnormal development of certain worm cells. The Ambros laboratory found that the developmental defects resulted from mutation of the *lin-4* (lin = lineage) gene, which encoded a 21 nucleotide (nt) regulatory RNA (1). This type of gene was unprecedented, but opportune work from the Ruvkun laboratory on another developmental gene, *lin-14*, provided the necessary clues for predicting how a tiny RNA product might control gene expression (1, 2). The *lin-4* RNA recognizes sites of imperfect complementarity in the 3′ untranslated region (UTR) of the *lin-14* messenger RNA (mRNA) and halts protein expression (Figure 1). Insufficient *lin-4* RNA or deletion of the target sites in the *lin-14* 3′ UTR leads to failed down-regulation of LIN-14 protein expression at the appropriate time and, thus, abnormal development (1-4). Although the LIN-14 protein disappears in response to the *lin-4* RNA, the *lin-14* mRNA remains and continues to associate with polysomes, indicating that translational inhibition is the mechanism at work (2, 5).

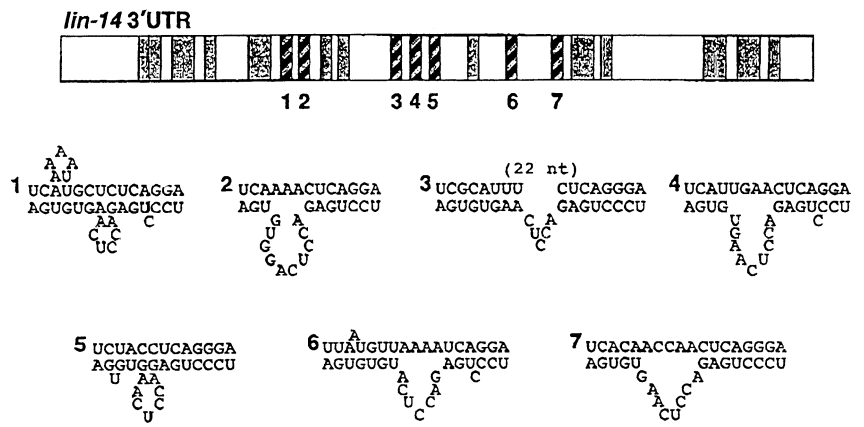


Figure 1. Conserved sites in the 3′UTR of the *C. elegans lin-14* mRNA are complementary to the *lin-4* miRNA (2). Shaded blocks in the *lin-14* 3′UTR indicate sequences of high homology (at least 10 nucleotides of exact conservation) between the related nematodes *C. elegans* and *C. briggsae*. The striped blocks 1-7, represent regions of partial complementarity to the *lin-4* miRNA. The duplexes are shown with the *lin-14* top strand reading 5′ to 3′ base-paired with the bottom strand *lin-4* miRNA.

Another target of *lin-4* regulation, the *lin-28* protein-coding gene, was also discovered by the Ambros laboratory, providing another example of a developmental gene under post-transcriptional control by the tiny RNA (6). Yet, the question of whether this novel mode of gene regulation was restricted to nematodes persisted until the turn of the century (7). The identification of the *let-7* (let = lethal) gene in *C. elegans* as another 22 nt RNA that regulates the expression of protein-coding genes containing 3′ UTR target sites raised the possibility that tiny RNA genes might abound in worms and beyond. Indeed, the remarkable conservation of the *let-7* RNA sequence enabled the Ruvkun laboratory to establish that this gene is expressed in diverse animals, including fruit flies, molluscs, sea urchins, zebrafish, and humans (8). Moreover, temporally regulated expression of the *let-7* RNA and potential target sites in *lin-41* homologues in all species assayed implied that this RNA gene may be essential for development of many animal species (8).

Around the time *let-7* RNA was discovered, another type of tiny RNA was gaining fame. In 1998, Fire and Mello reported that injection of double-stranded RNA (dsRNA) into *C. elegans* could elicit the degradation of homologous mRNA and, thus, potently inhibit gene expression in a process termed RNA interference (RNAi) (9). Shortly thereafter, several groups found that the dsRNA is cleaved to ~22 nt small interfering RNAs (siRNAs) that serve as the guides to target complementary mRNA sequences for destruction (10-14). It was clear that tiny RNAs could be powerful regulators of gene expression and soon hundreds of ~22 nt RNA genes were uncovered in animal and plant genomes (15-20). The small size of these regulatory RNAs inspired the name microRNA (21), and genes encoding these RNAs appear to be present in all multicellular organisms (22).

Transcription of miRNAs

Despite our relatively brief awareness of their existence, impressive progress has been made in understanding how miRNAs are expressed and function (22). The ~22 nt, mature forms of miRNAs arise from multiple processing steps of longer substrate RNAs. So far, little is known about the composition of the initial miRNA transcripts, called primary miRNAs (pri-miRNAs). Both Pol II and Pol III promoters have been used to drive ectopic expression of pri-miRNAs (23-25). Recently, the Kim laboratory presented direct evidence that Pol II transcribes several mammalian miRNAs (26). Additionally, a few complete pri-miRNAs that have been characterized show hallmarks of Pol II transcription—they apparently undergo 5′-end capping, 3′-end polyadenylation, and splicing (26-28). These first examples of pri-miRNAs are more than 1,000 nucleotides long—remarkably lengthy transcripts to serve as substrates for ~22 nt RNA products!

Many miRNAs are restricted to specific developmental periods or tissue types. At least in some cases, regulated expression of miRNAs is attributable to transcriptional control. Predicted promoter sequences of the *C. elegans let-7* gene can confine expression of green fluorescent protein (GFP) to late larval and adult stages—the same time in development when mature *let-7* is present (29). The transcriptional control sequences for *lys-6* miRNA, which regulates neuronal

asymmetry in *C. elegans*, restricts GFP expression to a subset of neurons (30). Although these examples of predicted miRNA promoters directing protein expression support the likelihood that they recruit Pol II, the specialized transcription factors that afford temporal or spatial control are yet to be identified for any miRNA gene.

Processing of miRNAs

Generation of the mature miRNA form requires multiple processing and cellular transportation events (Figure 2). In animals, the nuclear localized ribonuclease (RNase) Drosha clips the ~65 nt hairpin miRNA precursor (pre-miRNA) from the primary transcript (31). The pre-miRNA is shuttled by Exportin-5 to the cytoplasm for final processing by the RNase Dicer (32-38).

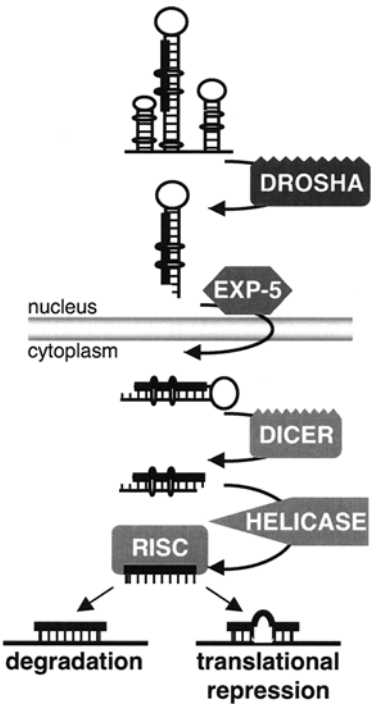


Figure 2. A model of miRNA biogenesis and function. The relatively long primary transcripts, called pri-miRNAs, are initially transcribed from miRNA genes (35). The pri-miRNAs are processed by the RNase Drosha to hairpin precursors (25, 31, 35). The precursors are recognized by Exportin-5 and delivered to the cytoplasm for maturation to ~22 nt RNAs by Dicer (25, 32-38). It has been proposed that a helicase activity separates the duplex (42, 43), and typically only one half is retained and incorporated into a multi-factor RNA-induced silencing complex (RISC) (12, 52-54). The degree of complementarity between a miRNA and its target site determines the regulatory mechanism: near perfect base-pairing directs RNA degradation and bulged duplexes mediate translational repression (22). This model is based primarily on work in animal systems; note that organismal differences exist for the protein factors and subcellular location of processing events (22).

This enzyme appears to measure ~22 nt from the 5′ and 3′ ends of the hairpin to position a staggered cut through both strands of the stem (25, 31, 39-41). Typically, only one half of the resulting duplex is retained. Thermodynamic arguments have been made to explain the choice for which strand persists. The 5′ end that is more easily peeled away from its antisense is favored for incorporation into a stable complex and, by default, the other half is unprotected and degraded (42, 43).

Pri-miRNAs contain sequences and structures important for processing and generation of the functional ~22 nt form. However, truncated pri-miRNA substrates, even the hairpin precursors, can suffice as substrates to produce mature miRNAs when overexpressed from heterologous constructs (23-25, 31, 35). In the endogenous situation, processing may be a critical control point in miRNA biogenesis. Deletion of *cis*-acting sequences in pri-miRNA transcripts or depletion of *trans*-acting processing factors can inhibit miRNA maturation (27, 31-34). In some cases, the miRNA substrates accumulate *in vivo*, indicating that transcription of a miRNA gene and production of the mature form are not necessarily coupled.

Function of miRNAs

Mature miRNAs inhibit expression of genes containing sequences of antisense complementarity. In animals, the primary mechanism of gene regulation concurs with the original model proposed for the *lin-4* miRNA and *lin-14* mRNA in *C. elegans* (1, 2). Imperfect base-pairing between the miRNA and sequences in the 3′ UTR of the target mRNA results in inhibited protein expression (2, 5, 44). It remains to be determined how partial base-pairing between miRNAs and target sequences results in blocked protein production.

In plants, many miRNAs exhibit perfect, or nearly complete, base-pair complementarity to their target mRNAs (45, 46). Not only does this feature of plant miRNAs make it simpler to predict specific targets, but it also results in target degradation (46-50). Animal miRNAs can also direct mRNA destabilization if they share near-perfect complementarity with target sequences (24, 51-54). In fact the vertebrate miRNA, miR-196, can form a complete duplex with sequences in *HOXB8* mRNA and direct degradation of this target (55). The general model holds that miRNAs can regulate gene expression by either translational inhibition or mRNA destabilization, depending on the nature of the duplex formed with the target sequences.

EXPERIMENTAL IDENTIFICATION OF miRNAS

The noncoding nature and the extraordinarily small size of miRNAs make their detection challenging. For a long time, the conventional cloning and identification techniques and the gene prediction databases were clearly biased for long protein-coding sequences. The discovery of tiny RNAs in *C. elegans* and elucidation of the RNAi mechanism led several groups to adopt novel or modified conventional methods to detect miRNAs.

Genetic Screening

The discovery of pioneer members of the miRNA family—*lin-4* and *let-7*—demonstrated the potential of classical genetic screens in detection of miRNAs (1, 2, 56). Although time-consuming and labor-intensive, identification of an miRNA through a genetic screen can readily give important clues about its function and gene targets. Also, rare and nonconserved miRNAs, which usually evade cloning and computational detection, can be identified by genetic methods. A loss-of-function screen led to the identification of *C. elegans* miRNA *lys-6,* which controls left/right neuronal asymmetry (30). Gain-of-function genetic screens based on mutations in negatively regulated targets or forcing altered expression of miRNAs have also contributed to the detection of new miRNAs (48, 51, 57, 58). These studies emphasized an important distinction of present-day genetic mapping—to look for noncoding, short stem-loop structures in addition to conventional open-reading frames (ORFs). Increasing efforts toward developing full genome databases will facilitate the identification of more miRNAs through genetic screenings. Taking into account the abundance of miRNAs, it would not be surprising if many of the previously uncharacterized loci in genetic screens could be ascribed to miRNAs.

Biochemical Cloning

Direct cloning of expressed miRNAs by the Ambros', Bartel's, and Tuschl's laboratories led to the identification of the first populations of miRNAs in worms, flies, and humans (15, 16, 20). Several unique as well as highly conserved miRNAs, like *let-7*, were detected in these initial cloning efforts. Northern blot analyses of cloned miRNAs revealed both tissue-specific and stage-specific miRNAs, emphasizing their role in developmental timing and tissue specifications. The phylogenetic distribution of miRNAs was further expanded by the cloning of plant miRNAs (17-20, 59). Cloning of tiny RNAs from specific ribonucleoprotein complexes also identified several novel miRNAs (60). To date, biochemical cloning has led to the identification of hundreds of distinct miRNAs (15-17, 19, 20, 60-64).

An important characteristic that emerged from biochemical cloning and complied with *lin-4* and *let-7* sequences is the existence of animal miRNAs as a part of ~70 nucleotide stem-loop precursors (1, 8, 56). Processing of mature miRNAs from hairpin precursors is now considered a signature of animal miRNA genes (21). Although plant miRNAs also derive from precursors, composition of these substrates is not as well defined (19). The miRNA sequence can reside on either arm of the stem-loop structure, and hence the location on the precursor is not a determinant of its excision by Dicer. Cloning of miRNAs that are clustered in the genome and identification of some in expressed sequence tag (EST) databases hinted that miRNA precursors might derive from longer primary transcripts (15, 20, 28, 35, 61, 65). After the discovery of Drosha, it was speculated that specific cleavage of primary transcripts determines the correct register of Dicer action and hence the mature ends of miRNAs are determined at the level of primary transcripts (31).

As the result of being RNase III Dicer products, miRNAs are cloned based on their three distinguishing features: a length of about 22 nt, a 5′-terminal monophosphate, and a 3′-terminal hydroxyl group (10, 33, 66). The general protocol for miRNA cloning involves size fractionation of an RNA population followed by ligation with adapter molecules (Figure 3) (10, 15, 16, 20). The chimeric RNA is then subjected to reverse-transcriptase polymerase chain reaction (RT-PCR), cloned, and sequenced. One of the advantages of biochemical cloning of miRNAs is that the expressed miRNA population from any tissue or at any stage of development can be readily detected. Cloning of mouse brain-tissue miRNAs revealed probable orthologs of *C. elegans lin-4* RNA, and the mouse sequences revealed probable *Drosophila* orthologs as well (61). Homologues of the *lin-4* gene had, thus far, not surfaced from informatic searches of other organisms. Although powerful in terms of revealing expressed miRNAs directly, detection by cloning has an inevitable drawback of selecting clones of breakdown products of abundant cellular RNAs. Hence, to qualify as an miRNA, a small cloned RNA should be able to form a stem-loop precursor structure with its flanking sequences and show conservation in related species (21). Endogenous siRNAs are usually distinguished from miRNAs by extended dsRNA structure of their precursors and by displaying less sequence conservation (21, 67).

Interestingly, cloning efforts in *C. elegans* and *Drosophila* led to the identification of new categories of noncoding RNAs designated as "tiny noncoding RNAs" (tncRNAs) and "repeat-associated small interfering RNAs" (rasiRNAs) (63, 64). The 24-26 nt rasiRNAs apparently derive from various repetitive sequence elements including retrotransposons, DNA transposons, satellite, and

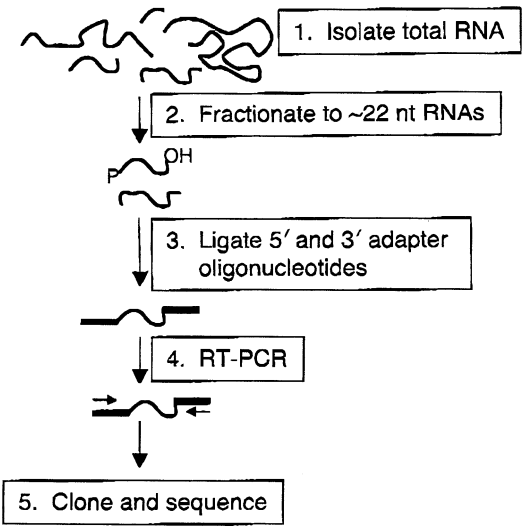


Figure 3. miRNA cloning strategy. Typically, total RNA is fractionated to ~22 nt size forms and miRNAs containing 5′ phosphate and 3′ hydroxyl groups are substrates for ligation to adaptor oligonucleotides (10, 15, 16, 20). The chimeric RNA is subjected to RT-PCR, cloning, and sequencing. Legitimate miRNAs match genomic sequences that support formation of a hairpin precursor (21).

microsatellite sequences, complex as well as vaguely characterized repetitive sequence motifs (64). The tncRNAs are similar in size to miRNAs but are not processed from stem-loop precursors and do not have orthologs in other species (63). Although some of tncRNAs exhibit temporal expression patterns, their exact role and significance await further experimentation.

Informatics

Although biochemical cloning led to the identification of several hundreds of new miRNAs, it is limited for identifying rare miRNAs or those that are triggered by specific environmental conditions. The availability of full genome databases of several organisms enabled the development of informatics approaches for identification of new miRNAs.

The fortuitous discovery of the first conserved miRNA, *let-7,* demonstrated the potential of simple homology searches using BLASTN (8). Homology searches with cloned miRNAs also revealed orthologs and paralogs in various organisms (15, 16, 20). A simple homology-based strategy originally involved the analysis of intergenic sequences among related organisms using the RNA folding program "mfold" (16, 68). The output was scanned by eye for miRNA characteristic stem-loop structures and the expression was confirmed by Northern blotting. The proximal location of several miRNA genes prompted the search for new miRNAs adjacent to the previously identified ones (20, 64, 69, 70). This approach is most suitable for identification of rapidly evolving miRNA genes, which are proximal to each other but are too divergent in sequence to be detected by general methods (22).

An important advance in detection of miRNA genes has been achieved by development of new computational approaches (63, 67, 71-76). All the programs primarily utilize sequence conservation, presence of stem-loop structures, and intergenic location of miRNAs as basic criteria. One of the more sensitive programs, "MiRscan," has been applied to vertebrate and nematode genomes to identify new miRNA genes (67, 74). The MiRscan program was developed by using the 50 cloned miRNAs from *C. elegans* as the training set (16, 20). Based on its similarity to the training set, a score is assigned to each putative genomic candidate that is identified by conserved stem-loop structures. The evaluation is based on seven features: 1) base-pairing of the miRNA portion of the fold-back, 2) base-pairing of the rest of the fold-back, 3) stringent sequence conservation in the 5′ half of the miRNA, 4) slightly less stringent sequence conservation in the 3′ half of the miRNA, 5) sequence biases in the first five bases of the miRNA, 6) a tendency toward having symmetric internal loops and bulges in the miRNA region, and 7) the presence of 2-9 consensus base-pairs between the miRNA and the terminal loop region with a preference for 4-6 base-pairs. The accuracy of MiRscan predictions has been further improved by the inclusion of conserved elements upstream of miRNA precursors (69). The successful application of this program to vertebrates, although developed using nematode miRNAs, demonstrated its universal application. It also emphasized that, despite sequence variations of miRNAs among diverse animals, their generic features are broadly conserved.

Using a reference set of *Drosophila* pre-miRNA sequences, another program called "miRseeker" identified novel miRNA genes (73). The miRseeker algorithm detects insect miRNA genes using a three-step filter strategy. The first step involves extraction of candidate genes using conserved and nongenic regions of *D. melanogaster* and *D. pseudoobscura* genomes. The next step identifies and ranks the stem-loop structured regions based on the helical length and free energy values. Finally, high-scoring regions averaged for two genomes are evaluated for divergence using the determinants of a reference set. In principle, miRseeker should be applicable to analysis of other sets of sequenced genomes of related organisms.

Recent informatics approaches specifically designed to detect plant miRNAs identified several new candidates (71, 76). These strategies are similar to MiRscan and miRseeker in terms of using homologous fold-back sequences conserved between *Arabidopsis* and *Oryza sativa*. However, the parameters constraining the selection of fold-back structures were specifically designed for plant miRNAs. The MIRcheck algorithm utilizes the sequences and structures of putative miRNA hairpins and 20mers within them (71). MIRcheck selects the candidates by restricting the number of unpaired, bulged, or asymmetrically unpaired, consecutive unpaired nucleotides, and the length of the hairpin. Unlike other programs, MIRcheck does not restrict based on pattern or extent of base pairing outside the 20mer sequence, a feature typical of plant miRNAs. Several of the plant miRNAs identified by this approach were confirmed by expression and target mRNA degradation (71).

ANALYSIS OF miRNAS

Many miRNAs exhibit diverse temporal and spatial expression patterns. Additionally, the relative levels of a particular miRNA can vary several orders of magnitude among different cell types. Adaptations of traditional molecular techniques as well as novel methods have been developed to analyze when, where, and how much of a specific miRNA exists and what is its biological function.

Expression Patterns

Northern blot and RNase protection assays yielded the first molecular evidence for the existence of a ~22 nt RNA product. A specific tiny RNA product was present in wild-type but not *lin-4* mutant worms, and this RNA reappeared upon rescue of the mutant with a transgene containing just 693 nt of *lin-4* genomic sequence (1). Typical analyses for miRNA expression by Northern blots utilize high percentage (10-15%) polyacrylamide gel electrophoresis (PAGE), which enables detection of the mature and precursor forms of the miRNA (Figure 4) (1, 77). The relative level of a mature miRNA can be readily assessed by sampling total RNA from particular tissues, developmental time points, or experimental conditions. However, Northern blotting to detect specific miRNAs can be labor intensive and insensitive to low-level miRNAs.

Computational prediction of miRNA genes avoids the cloning bias of detecting the more abundant species. Confirmation of a predicted miRNA can be

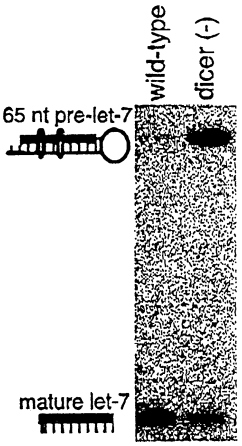


Figure 4. Northern analysis of miRNA expression. Total RNA from wild-type worms or worms depleted of dicer was isolated, separated by 11% polyacrylamide gel electrophoresis and subjected to Northern hybridization analysis to detect *let-7* RNA. The ~22 nt nucleotide mature form is predominant in wild-type worms, whereas the 65 nt precursor accumulates in Dicer(–) worms (32).

experimentally challenging, though, if the gene is weakly expressed or only activated under particular conditions. PCR-based approaches were developed to help validate the expression of miRNAs identified by informatics. Strong evidence for the existence of several elusive miRNAs was provided by a PCR protocol, which involves amplifying miRNA sequences from bulk miRNA cDNA libraries by way of the common adaptor sequences (67, 75). Real-time PCR assays have been employed for relatively high throughput analysis of miRNA precursor expression (78). In at least some cases, the level of precursor accurately reflected that of mature, as indicated by Northern analyses. More recently, an exceptionally sensitive and quantitative method was reported for detecting precursor or mature miRNAs (79). The Invader miRNA assay can detect as little as 20,000 molecules of a specific miRNA and has been used to show that the amounts of human let-7a miRNA vary over several orders of magnitude among different tissues (79).

Microarray technology offers an efficient and sensitive method to assess global changes in miRNA expression patterns. Microchips containing oligonucleotides corresponding to miRNA sequences have been used to screen various cell types to uncover the miRNA profile (80, 81). Additionally, this type of miRNA profiling was used to identify distinctions between normal human B cells and those derived from chronic lymphocytic leukemia cells (82). Since their discovery, miRNA genes have been considered possible disease candidates (83). High throughput profiling of miRNA expression patterns offers a powerful tool for correlating specific miRNAs with altered cell biological states.

Detection of miRNAs *in vivo* is particularly challenging considering their small size and potentially low abundance. Nonetheless, *in situ* hybridization results have indicated localized expression for a few miRNA transcripts. In

plants, there is an inverse correlation between expression of specific miRNAs and proposed targets in specific tissues, supporting the model that these RNAs negatively regulate protein expression to control development (84-86). Localized expression of miR-10 in *Drosophila* embryos indicates a role for this miRNA in regulating genes in the thoracic and abdominal primordia, although specific targets of miR-10 are yet to be identified (87).

An indirect method for analyzing temporal and spatial expression of miRNA genes is to fuse predicted miRNA promoter sequences to a reporter gene, such as GFP. This technique revealed tissue and developmental regulation of specific miRNA promoters in *C. elegans* that agreed with predictions about the function of the miRNAs (29, 30, 88). The *lys-6* miRNA was discovered as a gene that controls neuronal asymmetry in *C. elegans* by repressing expression of a transcription factor in a left taste neuron (30). Consistent with the proposed function of *lys-6*, a GFP reporter fused to the promoter for this miRNA gene is expressed in the left, but not right, neuron (30). These types of reporter experiments are very useful for predicting when and where an miRNA promoter functions as well as for studying its transcriptional control (29, 30, 88). However, the promoters and functions of most miRNAs are yet to be identified and, thus, caution is warranted for interpreting expression patterns based on fusions to miRNAs for which little is known about the natural biological role.

The *in situ* and reporter experiments described above can be used to indicate when and where an miRNA gene is active, but they do not demonstrate the production of functional miRNAs. Regulated processing and stabilization of some miRNAs may also influence their ability to control gene expression. An ingenious method to detect functional miRNAs *in vivo* was developed by the Cohen laboratory to show spatial and temporal expression of the *Drosophila bantam* miRNA (51). The "sensor" strategy is based on the demonstration that miRNAs will direct degradation of target mRNAs containing sites of perfect antisense complementarity (Figure 5) (24, 48, 52–54). A GFP-reporter gene containing *bantam* miRNA complementary sites was down-regulated in response to *bantam* expression. Thus, the presence of a functional miRNA can be assayed *in vivo* without the knowledge of its natural targets. Identification of specific miRNA expression patterns will greatly facilitate determination of biological functions.

Functional Roles

The first miRNAs, *lin-4* and *let-7*, were initially discovered as genes essential for regulating developmental timing in *C. elegans* (1, 56). Since the vast majority of RNAs to join the miRNA family were isolated by biochemical or computational means, biological functions are yet to be assigned. Considering their abundance, it is not surprising that miRNA genes are now being uncovered in mutant screens. Perhaps the lack of traditional gene structure allowed miRNAs to escape previous detection, but now mutations in miRNA genes account for broad-ranging phenotypes, including disrupted neuronal asymmetry, misregulated cell death, abnormal fat metabolism, and cellular patterning defects (18, 28, 30, 51, 58, 84, 88, 89). Isolation of genetic mutations in specific miRNA

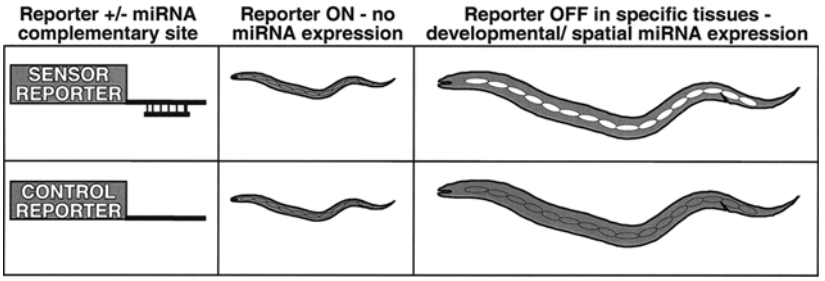


Figure 5. The "sensor" approach to analyze miRNA expression *in vivo* (51). In this example, constructs expressing a reporter protein, such as GFP, fused to 3′UTR sequences +/– miRNA complementary sites are introduced into worms. A ubiquitous promoter drives reporter expression. If the miRNA is absent, such as in early larval development (middle panels), GFP will be detected. However, if the miRNA turns on later in development or in particular cell types, the reporter mRNA will be specifically degraded and GFP will disappear (last panel, top row). This example shows a predicted pattern for *let-7* expression in *C. elegans*: early in development *let-7* is absent and thus GFP is expressed ubiquitously (gray shading of entire worm), including in the 10 hypodermal seam cells, but later in development *let-7* miRNA is produced and shuts off reporter expression, perhaps specifically in the 16 seam cells of adult worms (absence of gray shading) (29). Importantly, expression of a control reporter lacking the miRNA complementary sites is unaffected by miRNA expression (bottom panels).

genes not only aids in determining biological function but also is valuable for identifying direct targets of regulation. A genetic suppressor screen of the *let-7* mutant revealed *lin-41* as a target of negative regulation, which then led to the recognition of *let-7* complementary sites in the 3′UTR of *lin-41* mRNA (56, 90).

In many systems, targeted disruption or isolation of mutations in specific miRNAs is prohibitively laborious. Furthermore, homology among several groups of miRNAs suggests that redundancy may obscure phenotypes resulting from mutation of just one member. Overexpression or ectopic expression is an efficient alternative to study the function of particular miRNAs. The validity of this approach was established by introducing high copies of the *lin-4* gene to worms and observing developmental defects opposite of the *lin-4* loss-of-function phenotypes (4). More recently, ectopic expression of miR-181 in mouse hematopoietic stem cells biased their differentiation into B-lineage cells (23). Thus, direct targets of miR-181 may be predicted by focusing on distinct changes in gene expression in the B-lineage pathway.

The biological function of specific miRNAs can also be revealed by inhibition with antisense oligonucleotides. Injection of antisense DNA oligonucleotides corresponding to specific miRNAs into *Drosophila* embryos resulted in developmental defects (91). More recently, 2′-O-methyl oligonucleotides were shown to block potently the function of targeted miRNAs in *Drosophila*, human cell, and *C. elegans* systems (92, 93). The 2′-O-methyl modification protects the oligonucleotide against cellular RNases (94). Base-pairing of the oligonucleotide to an miRNA titrates the miRNA from its endogenous targets, thus revealing the loss-of-function phenotype. Although delivery of the antisense oligonucleotide can be technically prohibitive (92), this method of miRNA inhibition

offers an efficient means to uncover the biological roles of miRNAs for which only the mature sequence is known.

PAIRING OF miRNAS WITH TARGETS

The combination of cloning and computational approaches has likely enabled identification of the majority of miRNAs (22). However, as of yet, only a few miRNAs have been paired with their *bona fide* targets. Identification of direct miRNA targets is essential for understanding their diverse functions.

In addition to the discovery of pioneer miRNAs, the credit for the discovery of the first miRNA targets also goes to classical genetics. Long before the broad significance of tiny regulatory RNAs was appreciated, the functional pairing of *lin-4* RNA with its target *lin-14* mRNA was proposed (1, 2). The 3′UTR of *lin-14* mRNA had partial complementarity to *lin-4* RNA and was sufficient for temporal regulation of a reporter gene. The *let-7* target, *lin-41*, also supports the model, both in terms of partial complementarity and reporter gene regulation with the 3′UTR (56, 90, 95). The opposite phenotypes of *lin-4* and *lin-14* mutants helped pinpoint *lin-14* as a direct target of *lin-4* mediated negative regulation (96, 97). A handful of other *bona fide* miRNA target genes were identified through genetic screens (28, 30, 51, 57, 88, 89, 98, 99). However, for the majority of other miRNAs either mutants are not known or their mutant phenotypes are not apparent. Also, the small size and imperfect nature of base-pairing, particularly in case of animal miRNAs, hampers straightforward prediction of miRNA targets.

Target Identification and Validation for Plant miRNAs

Exact complementarity between miR171 and an mRNA target in *Arabidopsis* indicated that target prediction might be less complicated for plant compared with animal miRNAs (17, 19, 48). Indeed, "near perfect complementarity" appears to be a general rule for plant miRNA targets (46, 71, 76). Initially, targets were identified by searching annotated *Arabidopsis* mRNAs for 0-4 mismatches to specific miRNAs (19, 46). Conservation of the predicted mRNA target sequences in rice and low hits with a random cohort of tiny RNA sequences strengthened the validity of these proposed plant targets (46, 71). Plant miRNA targets show a clear bias toward transcription factors involved in cell differentiation and developmental patterning (46, 71). In comparison to sequences regulated by animal miRNAs, most plant miRNA-target interactions exhibit two general distinctions: 1) plant miRNA target sites are primarily found within open reading frames, and 2) multiple target sites within the same target mRNA are not detected in plants. These features may have significant functional implications for plant miRNAs—they favor an RNAi-like mechanism, as opposed to translational control, to inhibit gene expression (45).

A sensitive computational approach identified several novel plant miRNA targets belonging to families of transcription factors as well as other genes like ATP sulfurylase, laccase, and superoxide dismutase (71). This approach allowed for gaps and mismatches between mRNA:miRNA duplexes but

constrained the candidate targets to conservation between *Arabidopsis* and *Oryza*. Validation of predicted target sequences is facilitated by the fact that many plant miRNAs direct cleavage of their complementary mRNA targets. The 3′ cleavage product of the target, which maps to the 10th nucleotide of the miRNA and has a characteristic phosphate at its 5′ end, can be cloned and sequenced (47, 48, 50, 71, 76). Although absence of a 3′ cleavage product may suggest a false or alternatively regulated target, its presence is a convincing confirmation of regulation by a specific miRNA.

Informatics Approaches for Target Identification in Animals

New computational methods have matched animal miRNAs with numerous target genes, although many of the pairings still await experimental confirmation (100-104). The small number of validated miRNA targets in animals makes the development of reliable algorithms particularly challenging. As a starting point, most computational methods rely on conserved complementary sites within 3′ UTRs of potential target genes.

Identification of the *hid* gene as a *bantam* miRNA target exhibited the potential of computational approaches for identifying targets in *Drosophila* (51, 101). This approach was based on the presence of miRNA target sites in 3′ UTRs of target mRNAs and their relatively better complementarity to the 5′ end of miRNAs. The first step for identifying genome-wide *Drosophila* miRNA targets involved generation of a conserved database comparing 3′ UTRs of *D. melanogaster* and *D. pseudoobscura* (101). The candidate target genes were then scored based on their free energy of base-pairing with the miRNAs, as determined by *mfold* (68, 101). The combination of sensitive sequence databases with that of the RNA folding algorithm confirmed the known targets and identified several new ones (101). A striking feature of predicted targets was the presence of clusters of functionally related targets regulated by specific miRNAs. This included *Notch* target genes for *mir-7*, proapoptotic genes for *mir-2* family, and metabolic pathway enzymes for *mir-277* (101). Another computational method for target identification, miRanda, relies on evolutionary relationships between miRNAs and their targets using three insect genomes (104). The miRanda approach is a three-phase method involving sequence matching of miRNA:mRNA pairs, estimating the energetics of the physical interaction and using evolutionary conservation as an informational filter. This method suggested both multiplicity (one miRNA targets several genes) and cooperativity (one gene targeted by several miRNAs) as general features of miRNA-regulated gene expression (104).

The TargetScan algorithm predicted more than 400 target genes for mammalian miRNAs (102). TargetScan also combines thermodynamics-based modeling of RNA:RNA duplex interactions with comparative sequence analysis to predict miRNA targets conserved across multiple genomes. One of the criteria for filtering miRNA:mRNA pairs using this algorithm is exact complementarity between 2–8 bases of miRNAs counted from the 5′end of miRNA. The folding energy of each pair is calculated using RNAeval (105), after extending the pairing as far as possible. Each 3′ UTR is then scored based on the number of miRNA:mRNA matches, free energy of interaction, and number and affinity of

complementary sites. Comparative ranking of UTRs among different organisms sorted on the basis of this score then predicts the target. TargetScan revealed that, in contrast to plant miRNA targets, only a small fraction of predicted mammalian targets participate in developmental control; they seem to regulate broadly diverse biological processes (102). Another computational program, DIANA-microT, was developed to study the rules of single miRNA:MRE (target mRNA) pairing and to predict targets containing a single complementary site (103). Similar to other computational programs, DIANA-microT identifies the putative targets by estimating the binding energies between conserved miRNA:MRE pairs. A difference from other programs is that it also takes into account the G-U wobble dinucleotide pairs for calculating binding energies (103).

Computational identification, based on favorable energy statistics and evolutionary relationships, corroborated by experimental evidence provides reasonable substantiation of miRNA target validity. An important consideration in computational target prediction and confirmation is the use of correctly annotated genes. Already, ambiguity in annotated genes misguided attempts to validate a miRNA target (106, 107). Absent or incomplete annotations of 3′ UTRs also hinder the comprehensive analysis of miRNA targets.

Heterologous reporter assays are most commonly used for validation of miRNA targets (Figure 6). Typically, a reporter gene, such as luciferase or β-galactosidase, is fused to sequences containing the miRNA complementary region from a putative target. Expression of the reporter is observed in the presence or absence of the proposed regulatory miRNA (6, 56, 90, 98, 99, 101-103). Down-regulation of a reporter gene in the presence of the miRNA indicates the presence of regulatory sites in the fragment used for fusion. However, concerns of extraneous effects due to multiple copies of complementary sites or very long UTR regions should be kept in mind. Also, failure to demonstrate regulation of a heterologous reporter may reflect factors independent of the miRNA: 1) the cell system might not express additional co-factors or adequate levels of the miRNA to appreciably affect reporter expression, 2) additional mRNA elements are required but not included in the UTR segment of the reporter, and 3) steric hindrance imposed by the fusion of reporter on the putative sites blocks interaction with the miRNA.

SUMMARY AND OUTLOOK

Genetics introduced us to the existence of miRNAs. Biochemical and molecular methods were essential for establishing the existence of vast numbers of miRNA genes in diverse organisms. Bioinformatic approaches contributed to the identification of additional, elusive, miRNAs as well as to the prediction of miRNA target genes. Combined experimental and computational methods will be required to advance our rudimentary understanding of miRNA expression and function. Central questions remain: How are transcription and processing of miRNAs regulated? How do miRNAs find their appropriate targets? What is the mechanism by which miRNAs regulate expression of their targets? The discovery of miRNAs established a new paradigm for gene regulation and understanding the biological roles of these abundant RNA genes undoubtedly will be a challenging endeavor.

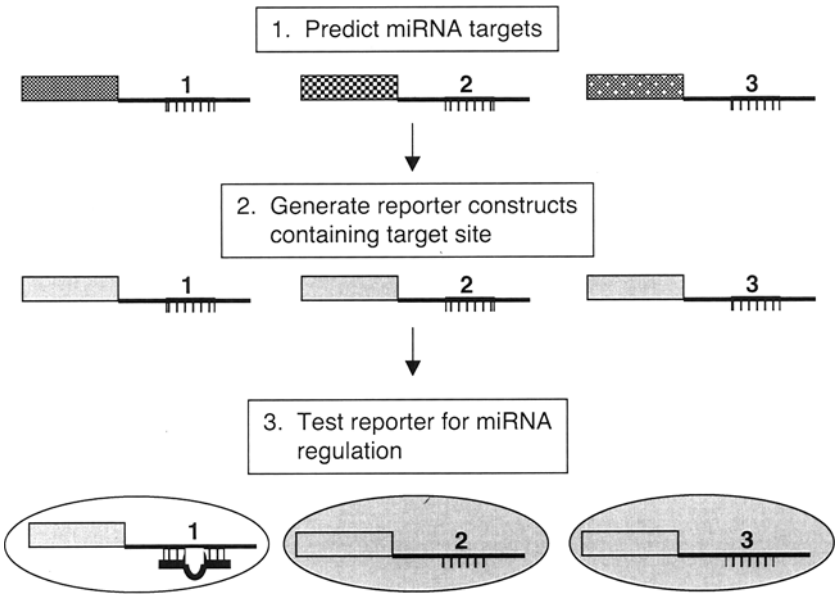


Figure 6. Validation of miRNA target predictions. Multiple targets with sites of partial complementarity are often identified for a single miRNA. To test the function of such predictions, sequences containing the complementary elements are fused to reporter genes, such as luciferase. The reporter is assayed for expression in the presence and absence of the candidate miRNA partner. In the depicted example, an miRNA is predicted to recognize sites in the 3′ UTRs of three different genes. Each UTR is fused to the reporter and the constructs are introduced to cells expressing the miRNA of interest. Only the site from gene "1" mediates reporter repression *via* interaction with the miRNA.

ACKNOWLEDGMENTS

We are grateful to members of the Pasquinelli laboratory for comments and inspiration. Research support is provided by the National Institutes of Health (GM071654-01) and the Searle Scholar, Peter Gruber, and V Foundations.

REFERENCES

1 Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) Cell 75, 843-854.
2 Wightman, B., Ha, I. and Ruvkun, G. (1993) Cell 75, 855-862.
3 Wightman, B., Burglin, T.R., Gatto, J., Arasu, P. and Ruvkun, G. (1991) Genes Dev. 5, 1813-1824.
4 Feinbaum, R. and Ambros, V. (1999) Dev. Biol. 210, 87-95.
5 Olsen, P.H. and Ambros, V. (1999) Dev. Biol. 216, 671-680.
6 Moss, E.G., Lee, R.C. and Ambros, V. (1997) Cell 88, 637-646.
7 Wickens, M. and Takayama, K. (1994) Nature 367, 17-18.
8 Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P.,

Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E. and Ruvkun, G. (2000) Nature 408, 86-89.
9 Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Nature 391, 806-811.
10 Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) Genes Dev. 15, 188-200.
11 Hamilton, A.J. and Baulcombe, D.C. (1999) Science 286, 950-952.
12 Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000) Nature 404, 293-296.
13 Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P. and Sharp, P.A. (1999) Genes Dev. 13, 3191-3197.
14 Zamore, P.D., Tuschl, T., Sharp, P.A. and Bartel, D.P. (2000) Cell 101, 25-33.
15 Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Science 294, 853-858.
16 Lee, R.C. and Ambros, V. (2001) Science 294, 862–864.
17 Llave, C., Kasschau, K.D., Rector, M.A. and Carrington, J.C. (2002) Plant Cell 14, 1605-1619.
18 Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) Curr. Biol. 12, 1484-1495.
19 Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. (2002) Genes Dev. 16, 1616-1626.
20 Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) Science 294, 858-862.
21 Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G. and Tuschl, T. (2003) RNA 9, 277-279.
22 Bartel, D.P. (2004) Cell 116, 281-297.
23 Chen, C.Z., Li, L., Lodish, H.F. and Bartel, D.P. (2004) Science 303, 83-86.
24 Zeng, Y., Wagner, E.J. and Cullen, B.R. (2002) Mol. Cell 9, 1327-1333.
25 Zeng, Y. and Cullen, B.R. (2003) RNA 9, 112-123.
26 Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H. and Kim, V.N. (2004) EMBO J. 23, 4051-4060.
27 Bracht, J., Hunter, S., Eachus, R., Weeks, P. and Pasquinelli, A.E. (2004) RNA 10, 1586-1594.
28 Aukerman, M.J. and Sakai, H. (2003) Plant Cell 15, 2730-2741.
29 Johnson, S.M., Lin, S.Y. and Slack, F.J. (2003) Dev. Biol. 259, 364-379.
30 Johnston, R.J. and Hobert, O. (2003) Nature 426, 845-849.
31 Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. and Kim, V.N. (2003) Nature 425, 415-419.
32 Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G. and Mello, C.C. (2001) Cell 106, 23-34.
33 Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. and Zamore, P.D. (2001) Science 293, 834-838.
34 Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. and Plasterk, R.H. (2001) Genes Dev. 15, 2654-2659.
35 Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) EMBO J. 21, 4663-4670.
36 Yi, R., Qin, Y., Macara, I.G. and Cullen, B.R. (2003) Genes Dev. 17, 3011-3016.

37 Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. and Kutay, U. (2004) Science 303, 95-98.
38 Bohnsack, M.T., Czaplinski, K. and Gorlich, D. (2004) RNA 10, 185-191.
39 Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004) Cell 118, 57-68.
40 Zhang, H., Kolb, F.A., Brondani, V., Billy, E. and Filipowicz, W. (2002) EMBO J. 21, 5875-5885.
41 Basyuk, E., Suavet, F., Doglio, A., Bordonne, R. and Bertrand, E. (2003) Nucl. Acids Res. 31, 6593-6597.
42 Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Cell 115, 209-216.
43 Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) Cell 115, 199-208.
44 Seggerson, K., Tang, L. and Moss, E.G. (2002) Dev. Biol. 243, 215-225.
45 Bartel, B. and Bartel, D.P. (2003) Plant Physiol. 132, 709-717.
46 Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. (2002) Cell 110, 513-520.
47 Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A. and Carrington, J.C. (2003) Dev. Cell 4, 205-217.
48 Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. (2002) Science 297, 2053-2056.
49 Xie, Z., Kasschau, K.D. and Carrington, J.C. (2003) Curr. Biol. 13, 784-789.
50 Tang, G., Reinhart, B.J., Bartel, D.P. and Zamore, P.D. (2003) Genes Dev. 17, 49-63.
51 Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. and Cohen, S.M. (2003) Cell 113, 25-36.
52 Doench, J.G., Petersen, C.P. and Sharp, P.A. (2003) Genes Dev. 17, 438-442.
53 Hutvagner, G. and Zamore, P.D. (2002) Science 297, 2056-2060.
54 Zeng, Y., Yi, R. and Cullen, B.R. (2003) Proc. Nat. Acad. Sci. U.S.A. 100 (17), 9779-9784.
55 Yekta, S., Shih, I.-H. and Bartel, D.P. (2004) Science 304, 594-596.
56 Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) Nature 403, 901-906.
57 Hipfner, D.R., Weigmann, K. and Cohen, S.M. (2002) Genetics 161, 1527-1537.
58 Xu, P., Vernooy, S.Y., Guo, M. and Hay, B.A. (2003) Curr. Biol. 13, 790-795.
59 Mette, M.F., van der Winden, J., Matzke, M. and Matzke, A.J. (2002) Plant Physiol. 130, 6-9.
60 Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002) Genes Dev. 16, 720-728.
61 Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) Curr. Biol. 12, 735-739.
62 Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. and Tuschl, T. (2003) RNA 9, 175-179.
63 Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. and Jewell, D. (2003) Curr. Biol. 13, 807-818.

64 Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J. and Tuschl, T. (2003) Dev. Cell 5, 337-350.
65 Smalheiser, N.R. (2003) Genome Biol. 4, 403.
66 Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Nature 409, 363-366.
67 Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. and Bartel, D.P. (2003) Genes Dev. 17, 991-1008.
68 Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) J. Mol. Biol. 288, 911-940.
69 Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P. and Burge, C.B. (2004) RNA 10, 1309-1322.
70 Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C. and Cavaille, J. (2003) Nat. Genet. 34, 261-262.
71 Jones-Rhoades, M.W. and Bartel, D.P. (2004) Mol. Cell 14, 787-799.
72 Bonnet, E., Wuyts, J., Rouze, P. and Van de Peer, Y. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 11511-11516.
73 Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. (2003) Genome Biol. 4, R42.
74 Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. and Bartel, D.P. (2003) Science 299, 1540.
75 Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G. and Kim, J. (2003) Mol. Cell 11, 1253-1263.
76 Wang, X.J., Reyes, J.L., Chua, N.H. and Gaasterland, T. (2004) Genome Biol. 5, R65.
77 Pasquinelli, A.E., McCoy, A., Jimenez, E., Salo, E., Ruvkun, G., Martindale, M.Q. and Baguna, J. (2003) Evol. Dev. 5, 372-378.
78 Schmittgen, T.D., Jiang, J., Liu, Q. and Yang, L. (2004) Nucl. Acids Res. 32, e43.
79 Allawi, H.T., Dahlberg, J.E., Olson, S., Lund, E., Olson, M., Ma, W.P., Takova, T., Neri, B.P. and Lyamichev, V.I. (2004) RNA 10, 1153-1161.
80 Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M., Alder, H., Bullrich, F. Negrin, M. and Croce, C.M. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 9740-9744.
81 Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K. and Kosik, K.S. (2003) RNA 9, 1274-1281.
82 Calin, G.A., Liu, C.-G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M., Dell'Aquila, M.L., Alder, H., Rassenti, L., Kipps, T.J., Bullrich, F., Negrini, M. and Croce, C.M. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 11755-11760.
83 McManus, M.T. (2003) Semin. Cancer Biol. 13, 253-258.
84 Chen, X. (2004) Science 303, 2022-2025.
85 Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A. and Timmermans, M.C. (2004) Nature 428, 84-88.

86 Kidner, C.A. and Martienssen, R.A. (2004) Nature 428, 81-84.
87 Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W. and Bier, E. (2004) Science 305, 846.
88 Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S. and Hobert, O. (2004) Nature 430, 785-789.
89 Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Nature 425, 257-263.
90 Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R. and Ruvkun, G. (2000) Mol. Cell 5, 659-669.
91 Boutla, A., Delidakis, C. and Tabler, M. (2003) Nucl. Acids Res. 31, 4973-4980.
92 Hutvagner, G., Simard, M.J., Mello, C.C. and Zamore, P.D. (2004) PLoS Biol. 2, 465-475.
93 Meister, G., Landthaler, M., Dorsett, Y. and Tuschl, T. (2004) RNA 10, 544-550.
94 Inoue, H., Hayase, Y., Imura, A., Iwai, S., Miura, K. and Ohtsuka, E. (1987) Nucl. Acids Res. 15, 6131-6148.
95 Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. and Slack, F.J. (2004) Genes Dev. 18 (2), 132-137.
96 Ambros, V. and Horvitz, H.R. (1984) Science 226, 409-416.
97 Ambros, V. (1989) Cell 57, 49-57.
98 Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A. and Rougvie, A.E. (2003) Dev. Cell 4, 625-637.
99 Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E. and Slack, F.J. (2003) Dev. Cell 4, 639-650.
100 Rajewsky, N. and Socci, N.D. (2004) Dev. Biol. 267, 529-535.
101 Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) PLoS Biol. 1, 397-409.
102 Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Cell 115, 787-798.
103 Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) Genes Dev. 18, 1165-1178.
104 Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) Genome Biol. 5, R1.
105 Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M. and Schuster, P. (1984) Monatshefte fur Chemie 125, 167-188.
106 Kawasaki, H. and Taira, K. (2003) Nature 423, 838-842.
107 Kawasaki, H. and Taira, K. (2003) Nature 426, 100.

DORMANCY AND THE CELL CYCLE

Michael A. Campbell

School of Science
Penn State Erie—The Behrend College
Erie, Pennsylvania 16563

INTRODUCTION

Dormancy can be viewed as a complex trait that encompasses physiological and developmental responses to environmental signals. There has been considerable research and emphasis in the past on the dormancy of seeds, which is to be expected in view of the significant impact to agricultural yields, but it has only been in the last few decades that significant research success has yielded information on the topic of bud dormancy. Thus, this review will focus on the recent results regarding bud dormancy and the interaction of dormancy to issues related to phytohormones and cell division.

TYPES OF DORMANCY AND THE CELL CYCLE

Biologically, dormancy is an avoidance response to drought, cold, or shortening days. The dormant situation is a complex set of physiological states and conditions in which plants respond to a series of stresses such as drought and

Genetic Engineering, Volume 27, Edited by J. K. Setlow

over-wintering by entering a state of growth suspension. This state of growth suspension can be exhibited by degrees of dormancy. The classification and degrees of dormancy found in plant vegetative structures have been defined as endodormancy, paradormancy, and ecodormancy (Figure 1) (1). Ecodormancy is a reduced growth response to an external stimulus such as drought or cold. Removal of the stimulus results in a resumption of growth. Paradormancy is a reduced growth response induced by a biochemical signal that is transported to a target tissue. Removal of the signal results in a resumption of growth. A good example of paradormancy is apical dominance, where auxin transported from an apical shoot suppresses the growth of lateral buds. Endodormancy is due to an endogenous signal that results in growth suppression. Older references often used the term "deep dormancy" to describe this phenomenon. In some situations, time is all that is required for endodormancy to terminate (e.g., the potato) and in others there is a need for a cold treatment in order to break the dormant state (such as most flowering temperate trees and shrubs).

Defining a dormant state in a plant can be difficult because a tissue may progress from one dormant condition to another without any phenotypic change. For example, a meristem may enter a phase of reduced growth in late summer in response to day-length or drought. Thus, the meristem could be considered to be in a state of ecodormancy. This situation could be followed by a physiological shift and a transition into a state of endodormancy. As winter progresses the endodormant state may terminate but growth will not resume due to unfavorable conditions and the meristem is now again in an ecodormant state. This shift between dormant states suggests that control of the growth cycle is a complex interaction between endogenous and environmental factors, and there is no simple genetic solution to increasing yields by manipulating dormancy. This concept is supported by breeding experiments and quantitative trait analysis, which is demonstrated by the significant genetic complexity controlling the onset and breakage of dormancy in poplar (2).

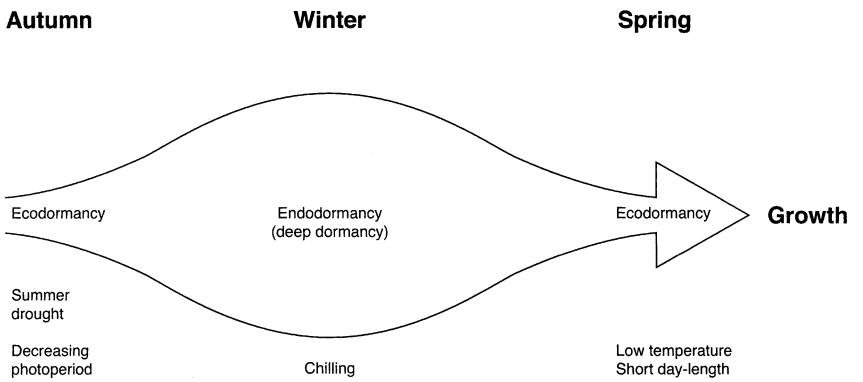


Figure 1. Diagram of dormancy states in a typical perennial bud. As the season progresses, dormancy types may shift between ecodormancy and endodormancy and back to ecodormancy. Redrawn from M. Lang et al. (1).

The variations and changes in the types of dormancy suggest that different biological and physiological mechanisms are involved with the changing of dormant states. As this pertains to cell cycle regulation, it is unclear whether different states of dormancy are associated with variations in cell cycle control. The cell cycle is regulated by a set of protein interactions between cell division kinases (CDKs), cyclins (CYCs), and inhibitors or regulators of the CDK/CYC complex [reviewed in (3)]. The activity of the CDK/CYC complex is associated with the establishment of restriction, or control, points throughout the cell cycle. These control points are positioned in the gap phases (G1 or G2) of the cell cycle (Figure 2). Thus, dormancy, which is characterized by very low or absent rates of cell division, must function by the established control points found in the G1 and G2 phases of the cell cycle (4). The arrest of cell division in endodormant ash buds (5), endodormant Helianthus buds (6), endodormant potato meristems (7), and paradormant axillary buds of pea (8) appear to be predominantly at the G1/S-phase of the cell cycle. This commonality of regulation, despite the specific type of dormancy, suggests that attempts to alter or control the dormant state in plants, with respect to the cell cycle state, would probably require manipulation of the factors or signal transduction mechanisms that interact with the G1/S restriction point. The interaction of various types of dormancy with the G1/S portion of the cell cycle demonstrates that artificial manipulation dormancy, by alteration of cell division or growth, will most likely involve genes or proteins that regulate the transition from the G1 to the S-phase of the cell cycle. At this point I do not suggest that direct alteration of genes regulating the G1/S restriction point is a practical solution to dormancy manipulation. Cell cycle control is

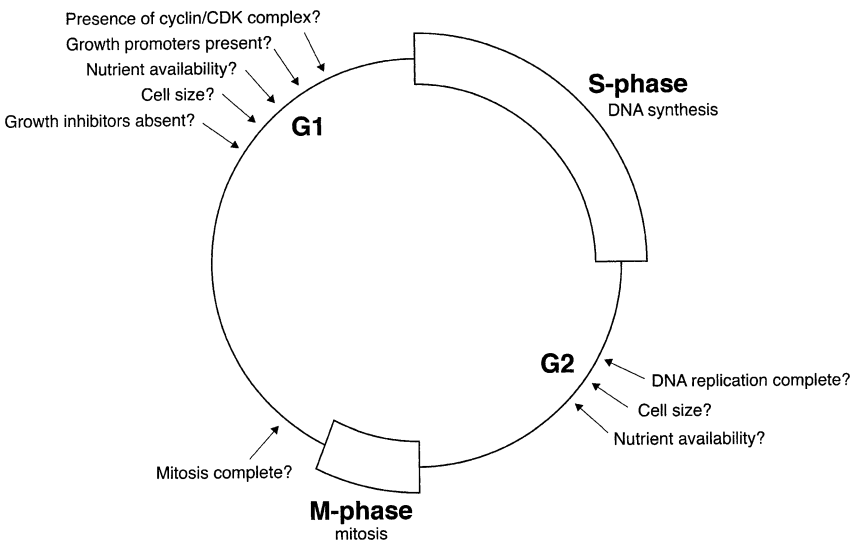


Figure 2. The major regulatory positions of plant cells.

fundamental to cell proliferation and survival. However, recent analysis of plant dormancy has narrowed the field to a number of potential targets that might impact dormancy onset and length without directly altering basic cell cycle machinery.

THE PARADORMANT STATE

Paradormancy, specifically apical arrest of lateral buds, is regulated by auxin but it is not clear whether this response is directly or indirectly associated with that phytohormone [reviewed by Horvath et al., 2003 (9, 10)]. Shimizu-Sato and Mori (11) recently reviewed the topic of dormancy regulation in lateral meristems and they proposed a scheme for hormonal regulation of growth (Figure 3). In this model, auxin, produced in the apical regions of the plant, does not directly inhibit cell division in the lateral meritems, but does result in abscisic acid (ABA)-induced genes in the node region and the lateral meristem. The model is supported by the work of Gocal et al. (12), who demonstrated that auxin levels in dormant meristems are low and they increase after release of apical repression. This suggests that, inasmuch as high levels of auxin are not present in paradormant lateral meristems, growth inhibition must result from some other signal. The increase in auxin as lateral buds are released from apical repression may have some association with entry into the cell cycle, since it has been shown that tobacco BY2 cells express D-type cyclins in response to auxin (13). Thus, a potential working model for paradormancy may be lateral meristem arrest by nodal ABA signals. Entry into the cell cycle begins following removal of the auxin/ABA inhibition and production of a phytohormonal signal, such as cytokinin, that initiates cell division.

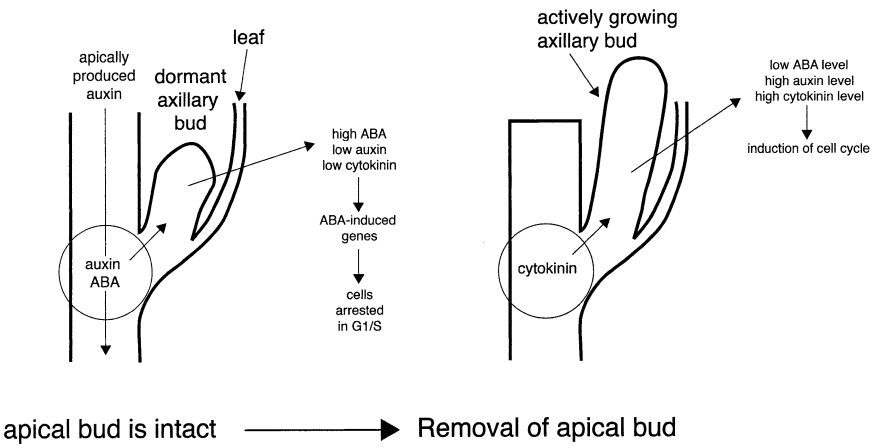


Figure 3. The mechanism of action in a paradormant axillary bud induced by apical dominance. Redrawn from Shimizu-Sato and Mori (11).

Transgenic plants that overproduce cytokinins exhibit a reduced level of lateral bud arrest (14, 15). Thus, it can be concluded that the interaction of cytokinins and auxins regulate lateral bud growth and paradormancy. These two hormones are probably functioning at two levels, an inhibitor process controlled by apically produced auxin/ABA and a growth-promoting process regulated by cytokinin.

THE ENDODORMANT STATE

In many perennial woody plants the breakage of the endodormant state requires chilling. It has been found that chilling results in a rearrangement of symplastic connections between cells of the apical meristem by the formation of 1,3-beta-D-glucan blockages (16). Thus, chilling induced capacity for cell division may be a result of removal of symplastic blockage by 1,3-beta-D-glucanase resulting in increased intercellular communication. It is possible that this increased communication results in the transport of hormonal signals throughout the meristem. However, it has not been shown that alterations of intercellular connections directly change hormonal transport or metabolism in meristems.

There has been recent research focused on the subject of vernalization, which is a requirement of some species for a cold treatment for the induction of flower development. Some parallels exist between the process of vernalization and cold-induced breakage of endodormancy. Vernalization has been shown to be an epigenetic process that requires prolonged exposure to cold temperatures, resulting in a developmental shift from vegetative to floral development [(17), reviewed in (18)]. Thus, vernalization and the breakage of endodormancy both require a similar temporal exposure to cold. Little is known about the molecular mechanisms associated with the breakage of endodormancy, but significant research advances have been accomplished regarding the process of vernalization. In Arabidopsis, vernalization has been shown to be an epigenetic process where extended exposure to cold results in the repression of the gene FLOWERING LOCUS C (FLC) with the use of chromatin remodeling induced by the genes *VRN1*, *VRN2,* and *VIN3* (19).

Exposure to cold can induce cold acclimation, vernalization, and the breakage of endodormancy in some species. Is there any similarity in the cold-induced regulation of these three processes? Cold treatment also induces elevated ABA levels and a series of cold-regulated genes controlled by the transcription factors CBF1, CBF2, and CBF3 (20, 21), but prolonged cold treatment reduces ABA in over-wintering endodormant buds. Liu et al. (22) demonstrated that vernalization is not regulated by ABA or the cold-induced transcription factors. While the onset of endodormancy appears to require ABA, it has been shown that cold is not necessary to induce the endodormant state in grape (23). However, to break endodormancy in some species, cold is a requirement and, in other species, such as potato, the breakage of endodormancy only requires time. A commonality between potato and species that require cold might be the reduction of ABA levels in meristems, shifting endodormancy to an ecodormant state. Another possibility is that the control of endodormancy is similar to vernalization, where chromatin remodeling, controlled by

ABA, time, and/or cold treatment is a requirement. Support for this hypothesis can be found in the work of Law and Suttle (24), who showed that in potato tubers demethylation of CCGG regions of DNA has been linked to the breakage of endodormancy. What remains to be determined is what regions of the genome are remodeled. Additionally, it would be important to discern whether a similar mechanism of remodeling is occurring in species that require cold treatment for the breakage of endodormancy and potato, which only requires a temporal exposure for dormancy loss. Is it possible that termination of endodormancy follows a pattern similar to that of vernalization: cold or time results in chromatin remodeling in an area of the genome that contains genes that suppress growth? Meristems that enter endodormancy have usually undergone a significant developmental shift with leaves replaced by bracts or bud scales at nodal regions and such developmental changes can be associated with chromatin remodeling (25).

IS THERE A UNIFYING THEORY TO DORMANCY?

Environmental stress, such as drought, induces the production of ABA, which results, with a complex set of responses, in growth arrest (ecodormancy). In paradormancy, in particular apical dominance, it has been demonstrated that there is an auxin-induced ABA response at nodal regions, which results in growth arrest in lateral meristems. Elevated ABA levels, at least in the potato system, induce endodormancy. Thus, a common theme in a number of plant systems, regardless of the type of dormancy, appears to involve an ABA response at some level.

However, the removal of ABA is not always sufficient to end the dormant state and initiate growth. Additional hormones, such as cytokinin and gibberellins, appear to be necessary for the resumption of cell division. Thus, dormancy, irrespective of the specific type, is controlled by a process of growth suppression and growth initiation. This might explain some of the genetic complexity found by breeders who are interested in the process of dormancy. The fact that both growth inhibitors and growth promoters regulate dormancy suggests that cell cycle control would follow a similar path; there would be cell cycle inhibitory mechanisms as well as cell cycle promotive mechanisms associated with dormancy. In potato endodormancy, growth arrest seems to occur upstream of the mechanisms of direct cell cycle control (7). This situation may be a result of endodormancy-inducing inhibitors of the cell cycle and that ABA has a central role in maintaining the endodormant state.

REGULATION OF THE G1 TO S TRANSITION OF THE PLANT CELL CYCLE

The cdk/cyclin protein complex regulates cell cycle transitions and, because dormancy appears to be a G1/S arrest, it is necessary to elucidate the components of the CDK/CYC complex associated with that arrest. Currently, specific proteins associated with dormancy G1/S arrest have not been found. In Arabidopsis, there are at least four different CDKs and the activity of one class

of the A-type of CDKs increases during the G1 to S-phase transition of the cell cycle [reviewed in (3, 26, 27)]. Potato meristems do not change in the levels of transcript for p34cdc2 kinase as endodormancy terminates (7). Thus, a working hypothesis is that dormancy regulates the activity, not the transcript levels, of a class of the A-type CDKs. The activity of a CDK requires the presence of specific cyclins, a specific phosphorylation state, and the absence of active inhibitors. This complex arrangement for CDK activity suggests that dormancy repression of the cell cycle at the G1 to S transition may result with the regulation of a number of different targets including cyclin levels, kinase activity, phosphatase activity, and the manipulation of inhibitors.

Plant cells contain a diverse population of cyclins, including A, B, D, and H-types (3, 28). More cyclins await description in plants, particularly in perennial species, but among the classes of cyclins known, the ones associated with G1/S cell cycle regulation are of direct interest to dormancy studies. The D-type cyclins have been shown to be associated with G1 to S-phase transitions in yeast (29). In Arabidopsis, genomic analysis has revealed that there are 49 different cyclins, which can be assigned to nine different subgroups: *CYC*A, *CYC*B, *CYC*C, *CYC*D, *CYC*H, *CYC*T, *CYC*L, *CYC*U, and *SDS* (30). Although function has not been determined for each of the cyclin-like genes, experimental evidence strongly suggests that the *CYC*D and *CYC*A classes are associated with the G1/S transition of the cell cycle [(27, 31), reviewed in (32-34)]. Thus, the *CYC*D and *CYC*A class of cyclins may be directly regulated by dormancy in plant tissues. It should be noted that cells not undergoing a cell cycle might exhibit low levels of many different classes of cyclins but dormant tissues, arrested in the G1 position, would first need to express the *CYC*D and *CYC*A proteins for entry into the S-phase.

The activity of the CDK/CYC complex is regulated by additional cellular and biochemical mechanisms. A class of proteins classified as CDK inhibitors (CKIs) interacts with the CDK/CYC complex and prevents cell cycle progression [reviewed in (33, 34)]. These inhibitors are interesting targets for investigating the interaction of dormancy and the cell cycle. In mammalian systems, G1/S-specific CKIs are represented by p21Cip1, p27Kip1, and p57Kip2 (17, 35-37). De Veylder et al. (38) have examined the activity of five Kip-related proteins (KRPs) in *Arabidopsis thaliana*. Thus, in comparison to mammalian systems, plants appear to have a greater diversity in KRP-type CDK inhibitors. Does this suggest that plant systems utilize a greater diversity of cell division inhibitors for spatial or temporal regulation of cell division? The results of De Veylder et al. (38) demonstrated more of a structural relationship between the KRPs and regulation of cell division. In Arabidopsis there are at least seven KRPs that appear to have diverse functions temporally and spatially in the shoot apex (39), but it is not clear how KRPs are associated with meristem activity and the process of dormancy. An additional class of cell division inhibitors called ICK1 and ICK2 has been identified in plants (40, 41). ICK1 has been shown to be induced by ABA (40), suggesting a relationship between a phytohormone associated with the dormancy response and a protein preventing entry into the cell cycle. The direct connection between ABA-induced dormancy and cell cycle inhibitors has yet to be adequately demonstrated.

THE PHYTOHORMONE CONNECTION

ABA has been shown to regulate response to drought, cold, salt stress, and seed dormancy through a complex set of fast and slow responses [reviewed in (42-45)]. The signal transduction mechanisms associated with ABA exposure in plants have recently been reviewed (42, 46), and currently there are about 50 genes associated with ABA responses in Arabidopsis (43), affecting more than 1,300 different transcripts (47). The interaction of the ABA signal transduction mechanism with genes or proteins that directly affect the dormancy response, which is a slow response, is not clear, and it has been difficult to separate ABA responses associated with stress and cold from those that are directly related to dormancy. ABA has been implicated with the onset and maintenance of endodormancy in potato (48), white birch (49), and lily (50). The interaction of ABA with the process of cell division is still not clear. Application or inhibition of ABA to meristematic tissues may alter the onset of endodormancy but it also results in developmental shifts resulting in the formation of bud scales in place of primordial leaf. Additionally, cross-talk between ABA and other hormones, particularly those that induce growth such as cytokinin, gibberellin, and auxin, complicates the experimental approaches necessary to elucidate specific responses. The role of ABA in seed dormancy and germination has progressed significantly (51), but due to a lack of a model system, ABA control of perennial meristem growth remains undefined. In Arabidopsis, an ABA application to germinating embryos results in reversible growth arrest. In tomato, ABA-deficient mutants exhibit an increase in cells arrested in G2/M, suggesting that ABA might regulate the G1/S restriction point. The ABA regulation at the G1/S restriction point may be due to cell cycle inhibitors such as ICK1, but there is some speculation that seedling dormancy might be a function of a p53-regulated process (52, 53). The idea that p53 might regulate cell division in seeds is based on the concept that seeds can be exposed to prolonged storage, resulting in environmentally induced DNA damage. This becomes an interesting issue in long-lived perennial species where lateral bud arrest (paradormancy) may occur on the order of hundreds or thousands of years and may result in significant DNA damage. However, the connection between ABA-induced stress and DNA damage has yet to be elucidated. ABA also results in reduced levels of metabolic activity, which might reflect a lack of nutrient mobilization. In animal systems it has been shown that serum starvation induces p53 activation and growth arrest by the ribosomal protein L11 (54).

In addition to the production of cell cycle inhibitors, ABA appears to interact with the phosphorylation cascade that is associated with the regulation of cell division [reviewed in (43, 46)]. ABA interacts with inositol polyphosphate 5-phosphatase (55), phospholipase C (55), cyclin-dependent kinase (56), protein phosphatase 2C (57, 58), and mitogen-activated protein kinase (59). Additionally, ABA is involved with the regulation of RNA metabolism including transcript abundance, RNA stability, transport and degradation [reviewed in (42)] and some of these transcripts may relate to cell cycle regulation.

Cytokinins have an important role in the breakage of dormancy (60-62). The resumption of cell division following dormancy is often associated with an increase in cytokinin levels. Within the last decade, significant progress has been made regarding the molecular and genetic mechanisms associated with cytokinin signal transduction [reviewed in (63)]. In endodormant tissues, such as potato, meristems change in their sensitivity to exogenous cytokinins; close to harvest, or deep dormancy, cytokinins have little effect on sprouting, but as tubers age the sensitivity to cytokinins increases (62). It has been suggested that endormant tubers do not respond to exogenous cytokinin because of the lack of a cytokinin receptor or inactivity of the signal transduction mechanism of cytokinin action (64). In Arabidopsis, it has been determined that receptors for cytokinins (AHK2, AHK3, CRE1/AHK4) are transmembrane histidine kinases (65-68). The expression of the cytokinin receptors appears to occur in all tissues of Arabidopsis (63) and the AHK receptors relay endogenous cytokinin signals resulting in shoot and root meristem growth (69). It is not known whether dormancy alters receptor levels. Interestingly, Arabidopsis cre1 mutants, which have a decreased response to endogenous cytokinin, exhibit a slight increased sensitivity to ABA (65), suggesting a level of cross-talk between the cytokinin and ABA signal transduction systems. Exogenous cytokinin induces D-type cyclin expression and cell division (70), and cytokinin-induced cell division can be replaced by overexpression of D-type cyclins (71). This suggests that as meristems become active following dormancy they enter the cell cycle by cytokinin-induced expression of D-type cyclins. Thus, entry into the cell cycle may not be controlled directly by the dormancy process but by phytohormone production after tissues have exited the dormant state.

In addition to being associated with the breakage of seed dormancy, gibberellins (GAs) have been linked with dormancy release in tulips [reviewed in (72), potato (73), and lily bulbs (74, 75)]. However, the complexity of GA types in cells and tissues, the possible presence of inhibitors, and the difficulty assessing the specific dormant state makes it unclear whether GA is involved with the breakage of dormancy or is a postdormancy growth response. Results by Horvath et al. (76) suggest that the application of GA_3 to leafy spurge resulted in G1 to S-phase transition in adventitious buds (Figure 4).

GA has been shown to bind to a GCR1 receptor in *A. thaliana* seeds (77). Additionally, G-protein-type receptors are associated with GA perception (78). These types of binding by GA probably result in a signal transduction cascade that has yet to be defined in its entirety, but ultimately there must be some impact on the cell cycle machinery. In deepwater rice, GA application induces cell division (79, 80). The regulation of GA on rice cell division appears to be largely at the G2/M restriction point (81). One of the responses of rice to exogenous GA is to increase the levels of transcripts for mitotic cyclins and a specific class of cyclin-division kinase (82, 83). In the meristems of dicots, specifically tomato, GA induces the expression of transcripts for expansins, proteins that alter cell wall extensibility and cell expansion (84). The change in expansion expression may suggest another avenue for GA impact on cell cycle machinery, since cell size has been associated with cell cycle regulation in a number of eukaryotic organisms.

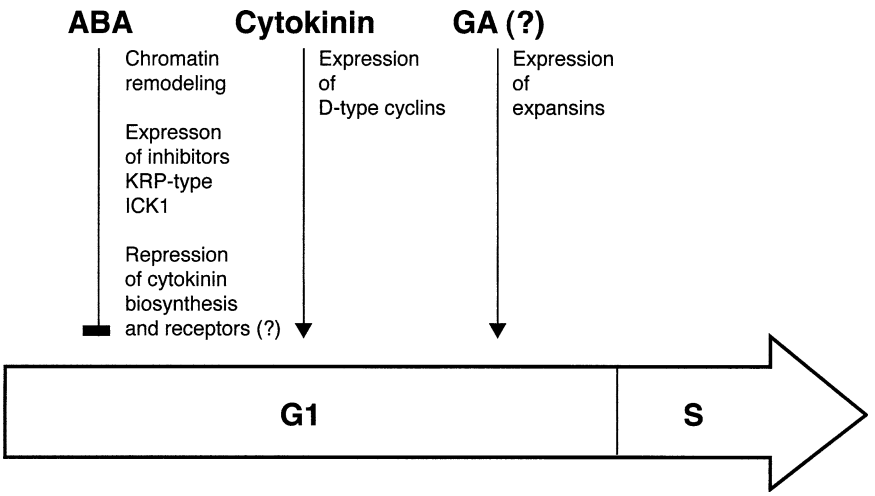


Figure 4. Possible regulatory steps in G1 by phytohormones that regulate dormancy.

CONCLUSIONS

The localization of the Kip-related proteins to meristematic regions suggests that these proteins might be associated with dormancy regulation. Additionally, the localization of chromatin remodeling to specific genes and loci may reveal the important regulation mechanisms for the dormant state. Developmental mutants that fail to develop bud scales and shift meristem programs toward an over-wintering bud would be informative in identifying.

A substantial body of work has been accomplished in identifying genes and proteins associated with cell division and the cell cycle in plants. However, most of the recent research has focused on the annual *A. thaliana* as a model. In order to progress rapidly in the area of plant meristem dormancy, a model system has to be adopted. Many perennial species are slow growing, genetically complex, and have little background genetic research that can be used to support dormancy studies. Recent advances into the genetic structure of poplar and recent interest in some perennial relatives of *A. thaliana* may create opportunities for models systems.

REFERENCES

1 Lang, G.E., Early, J.D., Martin, G.C. and Darnell, R.L. (1987) HortScience, 22(3), 371-377.

2 Chen, T.H.H., Davis, J., Frewen, B.E., Howe, G.T. and Bradshaw, H.D. (2000) in Dormancy in Plants: From Whole Plant Behavior to Cellular Control (J.-D. Viemont and J. Crabbe, eds.) CABI Publishing, New York, NY. pp. 319-330.

3 Dewitte, W. and Murray, J.A. (2003) Ann. Rev. Plant Physiol. Plant Mol. Biol. 54, 235-264.

4 Van't Hof, J. (1985) in The Cell Division Cycle in Plants (J.A. Bryant and D. Francis, eds.) Cambridge University Press, pp. 1-13.
5 Cottignies, A. (1979) Planta 147, 15-19.
6 Tepfer, S.S., Nougarede, A. and Rondet, P. (1981) Canadian J. Bot. 59, 1918-1940.
7 Campbell, M.A., Suttle, J.C. and Sell, T.W. (1996) Physiol. Plant. 98, 743-752.
8 Devitt, M.L. and Stafstrom, J.P. (1995) Plant Mol. Biol. 29 (2), 255-265.
9 Horvath, D.P., Anderson, J.V., Chao, W.S. and Foley, M.E. (2003) Trends Plant Sci. 8(11), 534-540.
10 Cline, M.G. (1994) Physiol. Plant. 90, 230-237.
11 Shimizu-Sato, S. and Mori, H. (2001) Plant Physiol. 127(4), 1405-1413.
12 Gocal, G.F.W., Pharis, R.P., Yeung, E.C. and Pearce, D. (1991) Plant Physiol. 95, 344-350.
13 Sorrell, D.A., Combettes, B., Chaubet-Gigot, N. and Murray, J.A. (1999) Plant Physiol. 119(1), 343-352.
14 Medford, H., Horgan, R., El-Sawi, Z. and Klee, H.J. (1989) Plant Cell 1(4), 403-413.
15 Romano, C.P., Hein, M.B. and Klee, H.J. (1991) Genes Dev. 5(3), 438-446.
16 Rinne, P.L.H., Kaikuranta, P.M. and Schoot, C. van der (2001) Plant J. 26(3), 249-264.
17 Lee, I. and Amasino, R.M. (1995) Plant Physiol. 108(1), 157-162.
18 Sung, S. and Amasino, R.M. (2004) Curr. Opin. Plant Biol. 7, 4-10.
19 Sung, S. and Amasino, R.M. (2004) Nature 427, 159-164.
20 Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M. and Thomashow, M.F. (1998) Plant J. 16(4), 433-442.
21 Medina, J., Bargues, M., Perez-Alonso, T.J.M. and Salinas, J. (1999) Plant Physiol. 119(2), 463-470.
22 Liu, J., Gilmour, S.J., Thomashow, M.F. and van Nocker, S. (2002) Physiol. Plant. 114, 125-134.
23 Fennell, A. and Hoover, E. (1991) J. Amer. Soc. Hort. 116, 270-273.
24 Law, R.D. and Suttle, J.C. (2003) Plant Mol. Biol. 51(3), 437-447.
25 Wagner, D. (2003) Curr. Opin. Plant Biol. 6, 20-28.
26 Hemerly, A., Engler, de A.J., Bergounioux, C., Van Montagu, M., Engler, G. Inze, D. and Ferreira, P. (1995) EMBO J. 14, 3925-3936.
27 Magyar, Z., Meszaros, T., Miskolezi, P., Deak, M., Feher, A., Brown, S., Kondo, E., Athanasiadis, A., Pongor, S., Bilgin, M., Bako, I., Koncz, C. and Dudits, D. (1997) Plant Cell 9, 223-235.
28 Vandepoele, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S. and Inze, D. (2002) Plant Cell 14(4), 903-916.
29 Dahl, M., Meskiene, I., Bogre, L., Ha, D.T., Swoboda, I., Hubmann, R., Hirt, H. and Heberle-Bors, E., (1995) Plant Cell 7, 1847-1857.
30 Wang, G., Kong, H., Sun, Y., Zhang, X., Zhang, W., Altman, N., DePamphilis, C. and Ma, H. (2004) Plant Physiol. 135, 1084-1099.
31 Meijer, M. and Murray, J.A.H. (2000) Plant Mol. Biol. 43, 621-633.
32 Stals, H. and Inze, D. (2001) Trends Plant Sci. 6, 359-364.

33 Rossi, V. and Varotto, S. (2002) Planta 215, 345-356.
34 De Veylder, L., Joubes, J. and Inze, D. (2003) Curr. Opin. Plant Biol. 6, 536-543.
35 Toyoshima, H. and Hunter, T. (1994) Cell 78, 67-74.
36 Zhang, P., Wong, C., DePinho, R.A., Harper, J.W. and Elledge, S.J. (1998) Genes Dev. 12, 3162-3167.
37 Zhang, P., Wong, C., Liu, D., Finegold, M., Harper, J.W. and Elledge, S.J. (1999) Genes Dev. 13, 213-224.
38 De Veylder, L., Beeckman, T., Beenster, G.T.S., Krols, L., Terrasa, T., Landrieu, I., Van Der Schueren, E., Maes, S., Naudts, M. and Inze, D. (2001) Plant Cell 13, 1653-1667.
39 Ormenese, S., de Almeida Engler, J., De Groodt, R., De Veylder, L., Inze, D. and Jacqmard, A. (2004) Ann. Bot. (London) 93(5), 575-580.
40 Wang, H., Qi, Q., Schorr, P., Cutler, A.J., Crosby, W.L. and Fowke, L.C (1998) Plant J. 15, 501-510.
41 Lui, H., Wang, H., DeLong, C., Fowke, L.C., Crosby, W.L. and Fobert, P.R. (2000) Plant J. 21, 379-385.
42 Kuhn, J.M. and Schroeder, J.L. (2003) Curr. Opin. Plant Biol. 6, 463-469.
43 Finkelstein, R.R., Gampala, S.S.L. and Rock, C.D. (2002) Plant Cell 14, 515-545.
44 Zhu, J., Gong, Z., Zhang, C., Song, C.P., Damsz, B., Inan, G., Koiwa, H., Zhu, J.K., Hasegawa, P.M. and Bressan, R.A. (2002) Plant Cell 14, 3009-3028.
45 Leung, J., Leung, J., Merlot, S., Gosti, F., Bertauche, N., Blatt, M.R. and Giraudat, J. (1998) Symp. Soc. Exp. Biol. 51, 65-71.
46 Himmelbach, A., Yang, Y. and Grill, E. (2003) Curr. Opin. Plant Biol. 6, 470-479.
47 Hoth, S., Morgante, M., Sanchez, J.P., Hanafey, M.K., Tingey, S.V. and Chua, N.H. (2002) J. Cell Sci. 115, 4891-4900.
48 Suttle, J.C. and Hultstrand, J.F. (1994) Plant Physiol. 105(3), 891-896.
49 Rinne, P., Tuominen, H. and Junttila, O. (1994) Tree Physiol. 14(6), 549-561.
50 Kim, K.S., Davelaar, E. and De Klerk, G.J. (1994) Physiol. Plant. 90, 59-64.
51 Koornneef, M., Bentsink, L. and Hilhorst, H. (2002) Curr. Opin. Plant Biol. 5(1), 33-36.
52 Whittle, C.A., Beardmore, T. and Johnston, M.O. (2001) Trends Plant Sci. 6(6), 248-251.
53 Korthout, H.A.A.J., Martien, P.M., Caspers, M.J., Kottenhagen, Q.H. and Wang, M. (2002) FEBS Lett. 526, 53-57.
54 Bhat, K.P., Itahana, K., Jin, A. and Zhang, Y. (2004) EMBO J. 23, 2402-2412.
55 Sanchez, J.P. and Chua, N.H. (2001) Plant Cell 13, 1143-1154.
56 Hemerly, A.S., Ferreira, P., de Almeida Engler, J., Van Montagu, M., Engler, G. nd Inze, D. (1993) Plant Cell 5, 1711-1723.
57 Sheen, J. (1998) Proc. Nat. Acad. Sci. U.S.A. 95, 975-980.

58 Tahtiharju, S. and Palva, T. (2001) Plant J. 26, 461-470.
59 Lu, C., Han, M.-H., Guevara-Garcia, A. and Fedoroff, N.V. (2002) Proc. Nat. Acad. Sci. U.S.A. 99(24), 15812-15817.
60 Hemberg, T. (1970) Physiol. Plant. 23, 850-858.
61 Arias, O. and Crabbe, J. (1975) Physiol. Veg. 13, 69-81.
62 Suttle, J.C. (1998) Physiol. Plant. 103, 59-69.
63 Heyl, A. and Schmulling, T. (2003) Curr. Opin. Plant Biol. 6, 480-488.
64 Suttle, J.C. (2000) in Dormancy in Plants from Whole Plant Behavior to Cellular Control (Viemont, J.-D. and Crabbe, J., eds.) CABI Publishing, New York, NY. pp. 211-226.
65 Inoue, T., Higuchi, M., Hashimoto, Y., Seki, M., Kobayashi, M., Kato, T., Tabata, S., Shinozaki, K. and Kakimoto, T. (2001) Nature 409, 1060-1063.
66 Suzuki, T., Miwa, K., Ishikawa, K., Yamada, H., Aiba, H. and Mizuno, T. (2001) Plant Cell Physiol. 42, 107-113.
67 Ueguchi, C., Sato, S., Kato, T. and Tabata, T. (2001) Plant Cell Physiol. 42, 751-755.
68 Yamada, H., Suzuki, T., Terada, K., Takei, K., Ishikawa, K., Miwa, K. and Mizuno, T. (2001) Plant Cell Physiol. 42, 1017-1023.
69 Nishimura, C., Ohashi, Y., Sato, S., Kato, T., Tabata, S. and Ueguchi, C. (2004) Plant Cell 16, 1365-1377.
70 Soni, R., Carmichael, J.P., Shah, Z.H. and Murray, J.A. (1995) Plant Cell 7, 85-103.
71 Riou-Khamlichi, C., Menges, M., Healy, J.M. and Murray, J.A (2000) Mol. Cell. Biol. 20(13), 4513-4521.
72 Saniewski, M., Kawa-Miszczak, L., Wegrzynowicz-Lesiak, E. and Okubo, H. (2000) in Dormancy in Plants: from Whole Plant Behavior to Cellular Control (Viemont, J.-D. and Crabbe, J., eds.) CABI Publishers: New York, NY. pp. 227-244.
73 Rappaport, L. and Wolf, N. (1969) Symp. Soc. Exp. Biol. 23, 219-40.
74 Ohkawa, K. (1979) Sci. Hort. 10, 255-260.
75 Nimi, Y., Endo, Y. and Arisaka, E. (1988) J. Jap. Soc. Hort. Sci. 57, 250-257.
76 Horvath, D.P., Chao, W.S. and Anderson, J.V. (2002) Plant Physiol. 128(4), 1439-1446.
77 Chen, J.-G., Pandey, S., Huang, J., Alonso, J.M., Ecker, J.R., Assmann, S.M. and Jones, A.M. (2004) Plant Physiol. 135, 907-915.
78 Iwasaki, Y., Fujisawa, Y. and Kato, H. (2003) J. Plant Growth Regul. 22, 126-133.
79 Metraux, M.C. and Kende, H. (1984) Planta 160, 73-77.
80 Raskin, I. and Kende, H. (1984) Plant Physiol. 76, 947-950.
81 Sauter, M. and Kende, H. (1992) Planta 190, 354-362.
82 Sauter, M. (1997) Plant J. 11, 181-190.
83 Fabian, T., Lorbiecke, R., Umeda, M. and Sauter, M. (2000) Planta 211, 376-383.
84 Vogler, H., Caderas, D., Mandel, T. and Kuhlemeier, C. (2003) Plant Mol. Biol. 53, 267-272.

THE OPT FAMILY FUNCTIONS IN LONG-DISTANCE PEPTIDE AND METAL TRANSPORT IN PLANTS

Mark Lubkowitz

Biology Department
Saint Michael's College
Colchester, Vermont 05477

INTRODUCTION

The physical properties of cellular membranes limit the compounds that can diffuse across membrane barriers, and most of the molecules important to life cannot cross without the aid of transport proteins. Transporters play a crucial role in many essential processes such as respiration, photosynthesis, and nutrient acquisition. As a group, they differ in function, structure, substrates, energy source, and cellular location, and more than 360 different families are recognized (1). This review covers the recently discovered and largely uncharacterized oligopeptide transporter (OPT) family.

Peptide transport is the protein-mediated process of translocating small peptides across a membrane (2). Peptide transport proteins have been found in three different families: the ATP binding cassette (ABC) superfamily, the peptide transporter (PTR) superfamily, and the oligopeptide transporter (OPT)family. The ABC and PTR groups are considered to be superfamilies because members are able to transport a diverse range of substrates, whereas it appears that OPTs

Genetic Engineering, Volume 27, Edited by J. K. Setlow

transport amino-acid-based compounds. The OPT family was named after the first identified substrates, peptides of three to five residues (3, 4), although the list of substrates has been expanded to include metal-binding secondary amino acids (5). This review begins by explaining how the OPT family was distinguished from ABC and PTR proteins and then focuses on the eight papers published since the last review (2), all of which were studies in plants, emphasizing the known and hypothesized biological roles. OPTs in fungi will only be briefly reviewed inasmuch as no new discoveries have been made since the last review in fungi (6).

WHAT DEFINES THE OPT FAMILY?

Jeff Becker's laboratory first defined the OPT family using a criterion of sequence similarity, common function (oligopeptide transport), and several conserved motifs to distinguish this group from other established transporters (4). This study found that *CaOPT1* from *Candida albicans, Spisp4* from *Schizosaccharomyces pombe,* and *ScOPT2* from *S. cerevisiae* all conferred to *S. cerevisiae* the ability to transport various peptides of three to five residues. Sequence analyses quickly revealed that these proteins lacked the characteristic signature motifs found in either the ABC or PTR protein families. This novel group of proteins was named the OPT family based on the length of the substrates (three to five residues) and to distinguish it from the PTR-type peptide transporters that translocate peptides of two to three residues.

More recently, bioinformaticists from InterPro (7), TIGR (8), and Pfam (9) have used *hidden Markov modeling* (HMM) to define more rigorously the sequence similarities of OPT proteins, and there are currently 75 full-length OPT sequences recognized in the public databases using the HMM criterion (Table 1). A close inspection of the organisms listed in Table 1 reveals that OPTs have only been found in fungi, bacteria, plants, and archaea. The number of known OPTs will undoubtedly grow as more genome projects are completed, but it seems unlikely that an OPT gene will be found in eukaryotes other than plants and fungi given the diverse range of representative species that have been sequenced.

Phylogenetic analyses (10) have revealed that the OPT family has two distinct clades, which are here termed the yellow stripe (YS) clade and the *peptide transport* (PT) clade, based on the first characterized members of each group (4, 5). OPTs in the YS clade are present in archaea, bacteria, plants, and fungi, whereas members of the PT clade only occur in plants and fungi. A partial phylogenetic tree of the OPT family, with the 26 representative OPTs used by Pfam to construct the OPT hidden Markov model, illustrates the clades [Table 1 (asterisks), Figure 1]. Of the OPTs listed in Table 1 and Figure 1, only a limited number have been experimentally characterized, but four general trends have emerged from these initial studies: 1) OPTs are energized by the symport of protons; 2) two distinct clades are present, the PT clade and YS clade; 3) substrates for the PT clade tend to be peptides of three to five residues and substrates of the YS clade are metal-chelating amino acids; and 4) OPTs function as importers, not efflux pumps. Given the small number of OPTs that have been studied, it is too early to determine whether these trends will hold true for all OPTs. The biological roles of individual OPTs are broad and will be addressed later in the review.

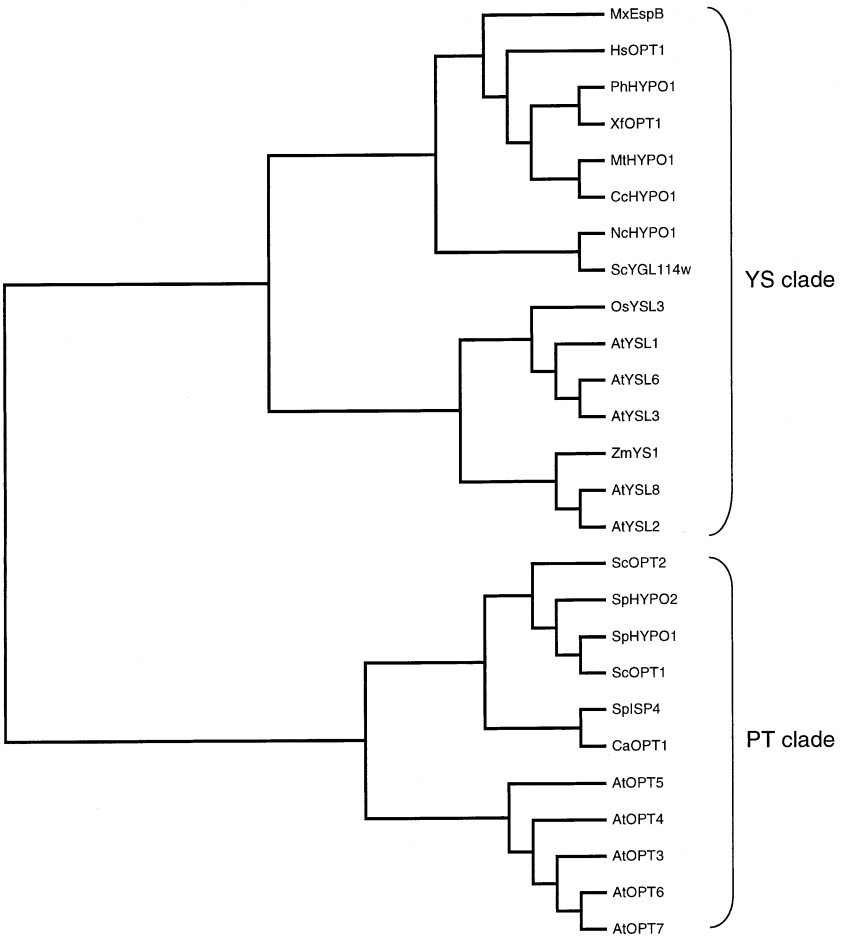


Figure 1. Partial phylogenetic tree of the OPT family. Twenty-six representative OPTs were used to construct an unrooted tree using the Clustal W algorithm. The tree reveals two clades, yellow stripe (YS) and peptide transport (PT).

Three studies have shown that transporter activity is pH dependent, with acidic conditions leading to higher transport rates. Bogs et al. (11) expressed *GlutathioneTransporter1* from *Brassica juncea* (*BjGT1*) in yeast and demonstrated that the optimal pH for transporter activity was 5.0 and that activity was greatly reduced in the presence of the protonophore *carbonylcyanid-m-chlorophenylhydrazone* (CCCP). Similarly, Schaaf et al. (12) showed that in *Zea mays* roots Yellow Stripe1 (ZmYS1) activity is highest at pH 4.5 and is also sensitive to CCCP. Finally, Hauser et al. (13) demonstrated that peptide transporter activity conferred by *ScOPT1* (also named *HGT1*/*GSH1*) from *S. cerevisiae* is highest at pH 5.5 and is sensitive to CCCP. The consensus from these three experiments is that OPTs use the proton motive force to drive importation. Given that

Table 1. OPT proteins identified by hidden Markov modeling. The family has been divided into two groupings, YS clade (yellow stripe) and PT clade (peptide transport), based on the first characterized transporters and separated into plant, fungi, archaea, and bacteria subgroupings. OPTs used to construct the hidden Markov model and phylogenetic tree in Figure 1 are designated by asterisks. Rice sequences that only appear in the TIGR database are indicated by (**) and the corresponding number is a TIGR index number.

Protein	Size (aa)	Organism	Genbank index number
YS clade			
Plants			
AtYSL1*	665	*Arabidopsis thaliana*	41352036
AtYSL2*	652	*Arabidopsis thaliana*	41352038
AtYSL3*	669	*Arabidopsis thaliana*	9759194
AtYSL4	670	*Arabidopsis thaliana*	41352040
AtYSL5	714	*Arabidopsis thaliana*	41352042
AtYSL6*	676	*Arabidopsis thaliana*	41352044
AtYSL7	688	*Arabidopsis thaliana*	41352046
AtYSL8*	724	*Arabidopsis thaliana*	41352048
NtA17	573	*Nicotiana tabacum*	27529843
OsYSL1	712	*Oryza Sativa*	32487645
OsYSL2	717	*Oryza Sativa*	32489855
OsYSL3*	708	*Oryza Sativa*	11034704
OsYSL4	724	*Oryza Sativa*	32489854
OsYSL5	675	*Oryza Sativa*	38345944
OsYSL6	695	*Oryza Sativa*	38344923
OsYSL7	724	*Oryza Sativa*	38347209
OsYSL8	678	*Oryza Sativa*	32483288
OsYSL9	684	*Oryza Sativa*	38345941
OsYSL10	679	*Oryza Sativa*	15624064
OsYSL11	694	*Oryza Sativa*	9630.t00144**
OsYSL12	755	*Oryza Sativa*	9630.t03900**
ZmYS1*	682	*Zea mays*	10770864
Fungi			
AfOPT1	843	*Aspergillus fumigatus*	20145239
NcHYPO1*	738	*Neurospora crassa*	9368956
ScYGL114w*	725	*Saccharomyces cerevisiae*	1322664

Archaea			
HsOPT1*	655	*Halobacterium sp*. NRC-1	10584311
PhHYPO1*	626	*Pyrococcus horikoshii*	3256752
Bacteria			
BbOPT1	693	*Bordetella bronchiseptica*	33577087
BpOPT1	682	*Bordetella pertussis*	33571903
BtOPT1	662	*Bacteroides thetaiotaomicron*	29338392
CbOPT1	669	*Coxiella burnetii*	29541704
CcHYPO1*	666	*Caulobacter crescentus*	13424225
CjHYPO1	665	*Campylobacter jejuni*	6967697
CpOPT1	638	*Clostridium perfringens*	18144939
HdOPT1	669	*Haemophilus ducreyi*	33149099
LsHYPO1	645	*Lactobacillus sakei*	1370208
MtHYPO1*	667	*Mycobacterium tuberculosis*	1655665
MxEspB*	592	*Myxococcus xanthus*	9313036
NmNme	672	*Neisseria meningitidis*	7378976
PaHYPO1	678	*Pseudomonas aeruginosa*	9950120
PgOPT1	659	*Porphyromonas gingivalis*	34396577
PsOPT1	578	*Pseudomonas syringae*	28850622
RsHYPO1	683	*Ralstonia solanacearum*	17430320
SvHYPO1	553	*Streptomyces viridochromogenes*	22095157
TtOPT1	647	*Thermoanaerobacter tengcongensis*	20515137
XaOPT1	656	*Xanthomonas axonopodis*	21107349
XcOPT1	656	*Xanthomonas campestris*	21112137
XfOPT1*	653	*Xylella fastidiosa*	9107417
PT Clade			
Plants			
AtOPT1	755	*Arabidopsis thaliana*	9758213
AtOPT2	734	*Arabidopsis thaliana*	15218331
AtOPT3*	637	*Arabidopsis thaliana*	18414644
AtOPT4*	729	*Arabidopsis thaliana*	9759417
AtOPT5*	753	*Arabidopsis thaliana*	4938497
AtOPT6*	736	*Arabidopsis thaliana*	4469024

(*Cont.*)

Table 1. (*cont.*)

Protein	Size (aa)	Organism	Genbank index number
AtOPT7*	766	*Arabidopsis thaliana*	3600039
AtOPT8	733	*Arabidopsis thaliana*	9759191
AtOPT9	741	*Arabidopsis thaliana*	9759190
BjGT1	661	*Brassica juncea*	41351489
OsOPT1	755	*Oryza Sativa*	20160527
OsOPT2	786	*Oryza Sativa*	6983868
OsGT1	766	*Oryza Sativa*	6983869
OsOPT4	769	*Oryza Sativa*	6983880
OsOPT5	733	*Oryza Sativa*	9636.t02194**
OsOPT6	737	*Oryza Sativa*	32489380
OsOPT7	726	*Oryza Sativa*	28144882
OsOPT8	914	*Oryza Sativa*	9630.t04393**
OsOPT9	752	*Oryza Sativa*	9636.t03769
Fungi			
CaOPT1*	783	*Candida albicans*	2367386
ScOPT1*	799	*Saccharomyces cerevisiae*	1015595
ScOPT2*	877	*Saccharomyces cerevisiae*	6325452
SpHYPO1*	851	*Schizosaccharomyces pombe*	7491035
SpHYPO2*	791	*Schizosaccharomyces pombe*	3136031
SpISP4*	776	*Schizosaccharomyces pombe*	218542
ScoMTD1	777	*Schizophyllum commune*	6716399
YlOPT1	836	*Yarrowia lipolytica*	18076958

BjGT1 and ZmYS1 are from two distant clades (Table 1), it seems likely that all OPTs use a proton gradient as an energy source. It should be noted that none of these experiments addressed the stoichiometric relationship between substrate and protons.

Conclusions regarding substrate preferences of OPTs are difficult to draw very simply because so few OPTs have been characterized. However, all substrates identified thus far, with one exception, fall into two categories: small peptides of three to five residues and secondary amino acids that bind metals, specifically phytosiderophores and *nicotianamine* (NA). The one exception is the recent finding that AtOPT3 from *Arabidopsis thaliana* may transport metals (14). Peptides and metal-binding secondary amino acids share an interesting link; both are synthesized from a small number of amino acids. Peptides, by definition, consist of

amino acids joined by a peptide bond and most peptide substrates identified for OPTs are three to five residues in length. On the other hand, phytosiderophores and NA are synthesized from three S-adenosylmethionines to produce a secondary amino acid. This commonality possibly reflects a conserved mechanism for substrate recognition and/or translocation.

OPTS OF THE YS CLADE TRANSPORT METAL-CHELATES

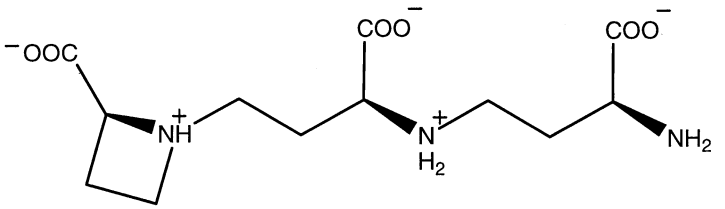
Studies from Elsbeth Walker's, Nicolaus von Wirén's, and Naoko Nishizawa's laboratories demonstrated that OPTs in the YS clade play a substantial role in metal acquisition and long-distance transport in plants (5, 12, 15-17). Metals such as iron, zinc, magnesium, manganese, and copper are necessary for many enzymatic functions but are also capable of inflicting extensive oxidative damage to cells due to their reactive nature. Therefore, long-distance transport requires nonreactive carrier molecules to shuttle metal ions from source (roots) to sink (such as leaves). Two of these carrier molecules, NA and phytosiderophores, are substrates for several OPTs when complexed with metals. The specific contribution to metal homeostasis of the two types of carrier molecules is quite different even though they are structurally very similar. Phytosiderophores are secreted by graminaceous monocots (i.e., grasses) into the soil to chelate metals, particularly iron. NA, which is not secreted into the soil, functions as a carrier molecule for long-distance metal transport in all plants and may be involved in intracellular transport (17).

NA is a secondary amino acid that is synthesized enzymatically from three molecules of S-adenosylmethionine. Furthermore, NA is the precursor for all phytosiderophores (Figure 2). Phytosiderophores are also secondary amino acids and belong to a family of compounds called *mugineic acids* (MAs). As discussed below, the primary role of MAs is to solubilize iron in the soil through chelation (18).

Iron is an important micronutrient because it plays a critical role in many enzymatic reactions and is a key component of the electron carriers used in photosynthesis and respiration. Even though iron is the fourth most abundant element in the earth's crust, iron concentrations in soils are often below the level necessary to support robust plant and microbial life because of low solubility (19). Furthermore, the concentration of soluble iron increases as the pH of the soil decreases, which means that in calcareous soils, or soils containing sodium bicarbonate, very little iron is available to organisms scavenging the environment. Organisms require ferrous iron for cellular functions but the valence state in the soil is Fe(III). Therefore, plants must surmount three challenges to acquire iron: they must transport iron across the plasma membrane, reduce Fe(III) to Fe(II), and transport the metal to the proper cellular and subcellular location while preventing the generation of iron-catalyzed hydroxyl radicals.

Flowering plants have evolved two mechanisms, called Strategy I and Strategy II, for scavenging iron from nutrient-poor soils (20). Strategy I plants include all dicots and all monocots except grasses, which are Strategy II plants. Strategy II plants secrete phytosiderophores to scavenge the soil whereas Strategy I plants increase the amount of soluble iron by acidifying the rhizosphere. Both

A) Nicotianamine



B) 2-Deoxymugineic Acid

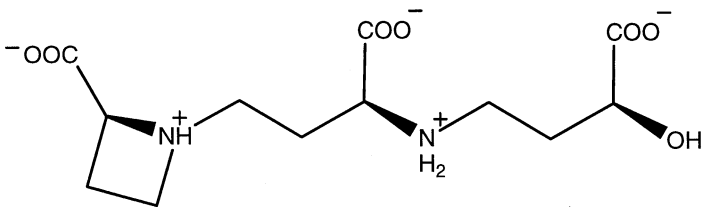


Figure 2. Structures of nicotianamine and deoxymugineic acid.

types appear to use OPTs of the YS clade for long-distance transport of metals, and Strategy II plants use OPTs of the YS clade to import iron-phytosiderophore chelates.

Strategy I iron acquisition has arguably been best studied in *A. thaliana,* where much of the pathway has been elucidated (21). In *A. thaliana, AHA2* encodes an ATPase that is thought to pump protons into the rhizosphere in order to solubilize iron. Because the rhizosphere is an aerobic environment, iron is in an Fe(III) valence state. Fe(III) is bound by ubiquitous chelates in the soil that originate from bacteria and fungi, and these Fe(III)-chelate complexes are reduced by the FRO2 membrane reductase present on the roots. The reduction step releases Fe(II) from the chelate, which is then transported into the plant by a high-affinity Fe(II) transporter (Figure 3A). None of the components involved in bringing iron into the roots of Strategy I plants belong to the OPT family. However, as discussed below, Strategy I plants have a large number of YS-type OPTs that participate in long-distance metal transport.

There has been great interest in elucidating the mechanisms of Strategy II iron acquisition given that graminaceous monocots, such as wheat, corn, rice, and barley, directly or indirectly provide the majority of our calories (22) and that approximately one-third of the earth's soils are calcareous with low levels of soluble iron (19). Phytosiderophores have a high affinity for iron, solubilize metals

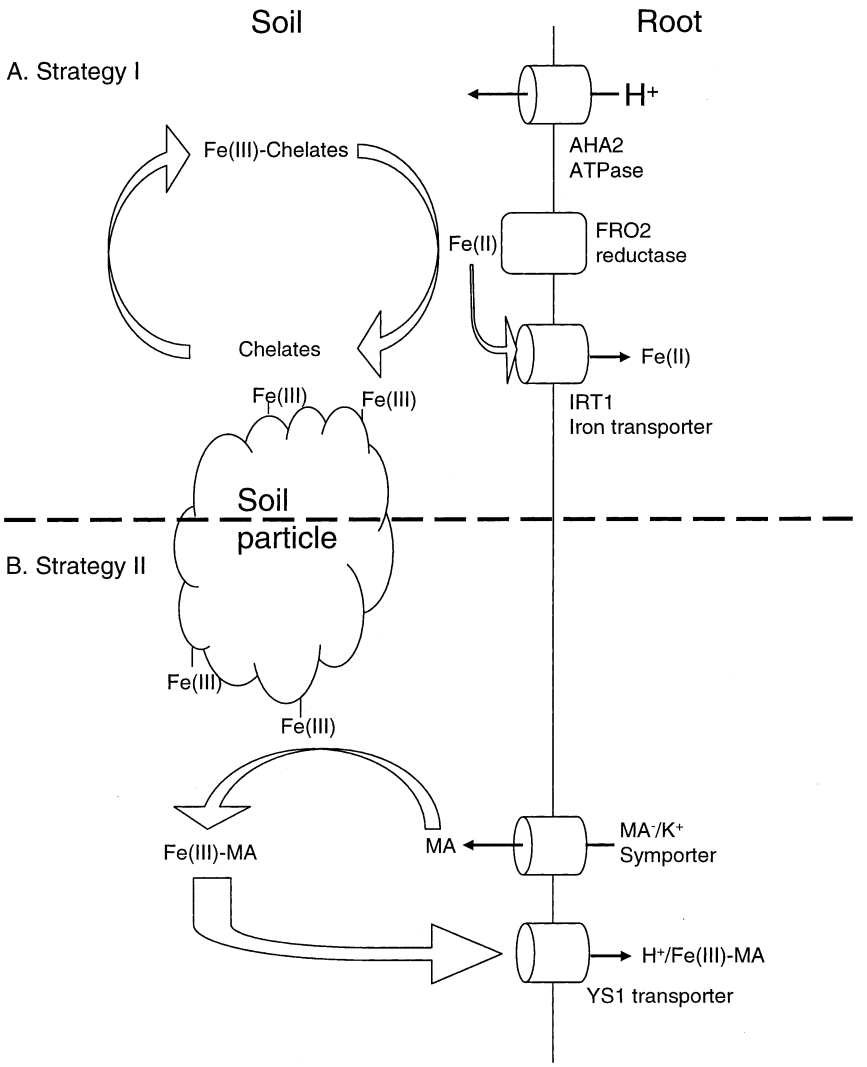


Figure 3. Diagram of iron acquisition by Strategy I (A) and Strategy II (B) plants grown in iron-deficient soil. MA: mugineic acids.

by chelation, and are thought to be secreted into the rhizosphere through a K^+/MA^- symporter (21). Additionally, different grasses secrete different MAs. Once in the rhizosphere, MAs chelate ferric ions by forming Fe(III)-MA complexes that are transported back into the roots through a high-affinity transporter that belongs to the OPT family (5) (Figure 3B).

The only phytosiderophore transporter cloned is Yellow Stripe1 (ZmYS1) from *Zea mays* (5). Given the close relatedness of grasses (23), it is likely that ZmYS1 orthologs will be the route of Fe-phytosiderophore importation by many

Strategy II plants. Several independent lines of evidence have definitively demonstrated that ZmYS1 functions in scavenging the soil for Fe(III)-MA complexes. First, *yellow stripe1* mutants exhibit an iron deficiency phenotype identified by interveinal chlorotic leaves and caused by the failure to import Fe-phytosiderophores (24). Second, ZmYS1 is able to complement an iron transport deficient yeast strain. The *fet3fet4* mutant strain is missing both high-affinity and low-affinity iron transporters and is not able to grow on iron-deficient media. ZmYS1 expressed in this strain confers the ability to grow on a medium containing Fe(III)-deoxymugineic acid (DMA) as the sole source of iron (5), and these growth experiments were substantiated by measuring the uptake of radiolabeled substrates (16). Furthermore, DMA is the only phytosiderophore secreted by maize (25). Third, uptake experiments with oocytes revealed that ZmYS1 has a high affinity for Fe(III)-DMA with an apparent binding constant (Km) of 5 μM. Finally, ZmYS1 mRNA expression is greatly influenced by the presence or absence of iron. For example, ZmYS1 mRNA levels are high in maize roots grown hydroponically in the absence of Fe, whereas expression is barely detectable in plants grown in iron-sufficient conditions (16). Furthermore, the high level of ZmYS1 mRNA expression in iron-deficient conditions is quickly reversed by the addition of soluble iron to the media. The observed response was rapid and the decrease in ZmYS1 mRNA levels was detectable one hour after the addition of iron and reached steady states by hour 18 (16). Interestingly, this same trend was not observed when protein levels were examined. Whereas mRNA levels responded quickly, proteins levels did not begin to change until 18 hours after the addition of iron. This difference in mRNA and protein levels suggests that *ys1* is regulated transcriptionally as well as post-transcriptionally (16). The clear consensus from these experiments is that ZmYs1 functions as an iron-phytosiderophore transporter involved in importing Fe-MAs complexes from the soil.

In addition to importing iron, experimental data from two separate groups argue that ZmYS1 plays two roles in long-distance transport (12, 16). To be distributed to the rest of the plant, iron must traverse the root cortex, be unloaded into the xylem stream, loaded into target organs, and then directed to the proper subcellular location. Furthermore, iron can be redistributed *via* the phloem. During these transport processes, carrier molecules prevent iron from catalyzing the formation of damaging hydroxyl radicals. Roberts et al. (16) found that ZmYS1 transcript and protein are readily detectable in leaves from plants grown in iron-deficient conditions and proposed that ZmYS1 unloads the xylem of Fe(III)-MA chelates acquired from the soil and transported through the xylem. If ZmYS1 were only involved in importing Fe-phytosiderophore complexes from the soil, then one would predict that expression would be restricted to the roots. Because leaves require relatively high amounts of iron for photosynthesis, up-regulation of ZmYS1 in leaves most likely reflects an effort to move iron into the organ of greatest need. Consistent with this model is the observation that ZmYS1 leaf expression coincides with the production of phytosiderophores and that Fe(III)-MA complexes imported from the soil are believed to be loaded into the xylem stream in the root (16, 26). Therefore, one probable role of ZmYS1 in long-distance transport is unloading the xylem sap of Fe(III)-MA chelates acquired during growth in iron-deficient soils.

The second proposed role of ZmYS1 in long-distance transport is the movement of iron-NA complexes. NA is able to bind Fe(III) (27) and Fe(II) (28), and has an approximate concentration of 130 μM in the phloem sap (29) and 20 μM in the xylem (30). Furthermore, NA may be involved in intracellular metal transport (17, 31). Because NA is a precursor for MAs and is structurally very similar (Figure 2), Roberts et al. (16) and Schaaf et al. (12) reasoned that Fe-NA complexes should be substrates for ZmYS1 if this transporter is involved in long-distance transport. Both groups tested this hypothesis by expressing ZmYS1 in a yeast deletion strain missing the iron transporters *fet3* and *fet4*. The *fet3fet4* strain was not able to grow on a medium that contained Fe(II)-NA as sole source of iron whereas transformants harboring ZmYS1 were able to grow, indicating that Fe(II)-NA is a substrate (12, 16). This finding was confirmed by measuring uptake of Fe(II)-NA in yeast (16) and *Xenopus* oocytes (12). Even though both groups concluded that Fe(II)-NA is a substrate, one experimental inconsistency merits comment. Roberts et al. (16) demonstrated that *fet3fet4* [ZmYS1] yeast were able to grow on a medium containing 5 μM Fe(II)-NA, while Schaaf et al. (12) found that *fet3fet4*[ZmYS1] transformants could not grow on the same medium containing 7.5 μM Fe(II)-NA but could grow at a concentration of 40 μM. This discrepancy in results is most easily explained by varying expression levels attributable to differences in yeast vectors. This slight inconsistency between research groups does not detract from the finding that ZmYS1 is able to transport Fe(II)-NA, suggesting a role in long-distance transport.

Fe(III)-NA also appears to be a substrate for ZmYS1, although to a lesser degree as demonstrated by uptake experiments and yeast growth assays (12, 16). Fe(III)-NA was not able to support growth of *fet3fet4*[ZmYS1] yeast strains at concentrations between 5 and 7.5 μM but was able to support growth at a concentration of 40 μM. Similarly, the rate of Fe(II)-NA accumulation by yeast expressing ZmYS1 was approximately 14 times higher than the accumulation of Fe(III)-NA (16), and significant Fe(III)-NA induced currents in oocytes were observed at 100 μM but not at 5 μM concentration of Fe(III)-NA (12). Iron is thought to be transported in the phloem as Fe(II)-NA (27), and, therefore, the finding that ZmYS1 has a higher affinity for Fe(II)-NA suggests this transporter may additionally function in the phloem.

ZmYS1 is not the only OPT implicated in long-distance iron transport. Eight ZmYS1 orthologs, named Yellow Stripe-Like1-8 (AtYSL1-8), have been identified in *A. thaliana*, a Strategy I plant (5). Because all plants use NA as a carrier molecule and Strategy I plants do not produce phytosiderophores, these proteins most likely function in long-distance metal-NA transport (5, 12, 15, 16). This idea is supported by the findings that AtYSL2 is able to transport Fe(II)-NA and Cu-NA and that AtYSL2 is expressed in vascular tissue (15).

One role of AtYSL2 appears to be to move iron-NA complexes out of the xylem (15). Furthermore, at least one other AtYSL has a similar role, inasmuch as loss-of-function mutants do not exhibit any phenotypes suggestive of iron deficiency. AtYSL2 is expressed in the lateral domain of xylem parenchyma and, unlike ZmYS1, AtYSL2 is down-regulated in iron-deficient conditions. The localization of AtYSL2 to the lateral domain of cells adjacent to xylem is significant because importation occurs *in planta* from the apoplasm that includes the

conducting cells of the xylem. Two plausible models have been proposed for how AtYSL2 functions in long-distance iron transport (15). In the first model, AtYSL2 in the xylem parenchyma laterally moves Fe(II)-NA away from conducting cells, where it is then distributed to the surrounding tissue, presumably symplasmically. AtYSL2 levels would decrease when iron is low in order to prevent uptake by mature tissues and thereby distribute the available iron by the xylem stream to younger tissues with higher demands. The second model proposes that AtYSL2 functions in metal cycling by acting as a conduit for transloading Fe(II)-NA from xylem to parenchyma to phloem. AtYSL2 expression would therefore be down-regulated to maintain iron levels in mature tissues by slowing cycling. Consistent with this latter model is the prediction by Takahashi et al. (31) that an Fe-NA specific transporter shuttles iron from xylem to phloem. Because the predominant iron chelate in the xylem is Fe(III)-citrate and not Fe(II)-NA (27), both models postulate a mechanism for citrate to NA exchange (31).

ZmYS1 orthologs also function in long-distance metal transport in Strategy II plants, as demonstrated by the recent characterization of OsYSL2 (17). OsYSL2 is expressed in phloem companion cells and is able to transport Fe(II)-NA and Mn(II)-NA. Furthermore, this transporter is not expressed in root epidermal and cortical cells and is not able to transport Fe(III)-DMA. The localization to the phloem and the substrates immediately suggest a role in long-distance iron and manganese transport and preclude a role in phytosiderophore uptake, especially because rice uses DMA to scavenge the rhizosphere for metals. Interestingly, unlike AtYSL2, OsYSL2 levels increased in plants grown in low iron, possibly indicating a difference in metal distribution processes between Strategy I and II plants. Alternatively, this difference could simply be attributable to varying roles of YSL orthologs.

The primary role of ZmYS1 is to transport iron-chelates but a secondary role may be the movement of other metals. MAs and NA can bind Cu(II), Ni(II), Co(II), Mn(II), and Zn(II) (27, 28), and ZmYS1 expressed in yeast can transport Cu-MA (16), Ni(II)-NA and Ni(II)-DMA (12) at physiologically relevant concentrations. However, two observations indicate that the movement of these substrates is a secondary role for this transporter. Namely, these complexes are poor competitors for the uptake of Fe(II)-chelates and, unlike iron, the lack of these metals in the growth medium does not induce ZmYS1 expression.

Similarly, AtYSL2 and OsYSL2 can transport metals other than iron. Specifically, AtYSL2 can transport Cu-NA and OsYSL2 can transport Mn-NA. AtYSL2 expression is repressed by iron deficiency and also repressed by excess copper. The fact that both metals affect expression suggests that this transporter plays a distinct role in maintaining homeostasis for both metals. Similarly, OsYSL2 expression is increased in the absence of iron although the effect of manganese has not been examined.

The sum of these studies is that YS type OPTs are involved in distributing transition metals. In Strategy II plants this role consists of importing MA-metal complexes from the rhizosphere and the long-distance transport *in planta* of metals complexed to NA or MAs. In Strategy I plants, the likely role is simply long-distance transport of NA-metal complexes. The large number of YS-type

OPTs in species as distant as rice and *Arabidopsis* makes it very probable that all plants have multiple YS-type transporters. This multiplicity most likely reflects a requirement for transporters with different affinities, capacities, expression patterns, and metal preferences in order to maintain metal homeostasis. Additionally, it is tempting to speculate that some OPTs might be involved in intracellular trafficking (31).

Whereas much progress has been made in plants, virtually nothing is known of the role of YS-type OPTs in fungi, archaea, and bacteria, where only a single study has been published, and it did not directly examine transporters. Cho and Zusman (32) found that *Myxococcus xanthus espB* mutants delay sporulation. This organism forms aggregate mounds prior to sporulation, and it was proposed that this OPT is required for the derepression of inhibitors that prevent premature sporulation. Peptide transporters belonging to the ABC superfamily are involved in quorum sensing in some prokaryotes (33) and perhaps EspB helps monitor mound formation through peptide signals.

THE PT CLADE: FUNGI

Six fungal OPTs in the PT clade have been described: CaOpt1p from *C. albicans,* SpIsp4 from *S. pombe,* ScOpt1p/Hgt1p/Gsh11p (34, 35) from *S. cerevisiae*, ScOpt2p from *S. cerevisiae,* YlOpt1 from *Yarrowia lipolytica* (36), and ScoMtd1 from *Schizophyllum commune* (37). Of these, CaOpt1p, SpIsp4, ScOpt1p, and ScOpt2p are known to transport peptides, whereas YlOpt1p and ScoMtd1 are hypothesized to use peptides as substrates (36, 37). Given that the 20 naturally occurring amino acids can form 160,000 and 3,200,000 possible tetra- and pentapeptides, respectively, and that only a small number of substrates have been tested, it is not clear how much substrate overlap exists between the fungal OPTs. However, expression studies and mutant phenotypes, as well as the identification of specific substrates, have shed light on the biological function of individual transporters.

CaOpt1p and SpIsp4 transport tetra- and pentapeptides and, to a lesser degree, tripeptides. Both of these transporters appear to have broad substrate specificity and are regulated by the availability of nitrogen in the growth medium (3, 4). Poor nitrogen sources induce expression whereas rich nitrogen sources repress expression, suggesting that the main function of SpIsp4 and CaOpt1p is scavenging the environment for nitrogenous compounds, namely oligopeptides. Interestingly, *isp4* was originally cloned through a subtractive hybridization screen designed to identify sporulation genes (38). Diploid *S. pombe* cells can be induced to sporulate through nitrogen starvation and, as might be predicted, sexual differentiation genes as well as genes under the control of the nitrogen catabolite repression system were identified in this experiment (38). Later, *isp4* was shown to function as an oligopeptide transporter under the control of the nitrogen catabolite repression system (4). Given that an *isp4* deletion strain exhibited wild-type levels of sporulation and spore viability, it seems unlikely that this gene plays a direct role in sporulation (4). An indirect role such as recycling cell wall components cannot be precluded and such a role has been proposed for the Mtd1 transporter from the split gill fungus *S. commune* (37).

ScOpt1p functions as a high-affinity transporter of the modified tripeptide *glutathione* (γ-Glu-Cys-Gly, GSH) with an apparent Km of 54 μM (34). ScOpt1p is also able to transport tetra- and pentapeptides including leucine enkephalin (Tyr-Gly-Gly-Phe-Leu) and methionine enkephalin (Tyr-Gly-Gly-Phe-Met) but with a much lower affinity (13). For example, ScOpt1p has a Km of 310 μM for leucine enkephalin. Enkephalins are opioids produced by mammals and because yeasts neither produce nor respond to opioids, the importation of these peptides most likely reflects the ability to transport numerous substrates. The high affinity for glutathione and poor competition by peptides for uptake has led to the suggestion that the primary function of this transporter is the importation of glutathione (34). GSH is an abundant tripeptide that is strongly nucleophilic due to the thiol on the cysteine residue. This nucleophilic nature allows this peptide to serve many cellular functions, including control of redox potential, oxidative stress protection, and mediation of environmentally induced stress responses. Furthermore, GSH can serve as a source of organic sulfur (39-41). *ScOPT1* expression is induced in the absence of sulfur and, in particular, when cysteine is depleted, suggesting that the primary function of this transporter is scavenging the medium for sulfur. Perhaps the function of ScOpt1p is simply to transport cysteine-containing peptides, which includes GSH. Interestingly, the role of ScOpt2p has remained elusive as deletion strains have failed to reveal informative phenotypes (4, 34), although ScOpt2p under the control of the *ADH1* promoter was able to transport a radiolabeled tetrapeptide (4).

In addition to scavenging the environment for nitrogenous or sulfur-containing compounds, one fungal OPT may be involved in sensing peptide signals. *Y. lipolytica* is a dimorphic yeast that secretes one of two different extracellular proteases, depending upon the pH of the medium (42). *Ylopt1* mutants failed to secrete the acidic protease Axp under inducing conditions as did mutants in the *Ylssy5* gene, which senses extracellular amino acids. Both mutant classes also did not form hyphae in an inducing serum medium. These phenotypes suggest that sensing extracellular peptides is requisite for Axp production and hyphael formation. There are two conceivable ways in which YlOpt1p could play a regulatory role. YlOpt1p could import a peptide(s) signal that induces the hyphael and protease pathway. This induction could result either directly through binding an intracellular receptor or indirectly by affecting the intracellular pool of amino acids. Alternatively, YlOpt1p could act as a sensor/receptor that binds peptides. There are examples of where a member of a transporter family has taken on the function of receptor/sensor. For example, Snf3p and Rgt2p are both 12 transmembrane domain proteins that function as glucose sensors in yeast and both of these proteins are members of the glucose transporter family (GLUT) (43). It is, therefore, conceivable that YlOpt1p functions as a peptide receptor/sensor. Both of these working models assume that YlOpt1p transports or binds peptides, and since this has not been experimentally demonstrated, it is plausible that the substrate(s) are different in nature.

THE PT CLADE: PLANTS

Plant members of the PT clade have been characterized from Indian mustard (*Brassica juncea*) (11), rice (44), and *A. thaliana* (11, 45, 46). These studies

have revealed that members of this clade transport small peptides, have different substrate preferences and expression patterns, and participate in a number of different functions including embryo development (47), GSH transport (11, 44), and heavy metal detoxification (11, 46). Furthermore, researchers have speculated that peptide transporters participate in seed germination (48) and long-distance nitrogen allocation (49). Less clear is the role that these proteins may play in translocating peptide hormones and signal molecules.

Complimentary genetic, molecular, and biochemical studies of AtOPT1-9 from *A. thaliana* have begun to unravel their biological roles in plants. These experiments have revealed similarities and differences in substrates as well as expression patterns. Koh et al. (45) tested whether peptides that are known substrates for fungal OPTs could be transported by AtOPTs. This approach was quite reasonable given the large number of possible oligopeptides of three to five residues and was successful in identifying substrates for AtOPT1, 4, 5, 6, and 7. Specifically, AtOPT4 was found to transport *lysyl-leucyl-leucyl-glycine* (KLLG), *lysyl-leucyl-glycyl-leucine* (KLGL), and *lysyl-leucyl-leucyl-leucyl-glycine* (KLLLG), whereas AtOPT1, 5, and 7 were only able to transport the pentapeptide KLLLG (45). None of these transporters was able to transport leucine containing di- or tripeptides, and substrates larger than five residues were not tested. It is not known whether these peptides are common in plants or whether their uptake by yeast simply reflects broad substrate specificity.

The limited number of PTs and peptides examined makes it difficult to generalize about substrates, and the failure of an OPT to transport a peptide does not preclude it from functioning as a peptide transporter. For example, AtOPT2 failed to transport the oligopeptides KLLG, KLGL, KLLLG, and leucine enkephalin (45), but the large number of natural tetra- and pentapeptides makes it still possible that this protein functions as an oligopeptide transporter. How does one systematically identify substrates when there are so many possibilities? My laboratory, in collaboration with Jeff Becker and Fred Naider, is testing the hypothesis that OPTs that are expressed in developing seeds or during germination in rice will transport peptides that are rich in amino acids that comprise seed storage proteins. This is a particularly compelling idea when one considers the amount of amino acids that must be imported to synthesize storage proteins, and that importation must precede protein synthesis. Similarly, one might predict that PTs that are expressed in vascular tissue throughout the plant transport peptides rich in glutamine or glutamate, since these amino acids are used to distribute nitrogen (49).

Serge Delrot's laboratory has recently published a study showing that AtOPT6 can transport a wide range of peptides, suggesting that some OPTs have broad substrate specificity (46). AtOPT6 permitted an *opt1* yeast mutant to grow on GSH as a sole sulfur source but did not rescue an *opt1met15* double mutant, indicating a stringency threshold. Furthermore, methionine, glutamine, *leucyl-leucine* (LL), KLLG, and KLLLG all competed for the uptake of [^{3}H]GSH. Both of these results were unexpected, simply because a previous study did not detect the uptake of [^{3}H]GSH and yeast transformed with AtOPT6 failed to grow on LL as a sole source of leucine (45). These differences illustrate several important points. First, the two groups used different expression vectors, implying that the

ability to measure transport in a heterologous system is influenced by vector choice. Second, growth assays, uptake assays, and competition experiments do not necessarily measure transport in the same manner.

In addition to AtOPT6, two other plant OPTs have been found to transport GSH: OsGT1 from rice (44) and BjGT1 from *B. juncea* (11). OsGT1, like AtOPT6, exhibits broad substrate specificity, as demonstrated by competition assays with [^{3}H]GSH (44). Interestingly, all three of these transporters do not follow simple Michaelis-Menten kinetics, but instead exhibit diphasic affinities for GSH. For example, AtOPT6 has a high-affinity phase of 400 μM and a low-affinity phase of 5 mM (46). This same trend, that is a high-affinity phase in the micromolar range and a low-affinity phase in the millimolar range, was also observed for OsGT1 and BjGT1. Zhang et al. (44) have postulated that these transporters may have two binding sites, one for peptides that bind GSH with low affinity and a second, high-affinity site for GSH. This would explain why these proteins are able to transport a broad range of peptides and maintain a high affinity for GSH. These are not the first transporters found to exhibit diphasic kinetics, but it is interesting to note that the fungal GSH transporter ScOpt1p follows simple saturation kinetics.

Serge Delrot's group has proposed an interesting model for the role of BjGT1 in mediating heavy-metal toxicity. Plants use glutathione as a precursor to produce a class of metal-binding compounds called phytochelatins (PCn: (γ-Glu-Cys)n-Gly where n = 2–11] (50). These compounds bind heavy metals, and the resulting complex is transported to the central vacuole. This detoxification process requires the coordination of GSH production and transport, as well as phytochelatin synthesis. Bogs et al. (11) noted that high levels of cadmium negatively affect BjGT1 leaf expression and proposed two plausible hypotheses to explain this observation. The first postulates that BjGT1 imports GSH into the phloem and BjGT1 expression is down-regulated by cadmium to prevent GSH export to sink tissues so that phytochelatins can be synthesized in the leaf. The other possibility is that BjGT1 expression is decreased in the leaf mesophyll to prevent the uptake of Cd-GSH complexes because GSH itself can bind cadmium. Regardless of which model proves true, the regulation by cadmium suggests that BjGT1 plays a role in heavy metal detoxification. Interestingly, AtOPT6 was also able to transport Cd-GSH complexes, supporting the idea that OPTs that transport GSH are involved in heavy metal stress responses. Furthermore, AtOPT6 expression is up-regulated after herbicide treatment but not after exposure to hydrogen peroxide, indicating that stress-induced *AtOPT6* expression is a specific and not general response (46).

AtOPT3 has also been implicated in metal transport and appears to play a critical role in seed development. Gary Stacey's group demonstrated that this transporter is expressed in vascular tissue throughout the plant and is temporally and spatially regulated during embryogenesis (47). *AtOPT3* is expressed in the vascular tissue of the funicle pre-fertilization but not in the corresponding ovule. Following fertilization, expression is induced to high levels in the embryo, suspensor, endosperm, and integuments until the heart stage of development, at which time expression is only present in the embryo. This expression pattern

coincides with the appearance of ingrowths in the embryo sac, a hallmark of increasing nutrient uptake by the filial tissues. Furthermore, homozygous plants for an insertion allele arrested embryonic development. Close examination revealed that the embryos were not able to undergo periclinal cell division and hence never formed a protoderm or hypophyseal cell, the progenitor of the root quiescent center. The finding that AtOPT3 can complement yeast copper and manganese transport mutants (14) suggests that this transporter provides metals to the embryo during seed development. Whereas the expression studies and genetic analysis are quite convincing, it does seem peculiar that AtOPT3 can transport free metals when all other OPTs transport either peptides or amino acid-metal chelates. Cagnac et al. (46) have cautioned that constituents in yeast media such as histidine may be able to bind metals and apparent transport of free metals may in reality be uptake of metal conjugates formed in the media. It will be interesting to see whether these substrates can be confirmed through competition experiments, *in planta,* or in another heterologous system such as *Xenopus*. Regardless, it is clear that *AtOPT3* is expressed in vascular tissue and that expression is influenced by metals, suggesting a role in long-distance metal transport.

The trend emerging from these studies is that OPTs in the PT clade function in long-distance transport and their primary substrates are most likely peptides, including GSH. In addition to the studies mentioned above, Gary Stacey's laboratory has found that all nine AtOPTs are expressed in vascular tissue and cotyledons, and that none of these transporters is expressed in the root tip or root hairs (G. Stacey, personal communication). Together, these studies suggest that these OPTs load or unload vascular tissue for long-distance transport and that these transporters are not involved in scavenging the soil. What is also clear is that these transporters use different substrates, which presumably reflects different biological roles. For example, the glutathione transporters may well be involved in heavy-metal detoxification and long-distance sulfur transport, whereas other OPTs may be involved in long-distance nitrogen transport. Large shifts in nitrogen partitioning occur during certain periods of a plant's life, for example, during seed formation, senescence, and germination, and OPT transporters provide a potentially energetically efficient and rapid means of moving amino acids in bulk.

One unexplored role of OPTs in plants is the translocation of peptide signal molecules. Peptide-based hormones and signal molecules are common in animals and five peptide hormones/signal molecules have been discovered in plants (51-53). They are 1) ENOD40, a 10-12-residue hormone that affects nodule formation in legumes; 2) systemin, an 18-residue hormone produced in response to wounding by herbivores; 3) RALF, a 50-residue signal molecule; 4) S-locus cysteine-rich protein, a 74-77-residue signal protein involved in pollen self-incompatibility; and phytosulfokine (PSK), a five-residue mitogen involved in dedifferentiation and cell cycle control (53, 54). Of all of these peptides, phytosulfokine is the best potential substrate for an OPT based on its size. Additionally, PSK is produced in two forms: PSK-α and PSK-β. PSK-α is a pentapeptide and the active form of the mitogen whereas PSK-β is a tetrapeptide formed through cleavage of the C-terminal residue of PSK-α

and is much less active (55). PSK-α binds to a membrane-bound receptor kinase and conceivably an OPT could act as an antagonist by importing PSK-α before it binds. Interestingly, OsOPT7 expression is cell cycle dependent, with highest expression occurring during G1, which corresponds with the mitogen activation period (56). Alternatively, an OPT could function in recycling the less active PSK-β form.

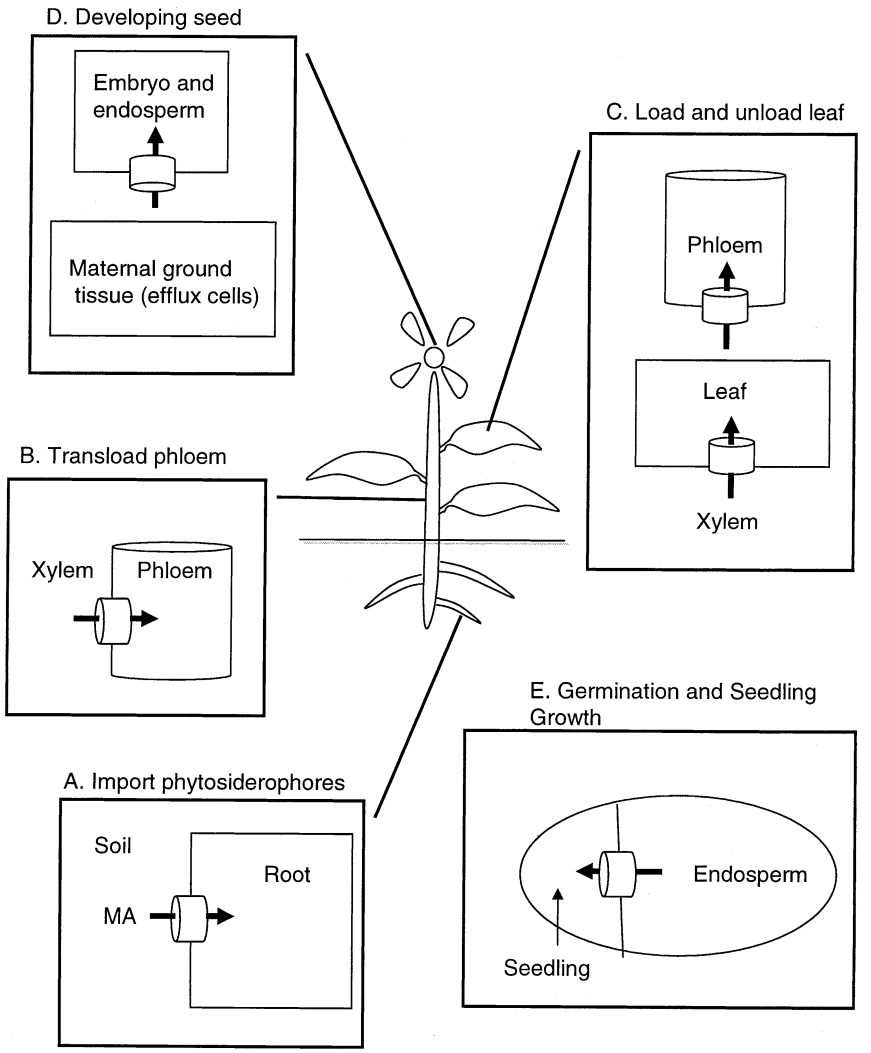


Figure 4. Working model for OPT function in plants. OPTs import metal-phytosiderophores (MA) in Strategy II plants (A), transload the phloem (B), and unload the xylem into the leaf (C). OPTs may also function in amino acid and metal-loading in the developing seed (D) and unloading during germination and seedling growth (E). Arrows indicate direction of peptide and metal movement.

SUMMARY

The OPT family was first described six years ago, and much progress has been made in understanding the role these transporters play in their respective organisms. Plants are the only organisms in which both YS- and PT-type transporters have been characterized, and all of these OPTs appear to be plasma membrane–bound proteins, suggesting that they import substrates from the apoplasm or the environment. YS1 is the only OPT known to translocate substrates from the rhizosphere, whereas all the other OPTs seem to function in long-distance transport of peptides or metals. The sum of all the studies covered in this review suggest the model for OPT function in plants depicted in Figure 4. Peptides, metal-NA, and metal-MAs complexes (Strategy II plants only) are loaded into the xylem stream in the root for long-distance transport. OPTs unload the xylem by importing substrates into sink tissues such as leaves and by transloading the phloem. Peptides and metal-NA complexes exit the leaf symplasmically or by importation into the phloem from the apoplasm by OPTs. The filial tissues (endosperm and embryo) are apoplasmically separated from the maternal tissues, and OPTs may also function in loading the developing seed. Similarly, seedlings are symplasmically disconnected from the endosperm and OPTs may help move nutrients to the growing plant.

Much progress has been made in the last two years toward understanding OPTs in plants, although several fundamental questions remain unanswered. Namely, what is the level of redundancy? Is there any substrate overlap between YS and PT OPTs? How crucial are their respective roles? Are there additional functions beyond peptide and metal transport? Given the recent pace of discovery, we may not have to wait long to find out the answers.

ACKNOWLEDGMENTS

The author thanks Elsbeth Walker and David Barnes for critical reading of the manuscript. This publication was made possible by the Vermont Genetics Network through NIH Grant Number 1 P20 RR16462 from the BRIN Program of the National Center for Research Resources.

REFERENCES

1 Busch, W. and Saier, M.H., Jr. (2002) Crit. Rev. Biochem. Mol. Biol. 37, 287-337.
2 Stacey, G., Koh, S., Granger, C. and Becker, J.M. (2002) Trends Plant Sci. 7, 257-263.
3 Lubkowitz, M.A., Hauser, L., Breslav, M., Naider, F. and Becker, J.M. (1997) Microbiology 143, 387-396.
4 Lubkowitz, M.A., Barnes, D., Breslav, M., Burchfield, A., Naider, F. and Becker, J.M. (1998) Mol. Microbiol. 28, 729-741.
5 Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S.L., Briat, J.-F. and Walker, E.L. (2001) Nature 409, 346-349.
6 Hauser, M., Narita, V., Donhardt, A. M., Naider, F. and Becker, J.M. (2001) Mol. Membr. Biol. 18, 105-112.

7 European Bioinformatics Institute. Available at: http://www.ebi.ac.uk/interpro/ (Entry IPR004813).
8 The Institute for Genome Research. Available at: http://www.tigr.org/TIGRFAMs/index.shtml (Entry TIGR00728).
9 Protein Families Database of Alignments and HMMs. Available at: http://www.sanger.ac.uk/Software/Pfam/ (Entry PF03169).
10 Yen, M.R., Tseng, Y.H. and Saier, M.H., Jr. (2001) Microbiology 147, 2881-2883.
11 Bogs, J., Bourbouloux, A., Cagnac, O., Wachter, A., Rausch, T. and Delrot, S. (2003) Plant Cell Environ. 26, 1703-1711.
12 Schaaf, G., Ludewig, U., Erenoglu, B.E., Mori, S., Kitahara, T. and von Wirén, N. (2004) J. Biol. Chem. 279, 9091-9096.
13 Hauser, M., Donhardt, A.M., Barnes, D., Naider, F. and Becker, J. M. (2000) J. Biol. Chem. 275, 3037-3041.
14 Wintz, H., Fox, T., Wu, Y.-Y., Feng, V., Chen, W., Chang, H.-S., Zhu, T. and Vulpe, C. (2003) J. Biol. Chem. 278, 47644-47653.
15 DiDonato, R.J., Jr., Roberts, L.A., Sanderson, T., Eisley, R.B. and Walker, E.L. (2004) Plant J. 39, 403-414.
16 Roberts, L.A., Pierson, A.J., Panaviene, Z. and Walker, E.L. (2004) Plant Physiol. 135, 112-120.
17 Koike, S., Inoue, H., Mizuno, D., Takahashi, M., Nakanishi, H., Mori, S. and Nishizawa, N.K. (2004) Plant J. 39, 415-424.
18 Tagaki, S., Nomoto, K. and Takemoto, T. (1984) J. Plant. Nutr. 7, 469-477.
19 Guerinot, M.L. and Yi, Y. (1994) Plant Physiol. 104, 815-820.
20 Marschner, H. and Rohmeld, V. (1994) Plant Soil 165, 375-388.
21 Curie, C. and Briat, J.-F. (2003) Ann. Rev. Plant Physiol. Plant Mol. Biol. 54, 183-206.
22 Shewry, P.R. and Halford, N.G. (2002) J. Exp. Bot. 53, 947-958.
23 Bennetzen, J.L. and Freeling, M. (1997) Genome Res. 7, 301-306.
24 von Wirén, N., Mori, S., Marschner, H. and Romheld, V. (1994) Plant Physiol. 106, 71-77.
25 Mori, S. and Nishizawa, N. (1987) Plant Cell Physiol. 28, 1081-1092.
26 Alam, S., Kamei, S. and Kawai, S. (2001) Plant Soil 47, 643-649.
27 von Wirén, N., Klair, S., Bansal, S., Briat, J.-F., Khodr, H., Shioiri, T., Leigh, R.A. and Hider, R.C. (1999) Plant Physiol. 119, 1107-1114.
28 Anderegg, G. and Ripperger, H. (1989) J. Chem. Soc. Chem. Commun. 10, 647-650.
29 Schmidke, I. and Stephan, U. (1995) Physiol. Plant 95, 147-153.
30 Pich, A. and Scholz, G. (1996) J. Exp. Bot. 47, 41-47.
31 Takahashi, M., Terada, Y., Nakai, I., Nakanishi, H., Yoshimura, E., Mori, S. and Nishizawa, N.K. (2003) Plant Cell 15, 1263-1280.
32 Cho, K. and Zusman, D. R. (1999) Mol. Microbiol. 34, 714-725.
33 Solomon, J.M., Lazazzera, B.A. and Grossman, A. D. (1996) Genes Dev. 10, 2014-2024.
34 Bourbouloux, A., Shahi, P., Chakladar, A., Delrot, S. and Bachhawat, A.K. (2000) J. Biol. Chem. 275, 13259-13265.

35 Miyake, T., Hiraishi, H., Sammoto, H. and Ono, B.-I. (2003) J. Biol. Chem. 278, 39632-39636.
36 Gonzalez-Lopez, C.I., Szabo, R., Blanchin-Roland, S. and Gaillardin, C. (2002) Genetics 160, 417-427.
37 Lengeler, K.B. and Kothe, E. (1999) Curr. Genet. 36, 159-164.
38 Sato, S., Suzuki, H., Widyastuti, U., Hotta, Y. and Tabata, S. (1994) Curr. Genet. 26, 31-37.
39 Meister, A. and Anderson, M. E. (1983) Ann. Rev. Biochem. 52, 711-760.
40 May, M., Vernoux, T., Leaver, T., Van Montagu, M. and Inze, D. (1998) Exp. Bot. 49, 649-667.
41 Leustek, T. and Saito, K. (1999) Plant Physiol. 120, 637-644.
42 Ogrydziak, D.M. and Mortimer, R.K. (1977) Genetics 87, 621-632.
43 Forsberg, H. and Ljungdahl, P.O. (2001) Curr. Genet. 40, 91-109.
44 Zhang, M.Y., Bourbouloux, A., Cagnac, O., Srikanth, C.V., Rentsch, D., Bachhawat, A.K. and Delrot, S. (2004) Plant Physiol. 134, 482-491.
45 Koh, S., Wiles, A.M., Sharp, J.S., Naider, F.R., Becker, J.M. and Stacey, G. (2002) Plant Physiol. 128, 21-29.
46 Cagnac, O., Bourbouloux, A., Chakrabarty, D., Zhang, M. Y. and Delrot, S. (2004) Plant Physiol. 135, 1378-1387.
47 Stacey, M.G., Koh, S., Becker, J. and Stacey, G. (2002) Plant Cell 14, 2799-2811.
48 Muntz, K. (1998) Plant Mol. Biol. 38, 77-99.
49 Williams, L. and Miller, A. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 659-688.
50 Cobbett, C. and Goldsbrough, P. (2002) Ann. Rev. Plant Biol. 53, 159-182.
51 Lindsey, K., Casson, S. and Chilley, P. (2002) Trends Plant Sci. 7, 78-83.
52 Franssen, H.J. and Bisseling, T. (2001) Proc. Nat. Acad. Sci. U.S.A. 98, 12855-12856.
53 Ryan, C.A. and Pearce, G. (2001) Plant Physiol. 125, 65-68.
54 Matsubayashi, Y., Ogawa, M., Morita, A. and Sakagami, Y. (2002) Science 296, 1470-1472.
55 Yang, H., Matsubayashi, Y., Nakamura, K. and Sakagami, Y. (1999) Proc. Nat. Acad. Sci. U.S.A. 96, 13560-13565.
56 Lubkowitz, M. (2004) Unpublished observation. Rice Expression Database (RED, available at: http://red.dna.affrc.go.jp/RED/)

PHOSPHOLIPID-DERIVED SIGNALING IN PLANT RESPONSE TO TEMPERATURE AND WATER STRESSES

Xuemin Wang

Department of Biology
University of Missouri
St. Louis, Missouri 63121
Donald Danforth Plant Science Center
St. Louis, Missouri 63132

INTRODUCTION

Cell membranes are typically the initial point of signal perception and transduction. Phospholipids provide not only the structural base of cell membranes, but are also rich resources for generating cellular mediators. Such mediators include phosphatidic acid (PA), lysophospholipids, free fatty acids (FFAs), oxylipins, *N*-acylethanolamine, diacylglycerol (DAG), phosphoinositides, and water-soluble inositol polyphosphates. Each category of the above lipid regulators encompasses many molecular species due to variations in fatty acids, head groups, and/or other modifications. These lipid and lipid-derived regulators are produced in response to biotic and abiotic cues, and the production is regulated by various lipid-signaling enzymes, such as different families of phospholipases, lipid kinases, and/or phosphatases (Table 1) (1).

Lipid-mediated signaling is an emerging, rapidly expanding area of research in plant biology. Significant advances have been made over the past years on the production and function of lipid-derived signaling processes.

Genetic Engineering, Volume 27, Edited by J. K. Setlow

Table 1. Lipid signaling enzymes and lipid-derived messengers.

Lipid signaling enzyme	Short name	Substrate	Products	Animals	Plants
Phospholipases					
Phospholipase D	PLD	PC, PE, PG, PS	PA + head group	PLD1, 2	α(3), β(2), γ(3), δ, ε, ζ(2)
Phospholipase C	PI-PLC	$PI(4,5)P_2$	DAG + IP_3	β(4), γ(2), δ(4), ε, ζ	ζ-like (7-9)
PC-Phospholipase C	PC-PLC	PC, PE, etc	DAG + choline-P_i	not cloned	putatively multiple
Phospholipase A_2	PLA_2	PC, PE, etc	2-lysoPC + fatty acid	$sPLA_2$, $cPLA_2$, $iPLA_2$	$sPLA_2$, $iPLA_2$
PA-Phospholipase A_1	PA-PLA_1	PA	1-lysoPA + fatty acid	n/d	SRG2
Lysophospholipase D	LPLD	lysoPC sphingosine-1-P-choline	lysoPA+ choline sphingosine-1-P_i + choline	autotaxin (ATX)	n/d
Sphingomyelinase	SMase	sphingomyelin	ceramide + choline-P_i	acidic, neutral SMases	n/d
Acyl hydrolase		PLs, galactolipids, DAG	FA, lysolipids	n/d	patatin-like, DAD1-like
Lipid kinases					
PI kinases	PIK	$PI(4)P_2$ $PI(3)P_2$ PI	$PI(4,5)P_2$ $PI(3,4)P_2$/$PI(3,5)P_2$ PtdIns(5)P	Type I Type II Type III	Type I Type II Type III
PI 4-kinases	PI-4K	PI	PtdIns(4)P	Type II, III(α, β)	Type II, III(α, β)
PI 3-kinases	PI-3K	PI(4)P/$PI(4,5)P_2$	$PI(3,4)P_2$,/$PI(3,4,5)P_3$	Class I, II	III (yeast ScVps34p-like)
DAG kinases	DAGK	DAG	PA	Type I-V	multiple
Sphingosine kinases	SPhK	sphingosine	sphingosine-1-P_i	SPhK1, 2	AtLCBK1
PA kinase	PAK	PA	DAG-PP_i	n/d	not cloned
Lipid phosphatases (pptase)					
Lipid phosphate pptase	LPP	PA, lysoPA, DAGPP sphingosine-1-Pi	DAG, MAG sphingosine	LPP1-3	2
$PI(3,4,5)P_3$ 5-pptase	SHIP	$PI(3,4,5)P_3$	$PI(3,4)P_2$	SHIP1-3	n/d
$PI(3,4,5)P_3$/		$PI(3,4,5)P_3$	$PI(4,5)P_2$		
$PI(3,4)P_2$-3-pptase	PTEN	$PI(3,4)P_2$	PI(4)P	PTEN	AtPTEN1-3

Abbreviations: AtLCBK1, Arabidopsis sphingolipid long-chain base kinase1; DAG, diacylglycerol; IP_3, inositol (1,4,5)-trisphosphate; MAG, monoacylglycerol; PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; Pi, inorganic phosphate; PL, phospholipid; PLD, phospholipase D; PLC, phospholipase C; PLA, phospholipase A; $PI(4,5)P_2$, phosphatidylinositol 4,5-bisphosphate; PS, phosphatidylserine. (Adapted from Wang, X., Ref. 1.)

Increasing evidence indicates that lipid mediators and associated enzymes are involved in regulating plant growth, development, and response to environmental challenges. They mediate various plant processes through roles in signal transduction, cytoskeletal rearrangements, membrane trafficking, and/or membrane metabolism. Interested readers are referred to recent reviews that have dealt with signaling aspects of various lipids and enzymes in plants (1-7). Lately, exciting progress has been made toward the identification of downstream targets of lipid messengers and understanding the mechanism by which specific lipid messengers and associated signaling enzymes in signaling cascades (8, 9). This chapter focuses on the role of phospholipid-mediated signaling in plant response to temperature and water stresses.

LOW-TEMPERATURE STRESS

Freezing injury is a major agricultural problem and results annually in substantial economic losses. In recent years, great progress has been made toward identifying genes involved in cold acclimation and freezing tolerance. Identification of the DREB family (also referred to as CBF) has revealed a set of transcriptional activators, together with their upstream and downstream genes, which play important roles in regulating plant response to low temperatures (10-12). DREB-independent processes that mediate plant tolerance to low temperature stress have also been identified (13, 14). Multiple pathways exist for plants to acquire freezing tolerance.

Studies in Arabidopsis show that members of the phospholipase D (PLD) family play important and different roles in plant freezing tolerance (15, 16). Phospholipases are phospholipid-hydrolyzing enzymes that are grouped into four major classes, PLD, PLC, PLA_2, and PLA_1, according to the site of hydrolysis (Figure 1). PLD cleaves the terminal phosphodiesteric bond of common membrane lipids to generate PA and a free head group. The plant PLD family consists of multiple members; Arabidopsis has 12 PLDs that are classified into six types, PLDα(3), β(2), γ(3), δ, ε, and ζ (2). The various types of PLDs display distinguishable catalytic and regulatory properties (17, 18), and the heterogeneity of PLDs plays an important role in their diverse functions in the cell. Unique cellular functions have been indicated for several PLDs, which include cell patterning, programmed cell death, abscisic acid response, and freezing injury and tolerance (1, 18, and references therein).

Genetic suppression of the most prevalent plant PLD, PLDα1, rendered *Arabidopisis* plants more tolerant to freezing (15). PLDα1-deficient plants had a higher level of phosphatidylcholine (PC) and a lower level of PA than that of wild-type plants after freezing. Comparative profiling of lipid molecular species revealed that PLDα1 used primarily PC as the substrate. Freezing induces a substantial hydrolysis of common membrane lipids, PC, phosphatidylethanolamine (PE), and phosphatidylglycerol (PG), and PLDα1 is responsible for more than 50% of PC hydrolyzed (15). PC is a bilayer-stabilizing lipid, whereas PA has a tendency to form hexagonal II phase in the presence of calcium. The propensity of cellular membranes to form the hexagonal phase has been suggested to be a key event in freezing injury. It has been proposed that suppression of PLDα1 may

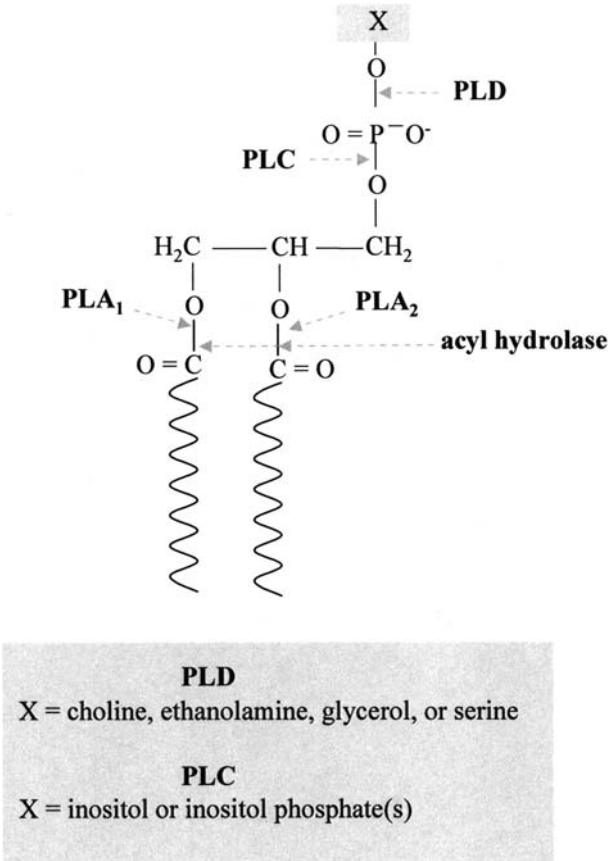


Figure 1. Type of phospholipid-hydrolyzing enzymes and the sites of hydrolysis by PLD, PLC, PLA_2, PLA_1, and acyl hydrolase. An acyl hydrolase can remove acyl groups from both sn-1 and sn-2 positions. X denotes head group, and the shaded box shows the common head groups released from the reactions of cloned PLDs and PLCs. Products from the respective enzyme activities are given in Table 1. The two waving lines denote hydrocarbon tails of fatty acids. The esterifed fatty acids may vary in the numbers of carbons and double bonds and the position of double bond, thus giving rise to different molecular species within each head-group class of phospholipids.

decrease freezing-induced production of PA and the propensity of membrane lipids to undergo a transition from lamellar to hexagonal II phase, thus increasing freezing tolerance (15). This study indicates that high PLDα1 activity has a detrimental effect on plants under severe freezing conditions.

By comparison, genetic knockout of the plasma membrane-associated PLDδ rendered Arabidopsis plants more sensitive to freezing, whereas overexpression of PLDδ increased freezing tolerance (16). Lipid analysis indicates that, unlike PLDα1, PLDδ activity does not result in substantial lipid hydrolysis, but it does produce selective PA species (16). These results indicate that PLDδ positively affects the cellular response to freezing. In addition, they demonstrate that manipulation of different PLDs can have different phenotypic alterations.

The differences in phenotypic and metabolic changes between PLDδ and PLDα1 also indicate that the two PLDs are involved in plant response to freezing via different mechanisms. Insights for the different function are gained from biochemical and molecular characterization of the two PLDs. PLDδ and PLDα1 differ from one another in terms of the requirements for Ca^{2+} and FFAs, substrate selectivity, intracellular location, and patterns of gene expression (18). For instance, PLDδ prefers PE to PC as substrate (19), whereas PLDα1 prefers PC to other lipids (15, 20). In addition to the different amounts of PA, the location and timing of PA produced by PLDα1 and PLDδ may play an important role in PA's cellular function. PLDδ is associated with the plasma membrane, whereas PLDα1 is with both the plasma and intracellular membranes (21). Another unique property of PLDδis that its activity is stimulated by the monounsaturated fatty acid, oleic acid (21). Oleate has been implicated in a decrease in cell death in both plants and animals. *PLDδ* has been shown to have an anti-death function (22). With *PLDδ*-knockout, Arabidopsis cells had increased sensitivity to H_2O_2 and PLDδ derived PA-mitigated H_2O_2-promoted cell death, whereas overexpresson of *PLDδ* increased the cell's tolerance of oxidative stress (22). The level of H_2O_2 increases in plant cells in response to various stress conditions, including freezing. Thus, the impaired response to oxidative stress in *PLDδ*-null plants may be a basis for the decreased freezing tolerance. It has been proposed that *PLDδ* positively mediates plant freezing tolerance, possibly through its role in signaling activation of MAP (mitogen-activated protein) kinases and the cell's resistance to damages promoted by reactive oxygen species (16, 22).

The *PLDδ* alterations resulted neither in any apparent change in the expression of the cold-regulated genes *COR47* or *COR78* nor in any change in cold-induced increases in proline or soluble sugars (16). The lack of changes in these molecules suggests that PLDs and associated membrane lipid hydrolysis are not directly involved in the DREB signaling pathways. The alterations in the freezing tolerance of the PLDδ-altered plants were manifested only after cold acclimation, suggesting that cold acclimation is required for *PLDδ* function during freezing. This requirement could result from cold regulation of *PLDδ* expression, other cellular factors, or a combination of both, brought about during the acclimation process. Analysis of PLDδ transcript levels indicates that increased expression of *PLDδ* alone is not sufficient to explain the requirement for cold acclimation in *PLDδ* action in freezing tolerance (16). It is likely that *PLDδ* enhances freezing tolerance in coordination with other cold-induced processes, including expression of cold-regulated genes, osmolytes, and lipid composition.

STOMATAL MOVEMENT AND WATER LOSS

Terrestrial plants lose water primarily by way of stomata, which are pores defined by pairs of guard cells on leaf surfaces. Stomata close during drought stress, and this change is crucial to maintaining hydration status in leaves and to plant survival. A number of cellular components have been implicated in regulating stomatal movement (23-26). Membrane lipid-derived mediators are part of the signaling cascades, and they include PA, *N*-acylethanolamine, inositol 1,4,5-trisphosphates (IP_3), phosphatidylinositol 3-phosphate (PI-3P) and phos-

phatidylinositol-4 phosphate (PI-4P), FFAs, and methyl jasmonate. These mediators are produced by several distinctive lipid-signaling pathways, including PLD, PLC, PLA, and phosphatidylinositol kinases (Table 1, Figure 1).

The hormone abscisic acid (ABA) plays an important role in regulating stomatal movement and plant water loss (23, 25, 26). ABA inhibits stomatal opening and promotes stomatal closure, and its level increases under water-deficit conditions. PLD and its derived PA have been implicated in plant response to ABA from studies using different systems. Antisense suppression of PLDα1 delayed ABA-promoted leaf senescence in Arabidopsis (27). In barley aleurone, PLD and its lipid product PA were involved in the ABA antagonism of gibberellin-induced events (28). Application of PA to *Vicia faba* guard cell protoplasts promoted ionic efflux and stomatal closure (29). In rice protoplasts, decreases in PA formation by *n*-butanol inhibited expression of several ABA-inducible genes that were regulated by ABI1-5, a basic leucine zipper factor (30). *N*-acylethanolamines, which inhibited the activity of PLDα1, but not β1 or γ1, retarded ABA-promoted stomatal closure in tobacco and *Commelina communis* (31).

Studies involving genetic manipulations of PLDs have provided molecular and physiological evidence for the role of a specific PLD in the ABA response and plant water loss (8, 32). Antisense suppression of PLDα1 decreased the plant's sensitivity to ABA and impaired stomatal closure. PLDα1-depleted plants exhibited an accelerated rate of transpirational water loss and a decreased ability to tolerate drought stress (32). In addition, overexpression of PLDα1 increases the leaf's sensitivity to ABA in promoting stomatal closure and decreases the rate of transpirational water loss. The role of PLDα1 in stomatal closure and transpirational water loss is further supported by a recent study using an Arabidopsis PLDα1 gene knockout (8). Stomatal aperture in the PLDα1-knockout plants was much less responsive to ABA than that in wild type. However, application of PA, the product of PLD, promoted stomatal closure in the PLDα1-knockout leaves (8).

A mechanism by which PLDα1 mediates ABA signaling has been revealed in a recent study (8); PLDα1-derived PA bound to ABI1, a protein phosphatase 2C (PP2C) that is a negative regulator of ABA responses in Arabidopsis. The PA binding decreased PP2C activity and also appeared to tether ABI1 to the plasma membrane, thus reducing the translocation of ABI1 into the nucleus in response to ABA (8). The translocation of ABI1 from cytosol to the nucleus is important to the phosphatase function (33). These results indicate that activation of PLDα1 inhibits the function of the negative regulator ABI1, thus promoting ABA signaling.

The activation of PLDα1 is interwoven with the function of Gα in Arabidopsis (34). Gα is the only canonical α subunit of the heterotrimeric G protein in Arabidopsis, and it is also involved in ABA signaling and transpirational water loss (35). PLDα1 bound to Gα through a sequence motif is analogous to the DRY motif normally conserved in animal G-protein-coupled receptors. Mutation of amino acid residues in the DRY motif abolished the PLDα1-Gα binding and also the Gα inhibition of PLDα1 activity. Meanwhile, the PLDα1-Gα interaction stimulated the intrinsic GTPase activity of Gα (34). Thus, this

interaction modulates reciprocally the activities of PLDα1 and Gα. These results, together with the PA interaction with ABI1 protein phosphatase 2C, indicate that PLD and the lipid messenger PA are intermediary links between important cellular regulators in plant cells (Figure 2).

Besides PLD, phosphatidylinositol 4,5-bisphosphate (PIP_2)-hydrolyzing PLC has also been implicated in regulating stomatal movement and transpirational water loss. Plants have multiple PI-PLCs [e.g., nine in Arabidopsis (36)], and their domain structures all resemble the latest, simplest animal PLC member PLCζ (1). The PLC activity produces DAG and the water-soluble lipid derivative IP_3 (Figure 1). An earlier study indicated that DAG induced both ion pumping in patch-clamped guard-cell protoplasts and opening of intact stomata (37). DAG is a potent activator of protein kinases in animals, but its signaling function remains an enigma in plants, as its direct target is unclear. Under several conditions, stress-induced DAG was rapidly converted by a DAG kinase to PA (2). IP_3, which is a potent Ca^{2+} mobilizer, has been shown to promote an increase in Ca^{2+} in guard cells and stomatal closure (38). Suppression of a recombinant NrPLC1 reduced ABA-promoted closure of stomata, consistent with a role for PLC, IP_3,

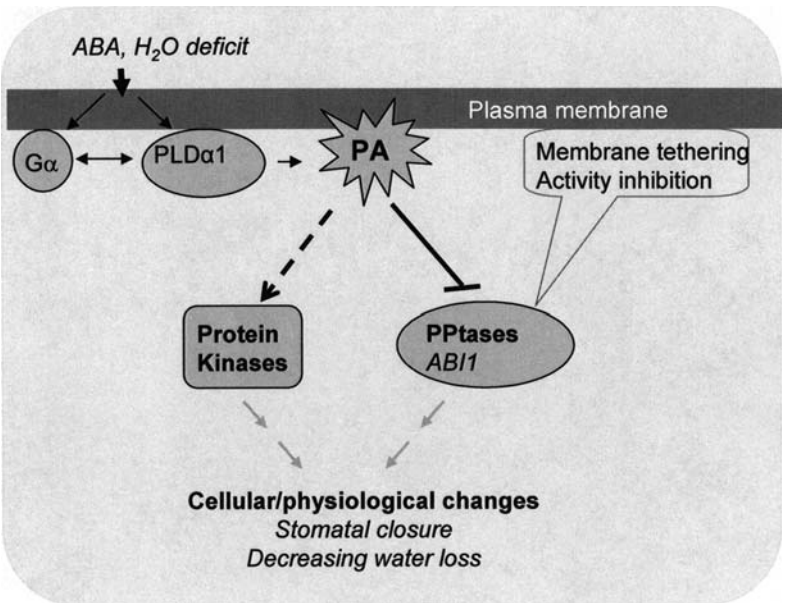


Figure 2. Simplified model depicting the interaction of PLD and PA with other signaling components in mediating ABA response, stomatal closure, and maintaining leaf water status. PLDα1 binds to GDP-bound Gα, and the binding inhibits PLD activity. GTP promotes dissociation of Gα from PLDα1 and releases the Gα inhibition of PLD. Activation of PLDα1 produces PA that binds to ABI1 protein phosphatase 2C (PPtase). The binding decreases ABI1 activity and tethers it to the plasma membrane, thus reducing its translocation to the nucleus, where ABI1 negatively regulates ABA response. Thus, PLDα1 and its derived PA promote ABA response and decrease transcriptional water loss. The dashed line indicates a hypothetical interaction.

and Ca^{2+} flux in stomatal movement (39, 40). The significance of IP_3 has also been investigated by overexpressing or ablating specific inositol phosphate phosphatases, which remove phosphate and, thus, have the potential to down-regulate IP_3 signaling functions. At5PTase1 was up-regulated in response to ABA and was suggested to act as a signal terminator of ABA signaling (41).

In addition to the water-soluble IP_3, phosphatidylinositol (PI) lipids have also been suggested to be involved in mediating stomatal movement. Manipulation of the levels of PI3P and PI4P modulated stomatal closing, and reductions in the levels of functional PI3P and PI4P enhanced stomatal opening (42). PI3P and PI4P are the products of PI 3-kinase (PI-3K) and PI 4-kinase (PI-4K) activities (Table 1).

UNIQUE AND MULTIFACETED FUNCTIONS OF MEDIATORS AND SIGNALING ENZYMES

Temperature and water are the two most crucial environmental factors that limit plant growth, productivity, and geographic distribution. As described above, phospholipid-mediated signaling plays important roles in plant responses to freezing temperature and water stresses. The distinguishable phenotypes resulting from genetic manipulation of different PLDs suggest that the loss of one PLD is not compensated for by the other 11 PLDs in Arabidopsis and that individual PLDs can occupy different steps in cell signaling and/or metabolism. Unique functions of different PLDs could occur by one or a combination of the following: 1) individual PLDs are regulated and activated differently in the cell; 2) they have different temporal and spatial patterns of expressions; 3) they are associated with different membranes; and 4) they have different substrate preferences and have the potential to generate different PA or other derivatives. These differences have been demonstrated for the PLDs characterized (18). Such distinctions ultimately regulate the location and timing of the PLD activities and lipid-derived mediators produced by the enzymes. Spatial and temporal regulation is important to all signaling events, but it is particularly critical to intracellular lipid messengers because of their limited mobility in the cell. Such different functions are expected to occur for some members of other phospholipid-signaling enzyme/gene families, such as PLCs and PLAs. Therefore, it is important to identify specific gene and enzymes involved when the role of a given type of lipid-signaling enzymes in a specific physiological response is addressed. On the other hand, overlapping functions are also likely to occur within each phospholipase family.

In addition, the results also demonstrate that a specific lipid-signaling enzyme can have multifaceted functions. In the case of PLDα1, its suppression increases freezing tolerance, but PLDα1-deficient plants lose more water. Overexpression of PLDα1 decreases water loss, but high PLDα1 activity could be detrimental to plants under severe stress conditions. This is because membrane lipid degradation occurs under many stresses and PLD is often blamed for initiating the degradation. Thus, the use of this gene in plant improvement is complicated, and a better understanding of its function in specific plant processes is needed to help design agronomically desirable applications.

Furthermore, the available data show that a specific physiological response, such as stomatal closure, can involve multiple phospholipid-mediated signaling processes. The aspects of cross-talk among lipid signaling steps have been discussed previously (1, 2, 43). The interaction of lipid messengers with "traditional" signaling components, such as protein phosphatases and kinases, has been documented (8, 9). Investigation of the signaling networks is an exciting field of future research, which has the potential to help advance greatly the current knowledge of the signaling cascades in plant growth and adaptation to adverse conditions.

ACKNOWLEDGMENT

Grant support from the National Science Foundation (IBN-0454866 and CMB-0455318) and the U.S. Department of Agriculture (2001-35304-15308 and 2005-35308-15253) is gratefully acknowledged.

REFERENCES

1 Wang, X. (2004) Curr. Opin. Plant. Biol. 7, 329-336.
2 Meijer, H.J. and Munnik, T. (2003) Ann. Rev. Plant. Biol. 54, 265-306.
3 Ryu, S.B. (2004) Trends Plant Sci. 9, 229-235.
4 Scherer, G.F. (2002) Plant Mol. Biol. 49, 357-372.
5 Farmer, E.E., Almeras, E. and Krishnamurthy, V. (2003) Curr. Opin. Plant Biol. 6, 372-378.
6 Worrall, D., Ng, C.K. and Hetherington, A.M. (2003) Trends Plant Sci. 8, 317-320.
7 Lindsey, K., Pullen, M.L. and Topping, J.F. (2003) Trends Plant Sci. 8, 521-525.
8 Zhang, W., Qin, C., Zhao, J. and Wang, X. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 9508-9513.
9 Anthony, R.G., Henriques, R., Helfer, A., Meszaros, T., Rios, G., Testerink, C., Munnik, T., Deak, M., Koncz, C. and Bogre, L. (2004) EMBO J. 23, 572-581.
10 Thomashow, M.F. (1999) Ann. Rev. Plant Physiol. Plant Mol. Biol. 50, 571-599.
11 Iba, K. (2002) Ann. Rev. Plant Biol. 53, 225-245.
12 Chinnusamy, V., Schumaker, K. and Zhu, J.K. (2004) J. Exp. Bot. 55, 225-236.
13 Xin, Z. and Browse, J. (1998) Proc. Nat. Acad. Sci. U.S.A. 95, 7799-7804.
14 Zhu, J., Shi, H., Lee, B.H., Damsz, B., Cheng, S., Stirm, V., Zhu, J.K., Hasegawa, P.M. and Bressan R.A. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 9873-9878.
15 Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H.E., Rajashekar, C.B., Williams, T.D. and Wang, X. (2002) J. Biol. Chem. 277, 31994-32002.
16 Li, W., Li, M., Zhang, W., Welti, R. and Wang, X. (2004) Nat. Biotechnol. 22, 427-433.

17 Qin, C. and Wang, X. (2002) Plant Physiol. 128, 1057-1068.
18 Wang, X. (2002) Curr. Opin. Plant Biol. 5, 408-414.
19 Qin, B., Wang, C. and Wang, X. (2002) J. Biol. Chem. 277, 49685-49690.
20 Pappan, K., Austin-Brown, S., Chapman, K. and Wang, X. (1998) Arch. Biochem. Biophys. 353, 131-140.
21 Wang, C. and Wang, X. (2001) Plant Physiol. 127, 1102-1112.
22 Zhang, W., Wang, C., Qin, C., Wood, T., Olafsdottir, G., Welti, R. and Wang, X. (2003) Plant Cell 15, 2285-2295
23 Assmann, S.M. (2003) Trends Plant Sci. 8, 151-153.
24 Zhu, J.K. (2002) Ann. Rev. Plant Biol. 53, 247-273.
25 Himmelbach, A., Yang, Y. and Grill, E. (2003) Curr. Opin. Plant Biol. 6, 470-479.
26 Kuhn, J.M. and Schroeder, J.I. (2003) Curr. Opin. Plant Biol. 6, 463-469.
27 Fan, L., Zheng, S. and Wang, X. (1997) Plant Cell. 9, 2183-2196.
28 Jacob, T., Ritchie, S., Assmann, S.M. and Gilroy, S. (1999) Proc. Nat. Acad. Sci. U.S.A. 96, 12192-12197.
29 Ritchie, S. and Gilroy, S. (1998) Proc. Nat. Acad. Sci. U.S.A. 95, 2697-2702.
30 Gampala, S.S., Hagenbeek, D. and Rock, C.D. (2001) J. Biol. Chem. 276, 9855-9860.
31 Austin-Brown, S.L. and Chapman, K.D. (2002) Plant Physiol. 129, 1892-1898.
32 Sang, Y., Zheng, S., Li, W., Huang, B. and Wang, X. (2001) Plant J. 28, 135-144.
33 Himmelbach, A., Hoffmann, T., Leube, M., Höhener, B. and Grill, E. (2002) EMBO J. 21, 3029-3038.
34 Zhao. J. and Wang, X. (2004) J. Biol. Chem. 279, 1794-1800.
35 Wang, X.Q., Ullah, H., Jones, A.M. and Assmann, S.M. (2001) Science 292, 2070-2072.
36 Mueller-Roeber, B. and Pical, C (2002) Plant Physiol. 130, 22-46.
37 Lee, Y, and Assmann, S.M. (1991) Proc. Nat. Acad. Sci. U.S.A. 88, 2127-2131.
38 Staxen, I. I., Pical, C., Montgomery, L.T., Gray, J.E., Hetherington, A.M. and McAinsh, M.R. (1999) Proc. Nat. Acad. Sci. U.S.A. 96, 1779-1784.
39 Hunt, L., Mills, L.N., Pical, C., Leckie, C.P., Aitken, F.L., Kopka, J., Mueller-Roeber, B., McAinsh, M.R., Hetherington, A.M. and Gray J.E. (2003) Plant J. 34, 47-55.
40 Mills, L.N., Hunt, L., Leckie, C.P., Aitken, F.L., Wentworth, M., McAinsh, M.R., Gray, J.E. and Hetherington, A.M. (2004) J. Exp. Bot. 55, 199-204.
41 Burnette, R.N., Gunesekera, B.M. and Gillaspy, G.E. (2003) Plant Physiol. 132, 1011-1019.
42 Jung, J.Y., Kim Y.W., Kwak, J.M., Hwang, J.U., Young, J., Schroeder, J.I., Hwang. I. and Lee Y. (2002) Plant Cell. 14, 2399-2412.
43 Wang, X., Wang, C., Sang, Y., Qin, C. and Welti, R. (2002) Physiol. Plant 114, 331-335.

ANIONIC NUTRIENT TRANSPORT IN PLANTS: THE MOLECULAR BASIS OF THE SULFATE TRANSPORTER GENE FAMILY

Hideki Takahashi,[1] Naoko Yoshimoto,[1] and Kazuki Saito[1,2]

[1]RIKEN Plant Science Center
Tsurumi-ku, Yokohama 230-0045
Japan

[2]Graduate School of Pharmaceutical Sciences
Chiba University
Chiba 263-8522
Japan

INTRODUCTION

Nitrogen, sulfur, and phosphorus are the major and essential elements for all organisms. For higher plants, they are available in the form of inorganic anions (i.e., nitrate, phosphate, and sulfate) in the soil solution (1). Ion transporter proteins are localized at the plasmalemma and organellar membranes to facilitate the uptake and internal transport of specific nutrient across the membrane barriers. In plants, nutrients are transported through consecutive layers of cells differentiated to form the structures of multicellular and multiorgan complexes. To fulfill the entire processes of the uptake and internal translocation of nutrients, higher plants have developed specialized components of membrane-bound transporters, prerequisite for the coordinate regulation of nutrient distribution within the whole plant system. Although the ion specificities of transport functions are strictly defined for each transporter, machineries of nitrate-, sulfate-, and phosphate-uptake systems are mechanistically similar. These major anionic

Genetic Engineering, Volume 27, Edited by J. K. Setlow

nutrients can be imported from rhizosphere to root cells, utilizing the proton gradient across the plasmalemma as a driving force. The uptake of anionic nutrient is generally coupled with the movement of protons. Anion and protons move through a membrane-bound symporter (co-transporter), and the membrane potential is depolarized by an excess import of protons; for instance, nitrate ion is imported to the cell with two or more protons. However, the membrane potential of plasmalemma is immediately reconstituted by proton-ATPase, exporting protons back to the extracellular space (2). Consequently, the mechanism of anion uptake is an energy-dependent process requiring ATP as an energy source. As mentioned above, the overall process of nutrient transport system is organized through the functions of specific transporters that facilitate proper distribution and allocation of nutrients to the developing cell types in plants. It is suggested that multiplicities of the members of transporter protein families in the plant genome may provide varying functional roles and specificities to meet with the emerging complexities of cell-type functions at the whole plant level.

Among the major nutrient transporters in higher plants, the sulfate transporters are the most well-characterized family that exemplifies the processes of anion transport and regulatory mechanisms (3, 4). The fundamental mechanism of plant sulfate uptake system has been demonstrated from the earlier physiological studies by Epstein and co-workers (5, 6). It has been clearly documented that transport of sulfate can be resolved into high- and low-affinity phases and that the capacity of uptake was stimulated by sulfur limitation (5-9). During the last decade, molecular studies have presented the functions of multiple isoforms of sulfate transporter proteins delineating the biphasic kinetics of the uptake in plants (10-17). As expected from the physiological studies, the expression of high-affinity sulfate transporter was positively regulated on sulfur starvation in root epidermal cells (13), showing drastic increase of uptake capacities and transition of kinetic parameters to the high-affinity phase (11, 15, 17). Accordingly, shortage of sulfate is the major environmental impact stimulating the regulatory scheme of sulfate uptake and assimilation pathways (18-21). In addition to the action of regulatory circuitries that optimize the gene expression of sulfate transporters, the activity of sulfate influx can be controlled at the levels of protein turnover or modification of transporters.

In contrast to a clear-cut view of the initial sulfate uptake system, which is linked to the functionalities and regulation of high-affinity transporters, the internal transport of sulfate appears to be rather complicated. The underlying complexities may derive from the fact that sulfate is transported through multiple cell layers and vascular tissues to reach the aerial organs. Within the cell, sulfate is transported to the vacuoles for storage, and partly allocated to the chloroplasts as a sulfur source for the metabolism (4). Multiple steps of transmembrane transport systems become necessary to control interorgan and intracellular distribution of sulfate in plants. For these reasons, the low-affinity transporters responsible for the steps of internal distribution of sulfate are suggested to have diverse functions (12, 22-24) and are distinguishable from the high-affinity components mediating the initial uptake.

After completion of the genome sequencing of *Arabidopsis thaliana* (25), systematic reverse genetics approaches allowed us to unveil the functions of sul-

fate transporters both from the standpoints of molecular and whole-plant nutrient physiology. This chapter will focus on the four major steps of sulfate transport systems regulated under sulfur deficiency in Arabidopsis as a model dicotyledonous plant: 1) initial uptake of sulfate from the soil; 2) root-to-shoot translocation of sulfate; 3) functional relevance of vacuolar transporter in the long-distance transport of sulfate; and 4) source-to-sink transport of sulfate. Recent studies provided evidence that transporters bearing specific functionality and cell-type localization will function as determinants controlling the uptake and distribution of sulfate in plants (10-17, 22-24). The data presented here are the up-to-date knowledge of the functions of sulfate transporter gene family in plants.

SULFATE TRANSPORTER GENE FAMILY IN EUKARYOTES

The studies on plant sulfate transporter gene family have proliferated after the discovery of yeast sulfate transporters that facilitate the uptake of sulfate from the external media (26). Plant sulfate transporters were first identified as membrane proteins capable of restoring the sulfate transport function of a selenate-tolerant mutant of *Saccharomyces cerevisiae*. Using this expression cloning system, plant sulfate transporters have been identified from the libraries of various plant species (10-17, 22, 23, 27). More recently, complete sets of sulfate transporter gene families have been identified through database searches of the genomic resources of Arabidopsis and rice (Table 1, Figure 1). Plant sulfate transporters identified so far were predicted to have 12 membrane-spanning domains (10-12) typical of the membranous structures of the members of the major facilitator family proteins. The C-terminus hydrophilic extension is referred to as STAS (after sulfate transporter and anti-sigma factor antagonist) domain (28), which is structurally related to a bacterial transcription effector component, anti-sigma factor antagonist. Both the Arabidopsis and rice genomes contained 12 homologues of sulfate transporters that are classified into four distinct groups (3, 4, 22) (Figure 1). As mentioned in the following sections, the phylogenic classifications corresponded to the kinetic properties and expression patterns of group members, representing close relationships with their physiological functions. A low-stringent similarity search identifies two additional membrane proteins relevant to this family but lacking the STAS domain (3, 4). The functionalities of this fifth divergent group remain unclear.

From the nutritional point of view, sulfur is essential for growth and development of all organisms. Within the sulfur cycle of natural ecosystem (29), plants fundamentally contribute in providing organic sulfur source to mammals. The proton/sulfate co-transport system facilitates the acquisition of inorganic sulfate to plant roots, and the key manufacturing processes of the biosynthesis of organic sulfur metabolites are initiated from the reductive sulfur assimilation pathways that predominantly take place in the chloroplasts, utilizing the energy and carbon sources generated through photosynthesis (30-33). Sulfur is present in major cellular components and stands as an active moiety mediating a number of fundamental biological processes. Thiols in cysteine residues serve in formation of disulfide bonds maintaining the protein structures. They are essential in

Table 1. Sulfate transporter gene family in Arabidopsis.

Group	Gene name	Locus number	Genbank Accession	References
Group 1	At*SULTR1;1*	At4g08620	AB018695	13, 15, 20
	At*SULTR1;2*	At1g78000	AB042322	16-19
	At*SULTR1;3*	At1g22150	AB049624	22
Group 2	At*SULTR2;1*	At5g10180	AB003591	12, 23
	At*SULTR2;2*	At1g77990	D85416	13
Group 3	At*SULTR3;1*	At3g51895	D89631	13
	At*SULTR3;2*	At4g02700	AB004060	13
	At*SULTR3;3*	At1g23090	AB023423	13
	At*SULTR3;4*	At3g15990	AB054645	
	At*SULTR3;5*	At5g19600	AB061739	23
Group 4	At*SULTR4;1*	At5g13550	AB008782	24
	At*SULTR4;2*	At3g12520	AB052775	24

The locus numbers derive from TIGR (the Institute of Genomic Research; http://www.tigr.org/tdb/e2kl/ath1) and MIPS (Munich Information Center for Protein Sequences, http://mips.gsf.de/proj/thal/db/index.html) databases.

controlling the enzyme reactions in various metabolic pathways, together with the action of redox equivalents as prosthetic groups. In addition, plants generate sulfolipids, sulfated cell-wall constituents, and various secondary sulfonated and sulfated metabolites (30-33), including a variety of repellants and flavored compounds found to have therapeutic and beneficial effects in improving inflammatory and carcinogenic disorders (34-36).

Apparently, acquisition of sulfate is not primarily necessary for mammals, because the plant-derived organic sulfur constitutes an important source for the diets. However, independent of the studies of plant and yeast family members, numbers of sulfate transporters and related sequences have been identified in mammals (37-43). They are categorized as the members of SLC26 family proteins (Figure 1) (44). Mammalian sulfate transporters were initially identified from the expression cloning of renal membrane–localizing transporters in oocyte cells (37) and from the positional cloning of genetic disorders related to the abnormality in cartilage formation (38) and syndromic deafness (41, 42). Recent studies on the functionalities of SLC26 transporters indicate that they are functional as anion exchangers, capable of transporting both the mono- and divalent anions, including sulfate, chloride, iodide, oxalate, and hydroxide (44). Apparently, unlike the case in plants, and as suggested from a broad range of anions they carry on, the physiological roles of the SLC26 members are diverse in mammalian cells. In terms of regulation, they are functionally coupled with SLC4 anion exchanger and CFTR (cystic fibrosis transmembrane regulator) for the balancing of overall anion homeostasis in the renal system (44).

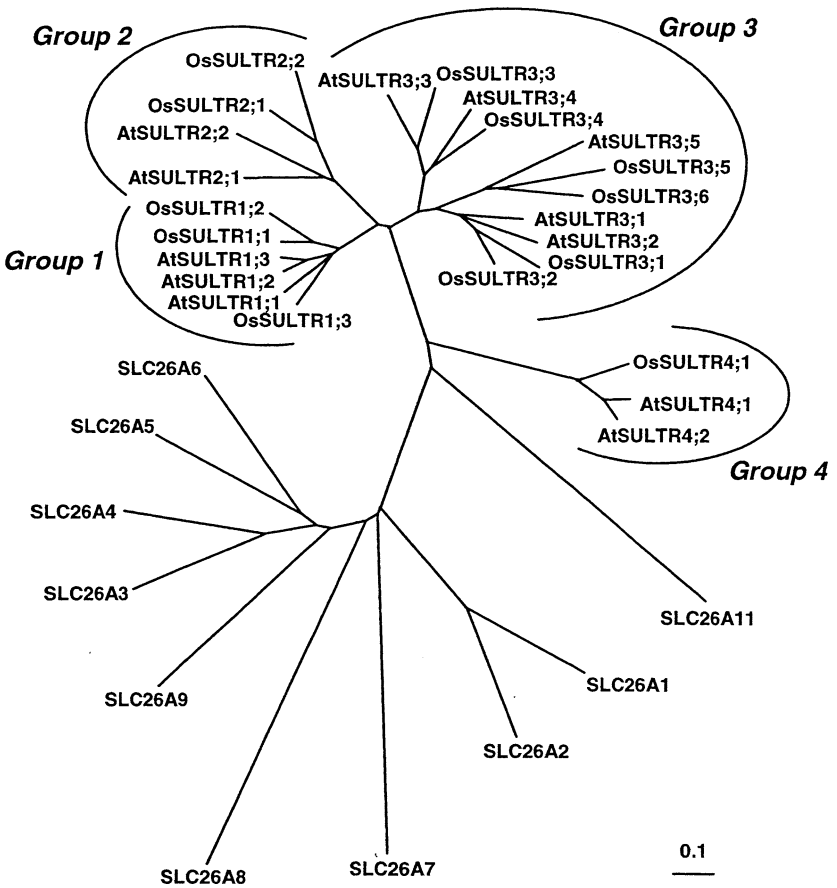


Figure 1. Phylogenic relationships of Arabidopsis and rice sulfate transporters and SLC26 family proteins in the human genome. Unrooted tree was constructed from the multiple alignment of the protein sequences of sulfate transporters from Arabidopsis (AtSULTR), rice (OsSULTR), and human SLC26 family proteins using ClustalW (http://www.ddbj.nig.ac.jp/search/clustalw-e.html) and TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) programs. The locus numbers and GenBank accessions for Arabidopsis genes are listed in Table 1. The accessions for the cDNA sequences and annotated open reading frames of rice sulfate transporters (OsSULTR) are as follows: OsSULTR1;1 (AK103007, AAN59767), OsSULTR1;2 (AAN59766), OsSULTR1;3 (BAC98594), OsSULTR2;1 (AAN59769), OsSULTR2;2 (AK067353, AAN59770), OsSULTR3;1 (NM_196532, AAP53801), OsSULTR3;2 (AK066932, AAN06871), OsSULTR3;3 (AK104831, CAE04513), OsSULTR3;4 (AK067270), OsSULTR3;5 (NM_192602, BAB92305), OsSULTR3;6 (AK121195, BAB68064), and OsSULTR4;1 (AK067676, BAD36818). The accessions of human SLC26 family proteins are as follows: SLC26A1 (AF297659), SLC26A2 (NM_000112), SLC26A3 (NM_000111), SLC26A4 (NM_000441), SLC26A5 (AY289133), SLC26A6 (AF416721), SLC26A7 (AF331521), SLC26A8 (AF314959), SLC26A9 (AF314958) and SLC26A11 (AF345195).

Furthermore, recent investigation indicates that STAS domain of SLC26 and R domain of CFTR are the functional moieties conferring the regulation of anion transport through phosphorylation-mediated interaction between the two transporter molecules (45).

UPTAKE OF SULFATE FROM THE SOIL

The primary sulfate uptake at the root surface is mediated by the Group 1 high-affinity sulfate transporters. The molecular characteristics of high-affinity sulfate transporters are studied extensively in various plant species (10, 11, 13-17, 22, 27, 46, 47), which generally corresponded to the high-affinity kinetics of sulfate uptake measurements in the earlier physiological studies. The first isolated cDNAs from *Stylosanthes hamata* were able to complement the yeast sulfate transporter mutant, and exhibited saturable kinetics of sulfate uptake with micromolar K_m values (10). In addition, the mRNA was accumulated in root tissues when plants were starved for sulfate. Genes identified from Arabidopsis showed similar characteristics (13, 15-17). The Arabidopsis *SULTR1;1* and *SULTR1;2* encoded high-affinity sulfate transporters showing the high-affinity kinetics of sulfate uptake. They were localized in the root hairs, epidermis, and cortex of roots and accumulated during sulfur limitation (13, 16, 20). *SULTR1;2* was abundantly expressed under conditions with adequate sulfur supply, playing a major role in facilitating the uptake of sulfate in Arabidopsis roots (16, 17). The absence of SULTR1;2 in the knockout mutant substantially affected the overall sulfur status; decrease in sulfate influx activities resulted in general induction of an array of sulfur assimilation enzymes and oxidative stress responsive genes (18). Induction of *SULTR1;1* mRNA was the typical response observed in sulfate starvation and by the knockout of *SULTR1;2* (13, 18). SULTR1;1 was co-expressed with SULTR1;2 in the same cell layers of root tissue (13, 16, 20); however, the up-regulation of *SULTR1;1* mRNA was not sufficient to fully recover the loss of sulfate uptake activities in the *sultr1;2* knockout mutant (18). When comparing the expression of *SULTR1;1* and *SULTR1;2*, the response of *SULTR1;1* was more sensitive to the fluctuation of sulfur conditions, suggesting that they were regulated in slightly different modes during sulfur limitation in the same cell layers (16, 21). Duplication of two high-affinity sulfate transporters presumably provides flexibility of the uptake system to adapt with a wide range of sulfur conditions in the environment.

REGULATION OF SULFATE UPTAKE

As mentioned above, *SULTR1;1* and *SULTR1;2* were both regulated by the external supply of sulfate (13, 15-17). In addition to sulfate, metabolites of sulfur assimilation pathways affected the gene expression of sulfate transporters. Exogenous application of cysteine and glutathione caused down-regulation, and *O*-acetyl-L-serine (OAS), the precursor of cysteine synthesis, positively affected the expression of high-affinity sulfate transporters, accompanied with decrease and increase in sulfate uptake activities, respectively (11, 14, 15, 19, 21, 48, 49). Apparently, OAS acts positively for the induction of sulfur-responsive genes, including sulfate transporters. This key metabolic intermediate is located at a pivotal center of cysteine synthesis, its content being affected by varying supplies of nitrogen, sulfur, and carbon skeleton (50-52). In addition, the activity of cysteine synthase complex is reversibly regulated by OAS and sulfide, postulated as a sensing unit of cysteine biosynthesis in plants (53). Regulation of sulfate transporters

by external supply of nitrogen and carbon may suggest a close linkage of upstream regulatory cascades with the mechanisms controlling N and C basal metabolisms (52, 54-58). In addition, the flux of sulfur can be constantly affected by the metabolic connections with the basal metabolisms. At present, the reality of intermetabolic regulatory signal is buried in the complex networks of nutrient assimilation metabolisms.

Besides nutrient-dependent and metabolic regulations, signal transduction pathways for the control of sulfur assimilation have not been demonstrated in photosynthetic organisms until the discovery of *Sac* genes in *Chlamydomonas reinhardtii* (59-61). Sac1 was a membrane-bound sulfur sensor-like protein, and its function was closely associated with the general control of utilization of photosynthetic energy in the major nutrient assimilatory pathways (60, 62). As for the regulation of sulfur assimilation, the expression of sulfur-responsive genes was generally under the control of Sac1, activating both the primary acquisition and remobilization of sulfur source under sulfur deficiency (63). Another key regulatory factor, Sac3, was a protein kinase, negatively regulating the expression of arylsulfatase and sulfate uptake activities under sulfur-replete conditions (61).

In Arabidopsis, recent findings suggested that a cytokinin-dependent signaling cascade participates in conducting the negative regulation of sulfate uptake in roots (19). Accumulation of *SULTR1;1* and *SULTR1;2* mRNAs in roots decreased by the addition of cytokinin to the medium. The regulatory pathway was suggested to involve a two-component phospho-relay system initiated by a cytokinin receptor histidine kinase (64-67). The *cre1-1* mutant of Arabidopsis lacks the receptor kinase, CRE1/WOL/AHK4 (68-70), and was unable to regulate the expression of high-affinity sulfate transporters in response to cytokinin (19). The current model suggests that two independent regulatory cues may control the uptake of sulfate, one requiring the cytokinin-derived signal, and the other modulated by sulfur. In addition, these two potential mechanisms worked independently for the control of sulfate uptake activities (19). When sulfate is limiting, plants activate the expression of high-affinity sulfate transporters for the acquisition of sulfur source. By contrast, cytokinin provides a negative signal to turn down the influx of sulfate, presumably working in parallel with the regulatory cascade that attenuates the acquisition of phosphate (71-73).

ROOT-TO-SHOOT TRANSLOCATION OF SULFATE

Following the primary uptake, sulfate is transported to the vasculature and distributed to the aerial tissues. First, the incorporated sulfate will go through the symplastic pathway across a number of cell layers in roots to reach the central cylinder (4). Inside the vascular tissue, sulfate in the symplast will be uploaded from the xylem parenchyma cells to the vessels. However, along with the pathway of symplastic transfer of sulfate in the vasculatures, sulfate may leak out to the apoplastic space probably through an unidentified anion channel, utilizing an outside positive membrane potential (Figure 2). Although the mechanism of xylem loading of sulfate has not been clarified at the molecular levels, recent studies suggested that low-affinity sulfate transporters localized at the plasma

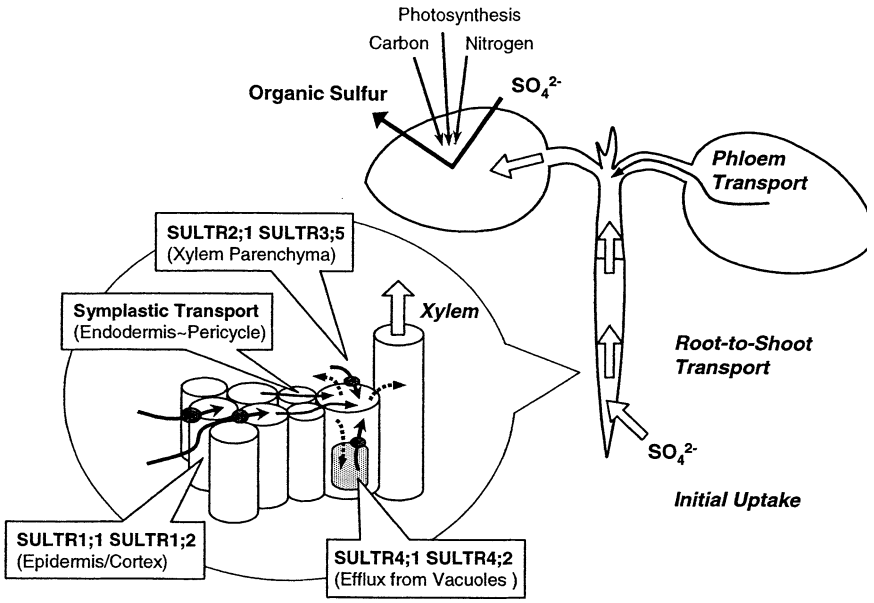


Figure 2. Hypothetical model of sulfate transport systems in Arabidopsis. The model illustrates the steps for the initial uptake and internal distribution of sulfate activated under sulfur deficiency. SULTR1;1-SULTR1;2 high-affinity sulfate transporters facilitate the uptake of sulfate from the soil at epidermis, cortex, and root hairs. Symplastic pathway mediates transport of sulfate through the endodermal cell layer. SULTR2;1-SULTR3;5 in the plasmalemma and SULTR4;1-SULTR4;2 in the tonoplast of xylem parenchyma cells maintain the symplastic flux of sulfate channeled toward the xylem vessels. Xylem loading, influx to vacuoles and efflux to apoplast, are the unidentified steps of sulfate transport, as indicated by dashed lines.

membranes of xylem parenchyma cells may participate in minimizing the leakage of sulfate from the symplastic flux of sulfate loaded to the xylem vessels (Figure 2) (23).

In Arabidopsis, *SULTR2;1* is known to localize in the vascular tissues of sulfate-starved plant roots (12, 13). The measurement of kinetic properties in yeast mutant suggested that SULTR2;1 represents the low-affinity transport system in Arabidopsis (13). The cellular localization of *SULTR2;1* and its responsiveness to sulfur limitation stress led us to postulate that the function of this low-affinity transporter is essential to activate the root-to-shoot transport of sulfate under sulfur deficiency. However, the situation was even more complicated. For the full expression of low-affinity sulfate transport activity, a second plasma membrane–localizing transporter molecule, SULTR3;5, was additionally required (23). SULTR3;5 was a novel nonfunctional subtype, classified as a member of Group 3 sulfate transporters. In yeast expression systems, SULTR3;5 itself showed no activity of sulfate uptake but was able to maximize the activity of SULTR2;1 in the co-expression system. These results suggest that low-affinity transport of sulfate in the vasculature is probably facilitated by the interplay of two transporters. In Arabidopsis, SULTR3;5 was co-localized with SULTR2;1 in

the pericycle and xylem parenchyma cells of roots but was expressed rather constantly regardless of the sulfur conditions. When SULTR2;1 was abundantly accumulated by sulfur limitation, the absence of SULTR3;5 in the knockout mutant affected the root-to-shoot transport of sulfate (23). SULTR3;5 functioned as an essential component of the low-affinity transport system in the root vasculature only in the presence of the inducible isoform, SULTR2;1.

VACUOLAR STORAGE AND RELEASE OF SULFATE REGULATES LONG-DISTRANCE TRANSLOCATION

Despite the evidence that the majority of sulfate accumulates in the vacuoles, the mechanisms of vacuolar sulfate transport systems have long been left undetermined (74, 75). At the tonoplast membranes, proton-ATPase and proton-pyrophosphatase generate an inside positive membrane potential (75, 76), which will be utilized as a driving force for the influx of negatively charged ions to the vacuoles (77-80). It is most probable that sulfate moves into the vacuoles utilizing this membrane potential (74); however, the specific facilitator protein for the influx has not been identified. As a storage compartment, the vacuole needs to release sulfate in response to the requirement of sulfur from the metabolisms. In general, vacuoles are acidified by an active import of protons through the action of ATPase and pyrophosphatase (75, 76), providing ideal circumstances for proton-coupled sulfate transporters to export sulfate back to the cytoplasm. Recently, the members of Group 4 sulfate transporters were demonstrated to represent this efflux system (24). Direct evidence for the efflux function was suggested from the increase of sulfate storage in the vacuoles of the *sultr4;1 sultr4;2* double knockout. In addition, SULTR4;1 and SULTR4;2 were both abundantly expressed in sulfur-starved plants of Arabidopsis, suggesting that remobilization of the vacuolar sulfate pool is significant when the external source of sulfur becomes limiting (24).

The significance of SULTR4-type vacuolar sulfate transporter was its contribution in controlling the turnover of sulfate pool in the root tissues. Both isoforms were predominantly expressed in the vascular tissues of roots and hypocotyls, and were suggested to participate in the step of root-to-shoot transport of sulfate by discharging the vacuolar sulfate reserve in the vasculature for xylem loading. The *sultr4;1 sultr4;2* double knockout accumulated sulfate in roots, whereas expression of SULTR4;1-GFP restored translocation of sulfate to shoots (24). It is suggested that the main function of SULTR4 is to prevent excessive accumulation of sulfate in the root vacuoles before the entry to the xylem stream. This mechanism is suggested to be important in maintaining the cytoplasmic flux of sulfate channeled toward the xylem vessels, supporting the function of SULTR2;1-SULTR3;5 low-affinity transport system at the plasma membrane (23). The current knowledge suggests that SULTR4;1-SULTR4;2 and SULTR2;1-SULTR3;5 were both essential components for the maintenance of loading flux (Figure 2) (23, 24). Under sulfur deficiency, the main portion of the induction of root-to-shoot sulfate transport activity is suggested to require an additional core component that corresponds to the function of sulfate efflux

facilitator; however, identification of the molecular mechanism of loading process at the root xylem parenchyma cells awaits further investigation (Figure 2).

SOURCE-TO-SINK TRANSPORT OF SULFUR

Phloem facilitates remobilization of nutrients to the sink organs. The evidence showing localization of specific sulfate transporters in the companion cell/sieve element complexes suggested contribution of the phloem transport systems in interorgan translocation of sulfate and organic sulfur metabolites in plants (13, 22, 48, 81). *SULTR1;3* was a unique member of Group 1, encoding a high-affinity sulfate transporter expressed in the transport phloems of roots, hypocotyls, and cotyledons (22). The role of SULTR1;3 was suggested from the analysis of the *sultr1;3* knockout. Radioactive sulfate was translocated from cotyledon to sink tissues in the wild type, but was rather immobilized to the fed organs in the *sultr1;3* knockout. SULTR1;3 was preferentially localized in the regions where the transport phloem facilitates the solute transfer between the source and sink organs. From these observations, it is suggested that the function of this transporter is not associated directly with the initial loading processes that mainly occur at the collection phloem, but rather participates in retrieval of sulfate to the companion cells of the transport phloem, preventing the leakage of sulfur source from the phloems during long-distance translocation. In addition, SULTR2;2 low-affinity sulfate transporter is reported to localize in the phloem of Arabidopsis (13); however, the role of this low-affinity component has not been determined in detail.

FUTURE PERSPECTIVES

Recent development in plant genomics research has aided us to discover the full sets of sulfate transporter gene families both from Arabidopsis and rice (4, 12, 13, 15-17, 22-24) (Table 1, Figure 1). One of the most prosperous achievements during the last decade was the use of reverse genetics tools for the analysis of individual transport components at the whole plant level, allowing us to determine the physiological roles of sulfate transporters *in vivo* (Figure 2) (17, 18, 22-24). These progresses certainly accelerated to bring the conventional knowledge of plant nutrient physiology into the focus of plant molecular biology. The roles of transporters defined from the molecular studies have re-evaluated the earlier findings in nutrient physiology in the past 50 years. Although the identification of facilitators for xylem loading and sulfate influx systems in the vacuoles and chloroplasts remain unidentified (74, 82, 83), the current findings described in this review provided us an overall view of sulfate transporters facilitating the uptake and vascular distribution of sulfate in plants. But at the same time, a considerable amount of questions arose, particularly on the aspects of regulation of transport processes. As for the regulation of gene expression, signal transduction cascades are driven by numerous external and internal signals, initiated by fluctuations of nutrients and metabolites, and by plant hormones (11, 14, 15, 19-21, 48, 49, 51, 52, 54-58). In addition, the multiplicities of functional transporter components may provide a fine-tuned control of transport activities in response

to the environmental impacts. Consequently, the functions of transporters must be properly integrated to the individual transport steps, playing specialized roles in various cell types differentiated during development. Furthermore, the overall transport processes are coordinately regulated to optimize distribution and allocation of sulfate within the plant body (Figure 2). The upcoming progresses are expected to provide us a more detailed picture of the regulatory mechanisms, which will form the basis for the genetic engineering of sulfate transport and assimilation in higher plants, and will be extended for the improvement of sulfur qualities of crop plant species (84-86).

REFERENCES

1 Marschner, H. (1995) Mineral Nutrition of Higher Plants, second edition, Academic Press, New York, pp. 231-277.
2 Palmgren, M.G. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 817-845.
3 Hawkesford, M.J. (2003) Physiol. Plant. 117, 155-163.
4 Buchner, P., Takahashi, H. and Hawkesford, M.J. (2004) J. Exp. Bot. 55, 1765-1773.
5 Leggett, J.E. and Epstein, E. (1956) Plant Physiol. 31, 222-226.
6 Epstein, E. (1966) Nature 212, 1324-1327.
7 Clarkson, D.T., Smith, F.W. and Vanden Berg, P.J. (1983) J. Exp. Bot. 34, 1463-1483.
8 Hawkesford, M.J., Davidian, J-C. and Grignon, C. (1993) Planta 190, 297-304.
9 Lass, B. and Ullrich-Eberius, C.L. (1984) Planta 161, 53-60.
10 Smith, F.W., Ealing, P.M., Hawkesford, M.J. and Clarkson, D.T. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 9373-9377.
11 Smith, F.W., Hawkesford, M.J., Ealing, P.M., Clarkson, D.T., Vanden Berg, P.J., Belcher, A.R. and Warrilow, A.G.S. (1997) Plant J. 12, 875-884.
12 Takahashi, H., Yamazaki, M., Sasakura, N., Watanabe, A., Leustek, T., de Almeida Engler, J., Engler, G., Van Montagu, M. and Saito, K. (1997) Proc. Nat. Acad. Sci. U.S.A. 94, 11102-11107.
13 Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., Hawkesford, M.J. and Saito, K. (2000) Plant J. 23, 171-182.
14 Vidmar, J.J., Schjoerring, J.K., Touraine, B. and Glass, A.D.M. (1999) Plant Mol. Biol. 40, 883-892.
15 Vidmar, J.J., Tagmount, A., Cathala, N., Touraine, B. and Davidian, J-C.E. (2000) FEBS Lett. 475, 65-69.
16 Yoshimoto, N., Takahashi, H., Smith, F.W., Yamaya, T. and Saito, K. (2002) Plant J. 29, 465-473.
17 Shibagaki, N., Rose, A., Mcdermott, J.P., Fujiwara, T., Hayashi, H., Yoneyama, T. and Davies, J.P. (2002) Plant J. 29, 475-486.
18 Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yamaya, T. and Takahashi, H. (2003) Plant Physiol. 132, 597-605.
19 Maruyama-Nakashita, A., Nakamura, Y., Yamaya, T. and Takahashi, H. (2004) Plant J. 38, 779-789.

20 Maruyama-Nakashita, A., Nakamura, Y., Watanabe-Takahashi, A., Yamaya, T. and Takahashi, H. (2004) Plant Cell Physiol. 45, 340-345.
21 Maruyama-Nakashita, A., Nakamura, Y., Yamaya, T. and Takahashi, H. (2004) J. Exp. Bot. 55, 1843-1849.
22 Yoshimoto, N., Inoue, E., Saito, K., Yamaya, T. and Takahashi, H. (2003) Plant Physiol. 131, 1511-1517.
23 Kataoka, T., Hayashi, N., Yamaya, T. and Takahashi, H. (2004) Plant Physiol. 136, (doi/10.1104/pp. 104.045625).
24 Kataoka, T., Watanabe-Takahashi, A., Hayashi, N., Ohnishi, M., Mimura, T., Buchner, P., Hawkesford, M.J., Yamaya, T. and Takahashi, H. (2004) Plant Cell 16, 2693-2704.
25 Arabidopsis Genome Initiative (2000) Nature 408, 796-815.
26 Smith, F.W., Hawkesford, M.J., Prosser, I.M. and Clarkson, D.T. (1995) Mol. Gen. Genet. 247, 709-715.
27 Howarth, J.R., Fourcroy, P., Davidian, J-C., Smith, F.W. and Hawkesford, M.J. (2003) Planta 218, 58-64.
28 Aravind, L. and Koonin, E.V. (2000) Curr. Biol. 10, R53-55.
29 Giovanelli, J., Mudd, S.H. and Datko, A.N. (1980) in The Biochemistry of Plants: A Comprehensive Treatise (P.K. Stumpf and E.E. Conn, eds.), Academic Press, New York, pp. 453-505.
30 Leustek, T. and Saito, K. (1999) Plant Physiol. 120, 637-643.
31 Saito, K. (2000) Curr. Opin. Plant Biol. 3, 188-195.
32 Leustek, T., Martin, M.N., Bick, J. and Davies, J.P. (2000) Ann. Rev. Plant Physiol. Plant Mol. Biol. 51, 141-166.
33 Saito, K. (2004) Plant Physiol. 136, 2443-2450.
34 Wittstock, U. and Halkier, B.A. (2002) Trends Plant Sci. 7, 263-270.
35 Shapiro, T.A., Fahey, J.W., Wade, K.L., Stephenson, K.K. and Talalay, P. (2001) Cancer Epidemiol. Biomarkers Prev. 10, 501-508.
36 Parcell, S. (2002) Altern. Med. Rev. 7, 22-44.
37 Bissig, M., Hagenbuch, B., Stieger, B,. Koller, T. and Meier, P.J. (1994) J. Biol. Chem. 269, 3017-3021.
38 Hästbacka, J., de la Chapelle, A., Mahtani, M.M., Clines, G., Reeve-Daly, M.P., Daly, M., Hamilton, B.A., Kusumi, K., Trivedi, B., Weaver, A., Coloma, A., Lovett, M., Buckler, A. and Lander, E.S. (1994) Cell 78, 1073-1087.
39 Schweinfest, C.W., Henderson, K.W., Suster, S., Kondoh, N. and Papas, T.S. (1993) Proc. Nat. Acad. Sci. U.S.A. 90, 4166-4170.
40 Höglund, P., Haila, S., Socha, J., Tomaszewski, L., Saarialho-Kere, U., Karjalainen-Lindsberg, M.L., Airola, K., Holmberg, C., de la Chapelle, A. and Kere, J. (1996) Nat. Genet. 14, 316-319.
41 Coyle, B., Coffey, R., Armour, J.A.L, Gausden, E., Hochberg, Z., Grossman, A., Britton, K., Pembrey, M., Reardon, W. and Trembath, R. (1996) Nat. Genet. 12, 421-423.
42 Sheffield, V.C., Kraiem, Z., Beck, J.C., Nishimura, D., Stone, E.M., Salameh, M., Sadeh, O. and Glaser, B. (1996) Nat. Genet. 12, 424-426.
43 Knauf, F., Yang, C.L., Thomson, R.B., Mentone, S.A., Giebisch, G. and Aronson, P.S. (2001) Proc. Nat. Acad. Sci. U.S.A. 98, 9425-9430.

44 Mount, D.B. and Romero, M.F. (2004) Pflugers Arch. 447, 710-721.
45 Ko, S.B., Zeng, W., Dorwart, M.R., Luo, X., Kim, K.H., Millen, L., Goto, H., Naruse, S., Soyombo, A., Thomas, P.J. and Muallem, S. (2004) Nat. Cell Biol. 6, 343-350.
46 Bolchi, A., Petrucco, S., Tenca, P.L., Foroni, C. and Ottonello, S. (1999) Plant Mol. Biol. 39, 527-537.
47 Buchner, P., Prosser, I. and Hawkesford, M.J. (2004) Genome 47, 526-534.
48 Lappartient, A.G., Vidmar, J.J., Leustek, T., Glass, A.D.M. and Touraine, B. (1999) Plant J. 18, 89-95.
49 Clarkson, D.T., Diogo, E. and Amancio, S. (1999) Plant Physiol. Biochem. 37, 283-290.
50 Kim, H., Hirai, M.Y., Hayashi, H., Chino, M., Naito, S. and Fujiwara, T. (1999) Planta 209, 282-289.
51 Hirai, M.Y., Fujiwara, T., Awazuhara, M., Kimura, Y., Noji, M. and Saito, K. (2003) Plant J. 33, 651-663.
52 Koprivova, A., Suter, M., Op den Camp, R., Brunold, C. and Kopriva, S. (2000) Plant Physiol. 122, 737-746.
53 Berkowitz, O., Wirtz, M., Wolf, A., Kuhlmann, J. and Hell, R. (2002) J. Biol. Chem. 277, 30629-30634.
54 Kopriva, S., Muheim, R., Koprivova, A., Trachsel, N., Catalano, C., Suter, M. and Brunold, C. (1999) Plant J. 20, 37-44.
55 Kopriva, S., Suter, M., von Ballmoos, P., Hesse, H., Krahenbuhl, U., Rennenberg, H. and Brunold, C. (2002) Plant Physiol. 130, 1406-1413.
56 Hesse, H., Trachsel, N., Suter, M., Kopriva, S., von Ballmoos, P., Rennenberg, H. and Brunold, C. (2003) J. Exp. Bot. 54, 1701-1709.
57 Wang, R., Okamoto, M., Xing, X. and Crawford, N. (2003) Plant Physiol. 132, 556-567.
58 Lejay, L., Gansel, X., Cerezo, M., Tillard, P., Müller, C., Krapp, A., von Wirén, N., Daniel-Vedele, F. and Gojon, A. (2003) Plant Cell 15, 2218-2232.
59 Davies, J.P., Yildiz, F.H. and Grossman, A.R. (1994) Plant Cell 6, 53-63.
60 Davies, J.P., Yildiz, F.H. and Grossman, A.R. (1996) EMBO J. 15, 2150-2159.
61 Davies, J.P., Yildiz, F.H. and Grossman, A.R. (1999) Plant Cell 11, 1179-1190.
62 Grossman, A. and Takahashi, H. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 163-210.
63 Takahashi, H., Braby, C.E. and Grossman, A.R. (2001) Plant Physiol. 127, 665-673.
64 Mizuno, T. (1998) J. Biochem. (Tokyo) 123, 555-563.
65 Schmülling, T. (2001) Trends Plant Sci. 6, 281-284.
66 Hutchinson, C.E. and Kieber, J.J. (2002) Plant Cell 14, S47-S59.
67 Kakimoto, T. (2003) Ann. Rev. Plant Biol. 54, 605-627.
68 Inoue, T., Higuchi, M., Hashimoto, Y., Seki, M., Kobayashi, M., Kato, T., Tabata, S., Shinozaki, K. and Kakimoto, T. (2001) Nature 409, 1060-1063.

69 Mähönen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N. and Helariutta, Y. (2000) Genes Dev. 14, 2938-2943.
70 Suzuki, T., Miwa, K., Ishikawa, K., Yamada, H., Aiba, H. and Mizuno T. (2001) Plant Cell Physiol. 42, 107-113.
71 Martin, A.C., del Pozo, J.C., Iglesias, J., Rubio, V., Solano, R., de la Pena, A., Leyva, A. and Paz-Ares, J. (2000) Plant J. 24, 559-567.
72 Franco-Zorrilla, J.M., Martin, A.C., Solano, R., Rubio, V., Leyva, A. and Paz-Ares, J. (2002) Plant J. 32, 353-360.
73 Karthikeyan, A.S., Varadarajan, D.K., Mukatira, U.T., D'Urzo, M.P., Damsz, B. and Raghothama, K.G. (2002) Plant Physiol. 130, 221-233.
74 Kaiser, G., Martinoia, E., Schroppelmeier, G. and Heber, U. (1989) J. Plant Physiol. 133, 756-763.
75 Martinoia, E., Massonneau, A. and Frangne, N. (2000) Plant Cell Physiol. 41, 1175-1186.
76 Maeshima, M. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 469-497.
77 Martinoia, E., Heck, U. and Wiemken, A. (1981) Nature 289, 292-294.
78 Blumwald, E. and Poole, R.J. (1985) Proc. Nat. Acad. Sci. U.S.A. 82, 3683-3687.
79 Massonneau, A., Martinoia, E., Dietz, K.J. and Mimura, T. (2000) Planta 211, 390-395.
80 Mimura, T., Dietz, K.J., Kaiser, W., Schramm, M. J., Kaiser, G. and Heber, U. (1990) Planta 180, 139-146.
81 Herschbach, C., van Der Zalm, E., Schneider, A., Jouanin, L., De Kok, L.J. and Rennenberg, H. (2000) Plant Physiol. 124, 461-473.
82 Hampp, R. and Ziegler, I. (1977) Planta 137, 309-312.
83 Mourioux D and Douce R. (1979) Biochimie 61, 1283-1292.
84 Blake-Kalff, M.M., Harrison, K.R., Hawkesford, M.J., Zhao, F.J. and McGrath, S.P. (1998) Plant Physiol. 118, 1337-1344.
85 Fitzgerald, M.A., Ugalde, T.D. and Anderson, J.W. (2001) J. Exp. Bot. 52, 1519-1526.
86 Sunarpi and Anderson, J.W. (1997) Plant Physiol. 114, 687-693.

PURIFICATION OF PROTEIN COMPLEXES BY IMMUNOAFFINITY CHROMATOGRAPHY: APPLICATION TO TRANSCRIPTION MACHINERY

Nancy E. Thompson, Debra Bridges Jensen,
Jennifer A. Lamberski, and Richard R. Burgess

McArdle Laboratory for Cancer Research
University of Wisconsin–Madison
Madison, Wisconsin 53706

INTRODUCTION

Many proteins must interact with other proteins to achieve cellular and subcellular structures, to acquire biological activity, and/or to regulate biological activity. Traditionally, protein purification methods have often resulted in the separation of protein complexes for the sake of "purity" or have inadvertently destroyed complexes as a result of the chemical or physical basis of the separation procedure. Sometimes the complex can be reconstituted by combining individually purified proteins, but, in many cases, activity cannot be recovered. In addition, it is often difficult to identify the binding partners once they are physically separated. This chapter will address the challenge of purifying protein complexes and will specifically concentrate on the method of immunoaffinity chromatography to achieve this purification.

Immunoaffinity purification of proteins takes advantage of the high specificity of an antibody for an antigen. This powerful method can result in a 1,000-fold purification of a protein in a single step. Generally, the antigen is the

Genetic Engineering, Volume 27, Edited by J. K. Setlow

object of purification, but the antigen can be used as a ligand to purify specific antibody as well. The power of this technique has been recognized for many years, but the actual implementation of the technique required that scientists overcome some major obstacles. One obstacle was the supply of monospecific antibody. This problem was largely overcome by the development of procedures to make monoclonal antibodies (mAbs), which allowed the production of almost an unlimited supply of epitope-specific antibody. Secondly, in order to recover the protein complex, the antigen-antibody interaction had to be dissociated without disrupting the complex or inactivating the biological activity. Again, the development of mAb technology contributed to overcoming this problem by yielding a homogeneous immunological reagent that reacts uniformly to elution conditions. Despite these advances, most antigen–mAb interactions remain very difficult to dissociate. There are many relevant reviews on the production of mAbs, characterization of mAbs, and methods for molecularly engineering mAbs (1-4). This chapter focuses on the critical issue of recovery of intact, biologically active protein complexes by immunoaffinity chromatography.

The simplest form of immunoaffinity purification of protein complexes is co-immunoprecipitation (co-IP). An elegant example of this is the experiments of Harlow et al. (5) describing the use of mAbs that react with the E1A viral protein and the co-IP of several cellular binding partners. These experiments were the preliminary experiments to others that identified the important "pocket proteins." Because these experiments were analytical in nature, activities were not recovered from the protein complex. As in many co-IP experiments, the results were simply analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). In some cases, activity can be recovered from SDS-PAGE (6), but the recovery, again, is on an analytical (μg) scale rather than a preparative scale. Recent genomic approaches (such as mass spectrometry) allow the identification of proteins from small-scale co-IP experiments; however, this information is limited by the lack of assignment of functions to many gene products. To date, only about one-third of the putative genes from the human genome have been assigned a function.

This chapter will briefly describe a procedure to identify mAbs that are useful for immunoaffinity chromatography; it will concentrate on the ability of mAbs to purify protein complexes, and it will discuss immunoaffinity chromatography procedures involving epitope tags to purify protein complexes.

Polyol-Responsive mAbs

To be useful for immunoaffinity chromatography, an antibody must have high affinity for its antigen in order to bind the antigen in a dilute solution. This property, however, generally results in a complex that is difficult to disrupt in order to recover the antigen in an active form. Most elution conditions require extremes of pH values, chaotropic salts, or denaturing reagents. A unique type of mAb has high affinity for a protein antigen, but whose interaction with the antigen can be dissociated under gentle elution conditions (7, 8). This type of mAb has been designated a polyol-responsive mAb (PR-mAb) because the eluting buffer is a neutral, aqueous buffer containing a low molecular weight polyhydroxylated compound (polyol) and a nonchaotropic salt, combinations that are

PR-mAbs are not unique to a specific class or subclass of immunoglobulin. IgG molecules are superior to IgM or IgA molecules because they are easier to purify and more stable. PR-mAbs that are IgG1, IgG2a, and IgG2b molecules have been identified.

Typically, PR-mAbs are derived through the mouse hybridoma methods originally described by Köhler and Milstein (24). By screening more than 200 hybridomas that produce mAbs reactive with subunits of Escherichia coli core RNA polymerase, it has been estimated that about 10% of the hybridomas produce PR-mAbs (8). Another method of producing mAbs, termed antigen-specific plasmacytomas (ASPCT or ASP), which results from infection of plasma cells with an engineered ABL-MYC retrovirus (25), can also produce PR-mAbs (www.neoclone.com). In addition, some mAbs produced by rat hybridomas are PR-mAbs (R. R. Burgess, unpublished data).

To help identify PR-mAbs, an ELISA-based screening assay has been developed and called an ELISA-elution assay (7, 8). This assay is highly reliable for the identification of PR-mAbs. Most of the PR-mAbs listed in Table 1 were identified by this assay specifically for use in immunoaffinity chromatography. mAb 8WG16, which was identified from a collection of existing mAbs, was used as a prototype to define the parameters of PR-mAbs and to develop the ELISA-elution assay (7). Other PR-mAbs have been identified from existing collections of mAbs (R.R. Burgess, unpublished data).

Figure 1A contains a schematic of the ELISA-elution assay. This is a simple ELISA that is modified by incorporating a wash step using a buffer containing salt and polyol after the reaction with the primary antibody and before the addition of the enzyme-conjugated secondary antibody. After the reaction with the substrate, the color intensity is roughly proportional to the amount of primary antibody remaining. PR-mAbs can be identified by a decrease in signal obtained from a well washed with the buffer containing the salt and polyol when compared with a well washed just with the buffer (Figure 1B). Originally, "polyol-responsiveness" was defined as a 50% or more reduction in signal in the ELISA-elution assay (8). Recently, it has been found that some mAbs that gave only a 30% reduction in the ELISA-elution assay worked well for immunoaffinity chromatography (R.R. Burgess, unpublished data). Because preliminary screening can be performed with only 50 μL of cell supernatant per well, hybridomas can be screened while they are still being cultured in the master wells. After the cells are cloned and more cell supernatant is available, the ELISA-elution can then be used to examine different salt and polyol combinations and concentrations.

Screening for PR-mAbs is generally performed with a general buffer such as TE buffer (Tris-HCl, pH 7.9; 0.1 mM EDTA) containing 0.75 M ammonium sulfate and 40% (v/v) propylene glycol as the eluting reagent. A PR-mAb will usually respond to a variety of salt and polyol combinations, but there are exceptions. mAb IIB8 only responds well to combinations of ammonium sulfate or sodium chloride and propylene glycol (17; unpublished data). Also, 8WG16 responds to a variety of salt and polyol combinations (7, 9), but it also responds to just 50% glycerol (7, 12). A PR-mAb that responds only to glycerol with or without the presence of salt is rare, although the PR-mAb described by Nagy et al. (23) responds only to ethylene glycol.

generally protein stabilizing. Several reviews have been published on the identification and use of PR-mAbs (9-11). PR-mAbs have been used to purify many components of the prokaryotic and eukaryotic RNA polymerase II (RNAP II) transcription complex; these PR-mAbs are listed in Table 1. Several other research groups have identified and used PR-mAbs for immunoaffinity chromatography of other protein complexes; these mAbs are also listed in Table 1.

Table 1. Compilation of PR-mAbs proven useful for immunoaffinity chromatography.

PR-mAb	Antigenic Protein	Epitope Sequence	Complex Purified	Reference
8WG16	rpb1, largest subunit of wheat-germ RNA polymerase II	YSPTSPSYSPTSPS	RNA polymerase II	7, 12
NT73	β′ subunit of E. coli RNA polymerase	SLAELLNAGLGGS	Bacterial RNA polymerase holoenzymes	8, 13
8RB13	β subunit of E. coli RNA polymerase	Unknown	Bacterial RNA polymerase core enzyme	14
4RA2	α subunit of E. coli RNA polymerase	a. a. 209-329	Bacterial RNA polymerase core enzyme; bacterial RNA polymerase holoenzyme	15, 16
IIB8	human transcription factor IIB (TFIIB)	TKDPSRVG		17, 18
1RAP1	human 30 kDa RNA-polymerase-associated protein (RAP30)	a. a. 1-118		19
1TBP22	human TATA-binding protein (TBP)	a. a. 55-99	TBP-containing complexes	20
Unnamed	human DNA polymerase δ	Unknown	2 subunit enzyme	21
301β	toluene dioxygenase	Unknown	α2β2 iron-sulfur protein	22
anti-Set1p	Set 1	Unknown	histone methylation complex	23

Occasionally, a mAb appears to be a PR-mAb in the ELISA-elution assay, but it is not useful for immunoaffinity chromatography because it is not able to pull the antigen out of solution (immunoprecipitate). An example of this is shown in Figure 1. mAb IIB14 appears to be polyol responsive in the ELISA-elution assay (Figure 1B) but does not bind TFIIB in solution (data not shown). This "false-positive" result can probably be attributed to the fact that the antigen can become distorted when it is immobilized on the ELISA plate, exposing epitopes that are not accessible when the protein is in solution. mAb IIB14 maps to the C-terminal domain of TFIIB (Figure 1C), which possesses two repeats of the cyclin fold (26) and is highly resistant to trypsin digestion (27, 28). Thus, this is a very compact domain in the native structure. A presumptive PR-mAb, identified by the ELISA-elution assay, that removes the antigen from solution, but does not release the antigen in the presence of salt and polyol, has not been observed.

For more detailed protocol, the reader is referred to two step-by-step reviews (9, 10). As noted in Table 1, several groups have developed purification systems for protein complexes, based on a PR-mAb immunoaffinity chromatography step (21-23). It should be noted that the ELISA-elution assay has been used for determining the response of an antigen-antibody interaction to various eluting conditions besides polyol/salt (29, 30). In addition, the ELISA-elution assay can be used to determine the elution of an antibody from an antigen by the use of a peptide containing the epitope for the mAb (7), and thus could be used to screen peptides for reactivity with the mAb and, consequently, to map the epitope of the antibody.

Challenge of Protein Complexes

RNA polymerases from some bacteriophage are single-subunit enzymes. However, RNA polymerases from bacteria and eukaryotes are multi-subunit enzymes, ranging from 4 to 15 subunits. There are many similarities between the structure and functions of RNA polymerases from bacterial and eukaryotic cells (31). In eukaryotic cells, there are three different nuclear RNA polymerases. In general, RNA polymerase I (RNAP I) transcribes ribosomal RNA, RNA polymerase II (RNAP II) transcribes messenger RNA, and RNA polymerase III (RNAP III) transcribes transfer RNA and other small RNAs. The eukaryotic RNA polymerases have similar structures containing at least 12 subunits, ranging in size from 10 to 220 kDa, with some of the small subunits represented in all three polymerases [for review, see (32)]. Immunoaffinity purification of RNA polymerases has been a reasonably stringent test for the power of immunoaffinity purification using PR-mAbs because of the lability of these complexes. The SDS-PAGE gel in Figure 2 shows the subunit structure of RNA polymerase isolated from E. coli and RNAP II isolated from yeast.

The Use of PR-mAbs that React with Other Species

The use of epitope tags to immunoaffinity purify proteins will be discussed below. In some systems (such as yeast) it has become routine to engineer an epitope tag into a protein of interest, but it is not always possible to use this approach. PR-mAbs that show broad cross-reactivity to RNA polymerase among species have been extremely useful tools. It was fortuitous that the first

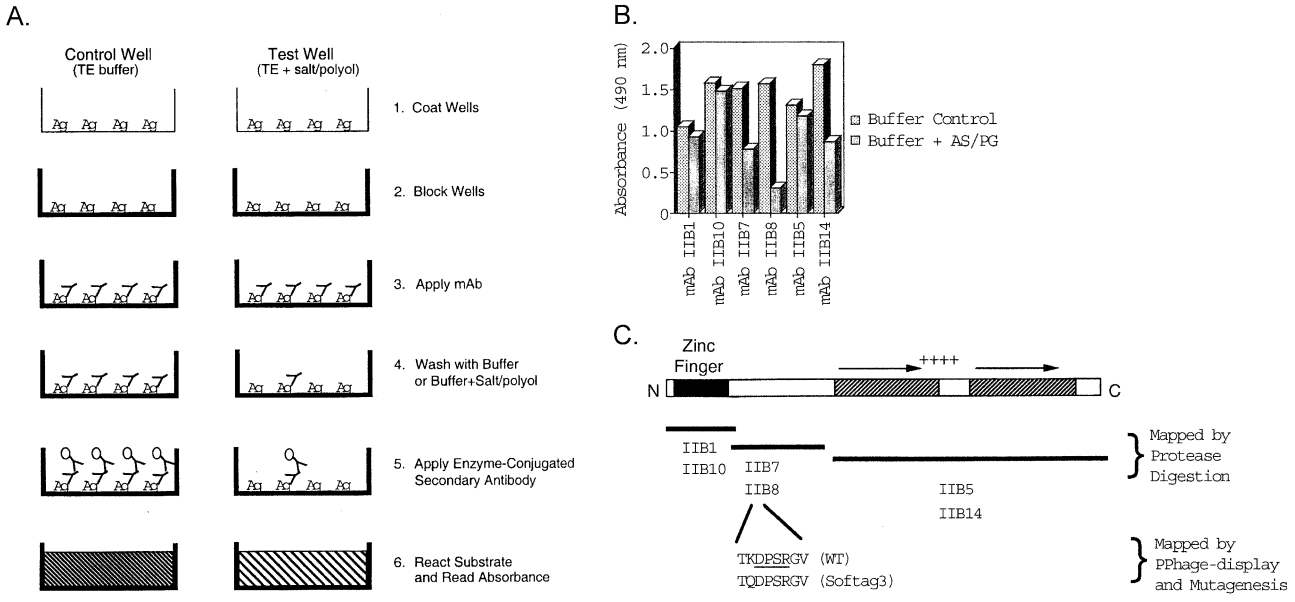


Figure 1. ELISA-elution assay for the identification of PR-mAbs. (A) Schematic of the ELISA-elution assay. Step 1: The wells of a microtiter plate are coated with purified antigen (50–100 ng/well). Step 2: The wells are blocked with a solution of 1% nonfat dry milk. Step 3: The mAb is applied, allowed to react, and then washed off (wash not shown). Step 4: Buffer or buffer containing salt and polyol is added, reacted for 20 min, and washed off (wash not shown). Step 5: The enzyme-conjugated secondary antibody is added, allowed to react, and washed off (wash step not shown). Step 6: Substrate is added to the wells, allowed to react, quenched, and read on a microplate reader. (B) Histogram showing the response of six mAbs that react with TFIIB. The wells were coated with human TFIIB (50 ng/well) and the elution was TE (50 mM Tris-HCl, 0.1 mM EDTA, pH7.9) buffer or TE buffer containing 0.75 M ammonium sulfate and 40% propylene glycol. (C) Map of human TFIIB showing the relative locations of the different mAbs as determined by protease digestion. The epitope for mAb IIB8 was determined by phage display, followed by site-directed mutagenesis mapping; the most critical residues are underlined. The mutation in the epitope sequence that is more sensitive to the salt/polyol elution, as described in the text, is designated Softag3. mAb IIB8 and mAb IIB7 are useful PR-mAbs and react with the same epitope. mAb IIB14 appears to be a PR-mAb (B) but does not immunoprecipitate TFIIB from solution; therefore, mAb IIB14 is a "false-positive," as described in the text.

successful PR-mAb (mAb 8WG16) reacted with a highly conserved domain of the largest subunit of RNAP II. This allowed the mAb to be used to purify RNAP II from calf thymus (7), yeast (12), and HeLa cells (33), in addition to wheat germ RNAP II, the original immunogen (7).

The PR-mAb that was isolated for the purification of E. coli RNAP (NT73) worked fabulously for E. coli (8); however, it did not react with RNAP from many bacteria outside the enteric group (Table 2). Later, a fusion was performed using splenocytes from a mouse that had been immunized with E. coli core RNA polymerase. The fusion was specifically screened for a PR-mAb that

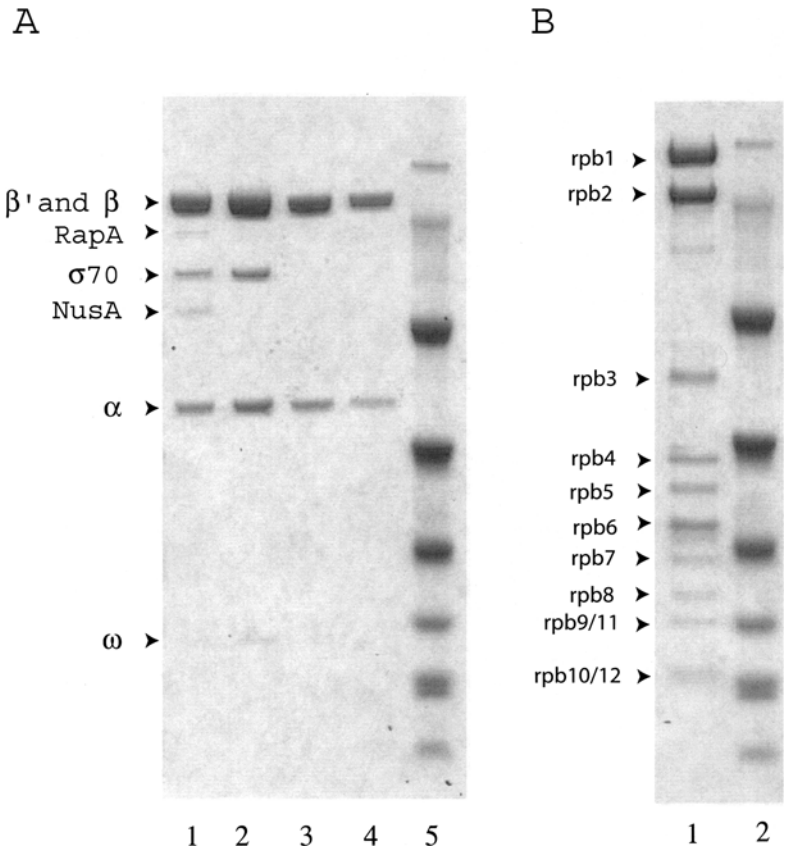


Figure 2. SDS-PAGE of RNA polymerases purified by immunoaffinity chromatography. Proteins were separated on a 4-12% NuPAGE gel (Invitrogen, Carlsbad, CA) in MES buffer, and the gel was stained with GelCode (Pierce, Rockford, IL). Markers were the prestained Multimark (Invitrogen). (A) E. coli RNA polymerase from the mAb NT73 immunoaffinity column (lane 1), σ70-holoenzyme (lane 2), and core RNA polymerase (lane 3) after separation on MonoQ (Amersham-Pharmacia). Lane 4 contains the core RNA polymerase loaded at 25% the load of lane 3 to show the separation of the β′ and β subunits. (B) Yeast RNA polymerase purified from wild-type yeast by immunoaffinity chromatography on 8WG16. The subunit designations are on the left. rpb9 and rpb11 co-migrate on this gel. rpb10 and rpb12 also co-migrate. rpb6 appears as a dimer because some subunits are phosphorylated.

Table 2. Cross-reactivity with other bacterial species of PR-mAb derived by immunization with E. coli RNA polymerase.

Bacterium	Strain Designation	NT73 (β′ subunit)	8RB13 (β subunit)	4RA2 (α subunit)	3RD3 (σ70 subunit)
Escherichia coli	MG1655	+	+	+	+
Klebsiella pneumonia	ATCC13883	+	+	+	+
Salmonella typhimurium	ATCC14028	+	NT	+	+
Shigella boydii	ATCC9207	+	+	+	NT
Serratia marcescens	ATCC13880	+	NT	+	+
Vibrio parahemolyticus	ATCC17802	+	NT	+	+
Vibro fischeri	ATCC7744	+	NT	NT	NT
Pseudomonas aeruginosa	ATCC10145	–	+	+	+
Pseudomonas putida	ATCC12633	NT[g]	+	+	NT
Rhodobacterium sphaeroides[a]	2.4.1	–	–	+	+
Agrobacterium tumefaciens	ATCC15955	–	+	+	–
Bordetella pertussis[b]	BP593	–	NT	+	+
Streptococcus faecalis	ATCC19433	–	NT	NT	+
Streptococcus pneumoniae[c]		NT	–	–	NT
Bacillus subtilis	ATCC6051	–	+	–	+
Arthrobacter globiformis	ATCC8010	–	NT	–	–
Borrelia burgdorferi[d]		–	NT	+	+
Streptomyces coelicolor[e]	A3(2)	NT	+	NT	NT
Anabena sp.[f]	7120	NT	+	NT	+

[a]Whole cell extract of R. sphaeroides obtained from Tim Donohue (University of Wisconsin-Madison).
[b]Whole cell extract of B. pertussis obtained from Philip Boucher (U. S. Food and Drug Administration).
[c]Whole cell extract of S. pneumoniae obtained from Elizabeth Campbell (Rockefeller University).
[d]Whole cell extract of B. burgdorferi obtained from John Dunn (Brookhaven National Laboratory).
[e]Culture of S. coelicolor obtained from David Hopwood (Norwick, UK).
[f]Purified Anabena sp. RNA polymerase obtained from Robert Haselkorn (University of Chicago).
[g]NT = not tested.

showed broad reactivity by screening the mAbs against Bacillus subtilis core RNA polymerase before it was screened for PR-mAbs by the ELISA-elution. The result was a highly useful PR-mAb (8RB13) that reacts with RNA polymerase from a broad range of bacterial species (Table 2), and has been used to purify RNA polymerase from E. coli, B. subtilis, Streptomyces coelicolor, and Pseudomonas aeruginosa (14). mAb 4RA2 shows broad cross-reactivity, reacting with the α subunit of the RNA polymerase from every Gram-negative bacterium tested. mAb 3RD3 also shows broad cross-reactivity among the σ70 subunits but has not been a particularly useful PR-mAb, as will be discussed below.

IMMUNOAFFINITY PURIFICATION OF RNA POLYMERASES AND ASSOCIATED FACTORS

Prokaryotic RNA Polymerase

In bacteria, RNA polymerase is composed of a core enzyme containing subunits α_2, β, β′, and ω (where the α subunit is a dimer), which is capable of non-specific synthesis of RNA from a DNA template. However, this enzyme has no promoter specificity. When a sigma factor interacts with the core polymerase, the complex is a holoenzyme that has specificity for promoter sequences. In general, bacteria have multiple sigma factors, and hence multiple holoenzymes, each of which recognizes a different DNA sequence in the promoter region of different classes of genes (34-36). E. coli has seven known sigma factors (σ70, σ54, σ32, σS, σF, σE, and σFecI) and the genomic sequence (37) does not predict any additional sigma factors. The genomic sequence of Streptomyces coelicolor predicts 64 sigma factors (38). Therefore, RNA polymerase in most bacteria is actually a heterogeneous population comprised of core RNA polymerase and different "holoenzymes" that contain different sigma factors. Under different growth conditions, the populations of holoenzymes can change in order to turn different classes of genes on or off to facilitate adaptation.

mAb NT73 reacts with the β′ subunit of E. coli RNA polymerase and has been a very useful antibody for purification of this enzyme (8). The epitope for this mAb has been mapped to 13 amino acids at the far C-terminus of this peptide (Table 1); this epitope has been used to develop an epitope tag (13) that will be discussed below. The purification of E. coli RNA polymerase using mAb NT73 results in a mixture of core RNA polymerase and holoenzymes (8) and also contains some RNA polymerase associated proteins, such as RapA and NusA (Figure 2A, lane 1). These proteins and the holoenzymes can be further separated by ion-exchange chromatography (MonoQ). The core fraction and the holoenzyme containing σ70 from the MonoQ column are shown in Figure 2A (lanes 2 and 3). mAbs for all seven sigma factors in E. coli (39) as well as the RNA polymerase-binding NusA anti-termination factor are available. By performing Western blots on the fractions from the MonoQ separation, the fractions containing the core RNA polymerase and the different holoenzymes and the NusA protein can be identified (Figure 3). The ability to follow the immunoaffinity chromatography step with a high-resolution HPLC is a powerful way to characterize different protein complexes.

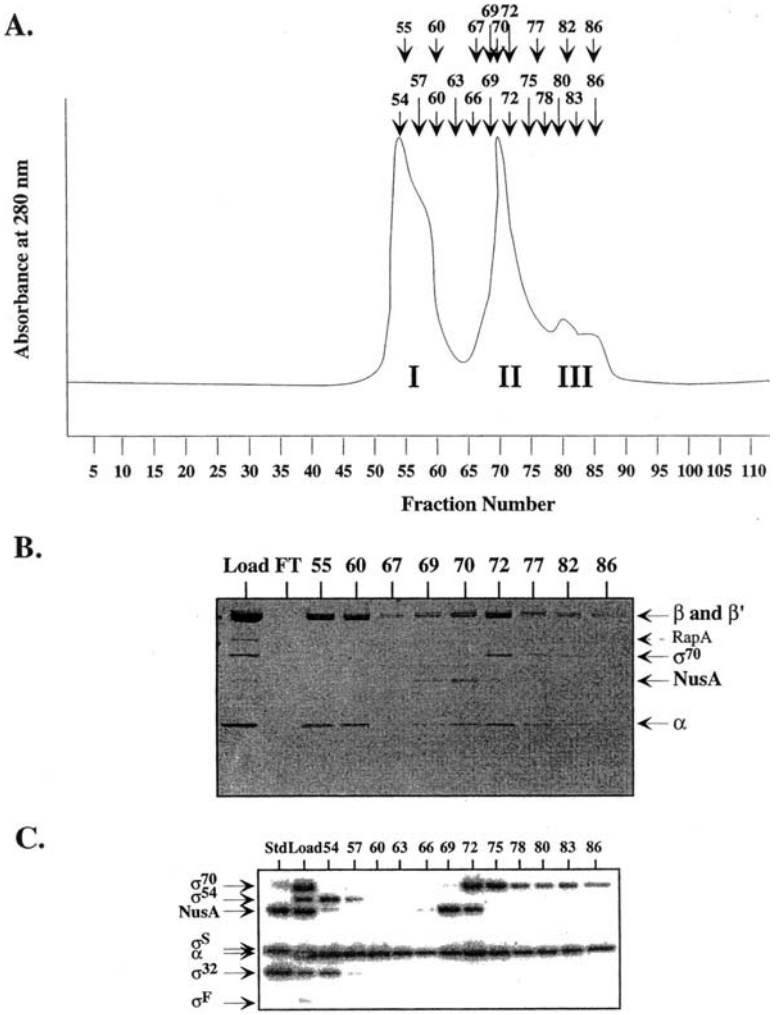


Figure 3. Fractionation of E. coli RNA polymerase (purified by immunoaffinity chromatography on mAb NT73) on a MonoQ ion-exchange column. (A) Chromatogram of the fractionation using a gradient elution of 0.3-0.45 M NaCl over 255 minutes at a flow rate of 0.5 mL/min. The numbered arrows on the top and bottom rows indicate the fractions that were analyzed in panels B and C, respectively. (B) Proteins were separated by SDS-PAGE (4-20% Tris-glycine, Invitrogen) and stained with Coomassie blue. The material that was loaded (Load) onto the column and the flow-through (FT) fractions were also electrophoresed. (C) Western blot containing the indicated fractions. The gel was a 8-12% Tris-glycine gel (Invitrogen). The mAbs used were 3D3 (σ70), 1NA3 (NusA), 4RA2 (α), 1RS1 (σS), 1RF13 (σF), 3RH3 (σ32), and 6RNA3 (σ54). Purified σ70, NusA, σS and σ32 were run as standards in the lane designated Std. σS (38 kDa) and α (40 kDa) co-migrate on this gel.

The interaction between the core RNA polymerase and the sigma subunit has been studied most extensively with σ70, the major, or housekeeping, sigma factor in E. coli. Although there are multiple interactions between the core RNA polymerase subunit and the sigma factor (40), a major binding site for σ70 has been identified in the N-terminus of the β′ subunit (41). In fact, all of the sigma factors in E. coli can bind to this region (42). The recently published crystal structures of bacterial RNA polymerases (43, 44) confirm that this is a major interaction site and indicate that the far C-terminal domain of β′ is not in contact with the sigma subunit, and seems to be accessible to the antibody. Thus, it is not surprising that this heterogeneous population would be isolated by immunoaffinity chromatography using mAb NT73.

Chromatography on PR-mAb 8RB13 gives a different result. This mAb recognizes an unmapped epitope on the β subunit of E. coli RNA polymerase. Surprisingly, this mAb isolates predominantly the core RNA polymerase (14). The epitope for this mAb is accessible in the core version of the enzyme, but probably becomes inaccessible to the mAb when it is contained in the holoenzyme form. In many cases it is useful to isolate core enzyme to which can be added a known or suspected sigma factor or mutant sigma factor. Because of the broad cross-reactivity of this mAb, the purification of core RNA polymerase can be accomplished easily from many species. All of the bacterial RNA core polymerases that have been isolated using 8RB13 have the typical $\alpha_2\beta\beta'\omega$ structure and are catalytically active.

Chromatography on PR-mAb 4RA2 results in yet another interesting result. This mAb reacts with an unmapped epitope in the C-terminus of the α subunit and has quite broad cross-reactivity. When 4RA2 is used to purify Rhodobacterium sphaeroides RNA polymerase, the immunoaffinity chromatography yields core RNA polymerase (15). However, the holoenzyme containing the major sigma factor is isolated when it is used to purify Bordetella pertussis RNA polymerase (16) or E. coli and P. aeruginosa RNA polymerase (N.E. Thompson, unpublished data). This indicates that either the C-terminal domain of the α subunit is positioned differently in some bacteria, or the sigma subunits are more easily dissociated in some bacteria.

mAb 3RD3 was also isolated by screening the hybridomas for reactivity with B. subtilis major sigma factor and is widely cross-reactive with the major sigma factor from other bacterial species (Table 2). This mAb maps to amino acids 475-528 of E. coli σ70 (45). This region of σ70 does not seem to be accessible in solution because mAb 3RD3 is not efficient at isolating either σ70 or the σ70-containing holoenzyme (N.E. Thompson, unpublished data).

Eukaryotic RNAP II

Like the bacterial RNA polymerases, eukaryotic RNA polymerases have catalytic activity, but cannot specifically identify a promoter. However, no distinguishable sigma factor has yet been identified in the eukaryotic nuclear RNA polymerases. Each eukaryotic polymerase has a set of factors that interact with the polymerase in the promoter region. Most of these factors are actually multi-protein

complexes themselves. In addition, the TATA-binding protein (TBP) forms different complexes that interact with the three different RNA polymerases. RNAP II usually requires the general transcription factors TFIIA, TFIIB, TFIID (containing TBP), TFIIE, TFIIF, and TFIIH. With the exception of TFIIB, all of the RNAP II general transcription factors are multi-subunit complexes (46). In addition, even more factors are required for activation of transcription from most promoters.

As mentioned above, PR-mAb 8WG16 reacts with the highly conserved C-terminal heptapeptide repeat on the largest subunit of eukaryotic RNAP II, and it is broadly cross-reacting because RNAP II from most eukaryotes have the consensus heptapeptide repeat. Although it is not seen in the crystal structure, the CTD is believed to be an extended, flexible domain when phosphorylated (47). During a conventional purification of RNAP II, this C-terminal domain (CTD) can be cleaved off by proteases present in the cell extract. Protease inhibitors reduce but do not eliminate the cleavage. The result from a conventional purification is a mixture of enzyme molecules that contain and lack the CTD. In the most simple immunoaffinity purification case, mAb 8WG16 binds to the CTD, and the resulting elution with polyol and salt isolates the multi-subunit enzyme containing the CTD. This enzyme is pure enough for preliminary 2-D crystal formation on lipid bilayers (12). With a little manipulation of the system, yeast RNAP II is essentially pure (Figure 2B), and this procedure has been used to purify enough enzyme for 3-D crystallographic studies (48, 49). Two subunits of the yeast RNAP II (rpb4 and rpb7) form a subassembly that dissociates easily, and is not represented stoichiometrically in the naturally isolated polymerase. Deletion of the nonessential rpb4 subunit gene results in a conditionally lethal yeast strain from which a 10-subunit RNAP II can be isolated. Consequently, the crystal structure of the 10-subunit enzyme (48) was determined before the 12-subunit enzyme (49, 50).

The CTD contains the heptapeptide YSPTSPS repeated up to 52 times. This sequence is a target for multiple modifications, particularly phosphorylations, which has been studied extensively [see (51)]. The sequence is often designated $Y_1S_2P_3T_4S_5P_6S_7$ in order to specifically identify the different serines that are phosphorylated by different kinases during different points in the transcription cycle. The CTD performs many functions in the transcription cycle, most of which also are dependent upon the interaction of the CTD with other proteins [see (52, 53)], and some of these interactions require that the CTD be specifically phosphorylated.

Early in the attempts to identify fractions that can mediate activated transcription, it was discovered that two forms of RNA polymerase exist in yeast extracts that differ greatly in size. A larger form could mediate transcriptional activation, but the smaller 12-subunit form could not. By binding the larger form to the 8WG16 resin, the 12-subunit form bound to the column and a multi-subunit complex was displaced (54). The displaced complex could facilitate activated transcription when added to the 12-subunit form in the presence of the basal transcription factors and an activator protein. Thus, this "negative" chromatographic step resulted in the identification of the "Mediator" complex, a complex that contains a variety of proteins that act as "coactivators" of transcription. The RNAP

II/Mediator complex was predicted by genetic analysis and is also referred to as the RNAP II "holoenzyme" (55).

Studies using 8WG16 for immunoaffinity chromatography purification of yeast RNAP II revealed another characteristic of the RNAP II transcription machinery. Using 8WG16 as an immunosorbent to immobilize yeast RNAP II, it was shown that other RNA polymerase-associated factors can be purified by differential washing of the immobilized complex with increasing salt concentrations (56). The CTD-associated proteins tended to elute with low salt, and the more tightly bound factors eluted at higher salt. Finally, very tightly bound factors could be eluted with the RNAP II during the salt/polyol elution. This approach allows the purification of factors that are associated with RNAP II under different growth conditions.

Several of the human RNAP II general transcription factors have also been purified from bacterial expression systems by the PR-mAb method. These factors include TFIIB (17), RAP30, a component of TFIIF (19), and TBP (20). Preliminary data suggest that the TBP PR-mAb (1TBP22) can also be used to purify large TBP-containing complexes from a HeLa cell nuclear extract (20). In the future, the fraction that comes off the 1TBP22 column will be subjected to a high-resolution HPLC step and analyzed by Western blotting to determine which TBP-containing complexes are present.

The Use of Epitope Tags

As indicated above, the use of an epitope tag for immunoaffinity purification of complexes has been an important development in the protein purification field. An early review of epitope tags is available (57). This section is not intended to be a comprehensive review of all of the possible tags for the purification of all possible proteins. However, it is interesting to note how some large transcription complexes can be studied by this method. The major limitation for this method is that the system must be able to accommodate genetic manipulation at the molecular level. This works well for bacteria, yeast, and some other model systems. In mammalian systems, the construct containing the tag is usually introduced by a retroviral vector, a technique that requires special considerations.

Two commonly used purification tags, which are more appropriately termed "affinity-purification" tags, are the hexahistidine (His6) tag and the glutathione-S-transferase tag (GST). These tags have been used in assays to either identify binding partners (by so-called "pull-down" assays) or for larger-scale purifications. The His6 tag suffers from the following: a degree of non-specificity due to the presence of other Ni+2-binding proteins in cell extracts and the fact that these columns must be run at high salt (0.5 M NaCl) to reduce ion-exchange effects (which can dissociate some complexes); elution from the Ni+2 column is with high concentration of imidazole, which can displace some metal ions. The GST tag suffers from the large size of the tag (26 kDa), which might interfere with the binding properties of the tagged protein. In addition, GST dimerizes, a condition that can result in artifactual results. Several other affinity purification tags have been used and are listed in Table 3. Large affinity tags are generally

Table 3. Commonly used nonantibody-based affinity purification tags.

Tag	Size (aa or kDa)	Affinity Ligand	Elution Conditions
His6	6 aa	Ni+2, other	imidazole
GST[a]	26 kDa	GSH[e]	reduced glutathione
S-tag	15 aa	S-protein	protease, chaotropic salt, low pH
Maltose B	40 kDa	Amylose	maltose
CBD[b]	11 kDa	Cellulose	ethylene glycol
Protein A	7-31 kDa	IgG[f]	low pH
Avidin, Streptavidin		Biotin	biotin, low pH
Chitin BP[c]	56 kDa	Chitin	
TAP-tag[d]	22 kDa	IgG, calmodulin	protease cleavage, EGTA

[a]GST = glutathione-S-transferase
[b]CBD = cellulose-binding domain
[c]BP = binding protein
[d]TAP = tandem affinity purification
[e]GSH = glutathione
[f]IgG = immunoglobulin G

removed from the purified protein by a specific protease whose cleavage sequence has been engineered between the tag and the protein of interest.

An epitope tag system for purification of proteins was first introduced by Field et al. (58), using a previously described epitope sequence (59) based upon immunological studies of the influenza virus hemagglutinin (HA). This work was seminal for several reasons. It utilized a short amino acid sequence (YPYDVPYA), a mAb was available for it (mAb 12CA5), a peptide was used to elute the tagged-protein from the mAb, and the result was an intact complex with biological activity. Within a year, this system was used by Kolodziej and Young (60) to show that tagged rpb3 could be used to purify yeast RNAP II on a mAb 12CA5 column, which confirmed the subunit structure of the enzyme. Recently, investigators have used two different tags, such as FLAG and HA tags on two different RNA polymerase subunits (61) or FLAG and His6 tags on the same subunit (62). Epitope tags that have been used to purify transcriptional machinery complexes and tags that react with PR-mAbs are listed in Table 4.

The TATA-binding protein (TBP) is required for transcription from all three eukaryotic nuclear RNA polymerases, and it is present in a complex with different TBP associated factors (TAFs) for the different polymerases. The TBP-containing complex required for transcription from RNAP II promoters is TFIID [reviewed in (64, 65)]. Zhou et al. (66) investigated this complex by using

Table 4. Epitope tags for immunoaffinity purification.

Epitope Tag	mAb	Sequence	Elution Conditions	References or Source
HA tag	12CA5	YPYDVPYA	peptide	58
myc tag	9E10	EQKLISEEDL	pH 3.5	63
FLAG tag	M1, M2	DYKDDDDK	peptide, EDTA	Sigma-Aldrich
T7 tag	T7-tag mAb	MASMTGGQQMG	low pH	Novagen
Softag1	NT73	SLAELLNAGLGGS	salt/polyol	8, 13
Softag2	8WG16	YSPTSPSYSPTSPS	peptide, salt/polyol, polyol	7, 9, 12, 33
Softag3[a]	IIB8	TQDPSRVG	salt/polyol	17, 18

[a]The wild-type epitope is TKDPSRVG; the mutant epitope is designated as "Softag3."

a HA-tagged TBP stably transfected into HeLa cells, and TFIID was isolated. A truncated TBP containing only the conserved C-terminal domain of human TBP was also tagged. Chromatography on the 12CA5 mAb revealed that all of the TFIID proteins were isolated when only the C-terminal domain was present. This indicated that the RNAP II-specific TAF proteins all associated with TFIID by interaction between either the C-terminal domain of TBP or with other TAFs.

Other large transcription complexes that have been studied by epitope tags are human TFIIH and human Mediator. The general transcription factor TFIIH is a multi-subunit complex that contains several catalytic activities, including ATPase, DNA helicase, DNA repair, and CTD kinase activities. The human TFIIH contains nine subunits. Jawhari et al. (67) have purified subcomplexes of human TFIIH by expressing FLAG-tagged subunits along with non-tagged subunits from baculovirus vectors in insect cells. In addition, they were able to isolate the entire nine-subunit complex by using a double tag system, tagging one subunit with the FLAG epitope and another with a His6 tag.

The Mediator, as mentioned above, is a large protein complex that contains co-activator activity. One of the major problems in studying the Mediator complex is that different preparations contain different subsets of proteins. Sato et al. (68) tagged six different human Mediator subunits with the FLAG epitope and expressed them independently in HeLa cells. After purification on the mAb column, the isolated complexes were analyzed by multidimensional protein identification technology (MudPIT). In this way, they established a consensus of human Mediator subunits.

In one study, TFIID, TFIIH, RNAP II, and Mediator were all isolated from HeLa cell nuclei by immunoaffinity chromatography using a FLAG-tagged subunit of each complex (69). By reconstituting the other general transcription

factors from recombinant subunits, the investigators were able to study the effect of different Mediator complexes on different steps in transcription.

Generally, when using an epitope-tag for immunoaffinity purification, the antigen is eluted with a synthetic peptide containing the sequence of the epitope. If necessary, the peptide is then removed by some type of dialysis or size exclusion step. At least in the case of 8WG16, the target protein can also be eluted from a PR-mAb with a peptide (7, 33). However, the polyol/salt combination is very inexpensive and, in most cases, protein-stabilizing. During the development of Softag3, a mutation was inadvertently introduced into the epitope sequence. This tag bound to the antibody as well as the wild-type sequence, but was more sensitive to elution at lower salt and polyol (18). This indicates that the PR-mAb system might be somewhat "tunable" to different levels of salt and polyol by the use of mutant epitopes. The mutation is shown in Figure 1C, and does not change any of the residues that are critical for mAb IIB8 binding (18).

Although the PR-mAb epitopes have only been used to tag GFP and have only purified this protein from bacterial expression systems, it is likely that these will be extremely useful for purifying complexes. NT73, in particular, is a high affinity antibody that elutes at relatively low polyol and salt (8); this PR-mAb has been tested with several eukaryotic cell extracts and does not show significant cross-reactivity with any eukaryotic protein (13). Therefore, Softag1 should be very useful as a tag in a eukaryotic system, and Softag3 should be useful in prokaryotic systems.

Single-Chain Antibodies

Unfortunately, the power of immunoaffinity chromatography is offset by the expense of the mAb reagents. It would be ideal to produce PR-mAbs in large quantities in bacteria or yeast. However, in our hands, converting some PR-mAbs into single-chain variable fragments (scFv) has not been very encouraging. mAb NT73 was selected to begin this endeavor because its initial high affinity would allow some affinity loss and still be a useful reagent. mRNA isolated from the hybridoma cell line was used in the RPAS Mouse ScFv Module (Amersham-Pharmacia, Piscataway, NJ) to generate the scFv. The scFv was selected by its reactivity with E. coli core RNA polymerase, and detected with anti-E-tag antibody that reacts with an epitope tag (E-tag) fused to the C-terminus of the scFv. The scFv was sequenced and recloned into expression vectors for expression in E. coli. The ELISA-elution data in Figure 4 show the responses of the single chain antibody compared to the response of the native mAb to salt, polyol, and the combination of salt and polyol. The scFv was responsive to salt and polyol, but also was responsive to just the polyol. However, the scFv has considerably lower affinity for the antigen than mAb NT73.

The scFv has been expressed from several bacterial vectors in different E. coli hosts. The best expression is from a pET22b vector (Novagen, Madison, WI) using the Rosetta (DE3)pLysS strain (Novagen) which supplies codons rarely expressed in E. coli. However, this scFv becomes insoluble when expressed in E. coli. To date, the scFv has not been successfully refolded in high

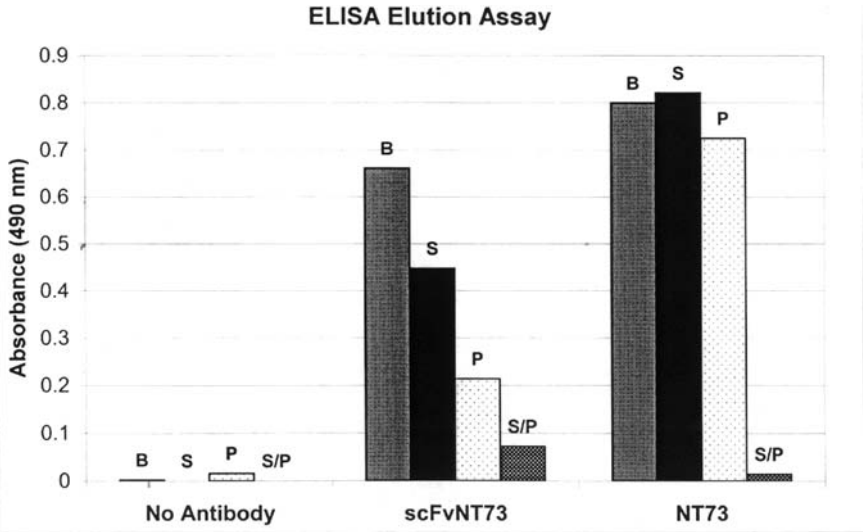


Figure 4. ELISA-elution assay showing the polyol-responsiveness of the NT73 scFv and mAb NT73. The wells were coated with core RNA polymerase (100 ng/well) and eluted with TE buffer (B) or TE buffer containing 0.75 M NaCl (S), 40% propylene glycol (P), 0.75 M NaCl and 40% propylene glycol (S/P). Reactions were run in duplicate.

enough yields to produce enough antibody for an immunoaffinity column. Perhaps a more productive approach would be to prepare a scFv library from the mRNA of immune splenocytes and screen for polyol-responsiveness by an ELISA-elution assay.

CONCLUSIONS

Immunoaffinity chromatography is a very useful tool for the purification of protein complexes. The use of mAbs has resulted in two methods to purify complexes gently and specifically.

The first is to use a PR-mAb that reacts with an epitope contained in a protein in the complex. It is helpful if this PR-mAb is cross-reactive with the homologous protein in different species. It is also possible to screen existing collections of mAbs that react with a protein of interest for polyol-responsiveness. One can also perform a fusion with the intention of screening specifically for PR-mAbs even before the hybridomas are cloned. Secondly, if the biological system of interest can be genetically manipulated, it is possible to incorporate an epitope tag onto one of the proteins in the complex. The complex can be recovered by eluting it with either a peptide homologous to the epitope or, if the mAb is polyol-responsive, by polyol/salt.

Some features of immunoaffinity chromatography should be considered when the target is part of a larger protein complex. First, the epitope might not be accessible when it is contained in the complex. Second, as mentioned above, depending upon where the epitope is located, some complex disruption might occur due to the high affinity reaction of the epitope with the antibody.

These gentle immunoaffinity purification methods are rapid and specific, and the structural integrity and biological activity of the complex are usually retained. In fact, in many cases, the complex is suitable for crystallization trials, although it might be advantageous to follow the immunoaffinity chromatography step with some type of polishing step, such as high-resolution ion exchange or a size-exclusion chromatography step.

ACKNOWLEDGMENTS

We thank Katherine Foley for help with the hybridoma work. This work was supported by Grants GM28575, CA60896, and CA14520 from the National Institutes of Health. In compliance with the University of Wisconsin–Madison's Conflict of Interest Committee policy, the authors (N.E.T. and R.R.B.) acknowledge financial interest in NeoClone LLC (Madison, WI), which markets mAbs 8WG16, NT73, 8RB13, 4RA2, and IIB8.

REFERENCES

1 Owens , R.J. and Young, R.J. (1994) J. Immunol. Methods 168, 149-165.
2 Peterson, N.C. (1996) Lab. Animal Sci. 46, 8-14.
3 Little, M., Kipriyanov, S.M., Gall, F.L. and Moldenhauer, G. (2000) Immunol. Today 21, 364-370.
4 Nelson, P.N., Reynolds, G.M., Waldron, E.E., Ward, E., Giannopoulos, D. and Murray, P.G. (2000) J. Clin. Pathol. Mol. Pathol. 53, 111-117.
5 Harlow, E., Whyte, P., Franza, Jr., B.R. and Schley, C. (1986) Mol. Cell. Biol. 6, 1579-1589.
6 Hager, D.A. and Burgess, R.R. (1980) Anal. Biochem. 109, 76-86.
7 Thompson, N.E., Aronson, D.B. and Burgess, R. R. (1990) J. Biol. Chem. 265, 7069-7077.
8 Thompson, N.E, Hager, H.A. and Burgess, R.R. (1992) Biochemistry 31, 7003-7008.
9 Thompson, N.E. and Burgess, R.R. (1996) Methods Enzymol. 274, 513-526.
10 Thompson, N.E. and Burgess, R.R. (2001) Curr. Protocol Molec. Biol., SecVI, 11.18.1-11.18.9.
11 Burgess, R. R., and Thompson, N. E. (2002) Curr. Opin. Biotechnol. 13, 304-308.
12 Edwards, A.M., Darst, S.A., Feaver, W.J., Thompson, N.E., Burgess, R.R. and Kornberg, R.D. (1990) Proc. Nat. Acad. Sci. U.S.A. 87, 2122-2126.
13 Thompson, N.E., Arthur, T.M. and Burgess, R.R. (2003) Anal. Biochem. 323, 171-179.
14 Bergendahl, V., Thompson, N.E., Foley, K.M., Olson, B.M. and Burgess, R.R. (2003) Protein Expr. Purif. 31, 155-160.
15 Anthony, J.R., Green, H.A. and Donohue, T.J. (2003) Methods Enzymol. 370, 54-65.
16 Boucher, P.E., Murakami, K., Ishihama, A. and Stibitz, S. (1997) J. Bacteriol. 179, 1755-1763.

17 Thompson, N.E. and Burgess, R.R. (1994) Protein Expr. Purif. 5, 468-475.
18 Duellman, S.J., Thompson, N.E. and Burgess, R.R. (2004) Protein Expr. Purif. 35, 147-155.
19 Thompson N.E. and Burgess, R.R. (1999) Protein Expr. Purif. 17, 260-266.
20 Thompson, N.E., Foley, K.M. and Burgess, R.R. (2004) Protein Expr. Purif. 36, 186-197.
21 Jiang, Y., Zhang, S.J., Wu, S.M. and Lee, M.Y. (1995) Arch. Biochem. Biophys. 320, 297-304.
22 Lynch, N.A., Jiang, H. and Gibson, D.T. (1996) Appl. Environ. Microbiol. 62, 2133-2137.
23 Nagy, P.L., Griesenbeck, J., Kornberg, R.D. and Cleary, M.L. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 90-94.
24 Kohler, G. and Milstein, C. (1975) Nature 256, 495-497.
25 Largaespada, D.A., Jackson, M.W., Thompson, N.E., Kaehler, D.A., Bryd, F.G. and Mushinski, J.F. (1996) J. Immunol. Meth. 197, 85-95.
26 Bagby, S., Kim, S., Maldonado, E., Tong, K.I., Reinberg, D. and Ikura, M. (1995) Cell 82, 857-867.
27 Malik S., Lee, D.K. and Roeder, R.G. (1993) Mol. Cell. Biol. 13, 6253-6259.
28 Thompson, N.E., Strasheim, L.A., Nolan, K.M. and Burgess, R.R. (1995) J. Biol. Chem. 270, 4735-4740.
29 Yang, J., Moyana, T. and Xiang, J. (1999) J. Chromatog. B 73l, 299-308.
30 Kummer, A. and Li-Chan, E.C. (1998) J. Immunol. Methods 21, 125-137.
31 Ebright, R.H. (2000) J. Mol. Biol. 304, 687-698.
32 Archambault, J. and Friesen, J.D. (1993) Microbiol. Rev. 57, 703-724.
33 Maldonado, E., Drapkin, R. and Reinberg, D. (1996) Methods Enzymol. 274, 72-100.
34 Ishihama, A. (2000) Annu. Rev. Microbiol. 54, 499-518.
35 Burgess, R.R. (2001) in Encyclopedia of Genetics (S. Brenner and J.H. Miller, eds.), Academic Press, New York, NY. pp. 1831-1834.
36 Paget, M.S. and Helmann, J.D. (2003) Genome Biol. 4, 203-208.
37 Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) Science 277, 1453-1474.
38 Hahn, M.-Y., Bae, J.-B., Park, J.-H. and Roe, J.-H. (2003) Methods Enzymol. 370, 73-94.
39 Anthony, L.C., Foley, K.M., Thompson, N.E. and Burgess, R.R. (2003) Methods Enzymol. 370, 181-192.
40 Sharp, M.M., Chan, C.L., Lu, C.Z., Marr, M.T., Nechaev, S., Merritt, E.W., Severinov, K., Roberts, J.W. and Gross, C.A. (1999) Genes Dev. 13, 3015-3026.
41 Arthur, T.M. and Burgess, R.R. (1998) J. Biol. Chem. 273, 31381-31387.
42 Burgess, R.R., Arthur, T.M. and Pietz, B.C. (1998) Cold Spring Harb. Symp. Quant. Biol. 63, 277-287.

43 Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. and Yokoyama, S. (2002) Nature 417, 712-719.
44 Murakami, K.S., Masuda, S. and Darst, S.A. (2002) Science 296, 1280-1284.
45 Rao, L., Jones, D.P., Nguyen, L.H., McMahan, S.A. and Burgess, R.R. (1996) Anal. Biochem. 241, 173-179.
46 Woychik, N.A. and Hampsey, M. (2002) Cell 108, 453-463.
47 Kamenski, T., Heilmeier, S., Meinhart, A. and Cramer, P. (2004) Mol. Cell 15, 399-407.
48 Cramer, P.A., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R. and Kornberg, R.D. (2000) Science 288, 640-649.
49 Armache, K.J., Kettenberger, H. and Cramer, P. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 6964-6968.
50 Bushnell, D.A. and Kornberg, R.D. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 6969-6973.
51 Orphanides, G. and Reinberg, D. (2002) Cell 108, 439-451.
52 Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Cell 108, 501-512.
53 Zorio, D.A.R. and Bentley, D.L. (2004) Experiment. Cell. Res. 296, 91-97.
54 Kim, Y.-J., Bjorklund, S, Li, Y., Sayre, M.H. and Kornberg, R.D. (1994) Cell 77, 599-608.
55 Koleske, A. and Young, R.A. (1994) Nature 368, 883-894.
56 Wade, P.A., Werel, W., Fentzke, R.C., Thompson, N.E., Leykam, J.F., Burgess, R.R., Jaehning, J.A. and Burton, Z.F. (1996) Protein. Expr. Purif. 8, 85-90.
57 Jarvick, J.W. and Telmer, C.A. (1998) Annu. Rev. Genet. 32, 601-661.
58 Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R.A. and Wigler, M. (1988) Mol. Cell. Biol. 8, 2159-2165.
59 Munro, S. and Pelham, H.R.B. (1984) EMBO J. 3, 3087-3093.
60 Kolodziej, P. and Young, R.A, (1989) Mol. Cell. Biol., 5387-5394.
61 Nakatani, Y. and Ogryzko, V. (2003) Methods Enzymol. 370, 430-444.
62 Hasegawa, J., Endou, M., Narita, T., Yamada, T., Yamaguchi, Y., Wada, T. and Handa, H. (2003) J. Biochem. 133, 133-138.
63 Bill, R.M., Flitsch, S.L. and Bicknell, R. (1997) Biochim. Biophys. Acta 1340, 13-20.
64 Green, M.R. (2000) Trends Biochem. Sci. 25, 59-63.
65 Pugh, B.F. (2000) Gene 255, 1-14.
66 Zhou, Q., Boyer, T.G. and Berk, A.J. (1993) Genes Dev. 7, 180-187.
67 Jawhari, A., Uhring, M., Crucifix, C., Fribourg, S., Schultz, P., Poterszman, A., Egly, J.M. and Moras, D. (2002) Protein Expr. Purif. 24, 513-523.
68 Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybailov, B., Swanson, S.K., Banks, C.A.S., Jin, J., Cai, Y., Washburn, M.P., Conaway, J.W. and Conaway, R.C. (2004) Mol. Cell 14, 685-691.
69 Wu, S.-Y., Zhou, T. and Chiang, C.-M. (2003) Mol. Cell. Biol. 23, 6229-6242.

BIOGENESIS OF IRON-SULFUR CLUSTER PROTEINS IN PLASTIDS

Marinus Pilon, Salah E. Abdel-Ghany, Douglas Van Hoewyk, Hong Ye, and Elizabeth A. H. Pilon-Smits

Biology Department
Colorado State University
Fort Collins, Colorado 80523

SUMMARY

Iron-sulfur (Fe-S) clusters are co-factors of proteins that perform a number of biological roles, including electron transfer, redox and non-redox catalysis, regulation of gene expression, and as sensors within all living organisms, prokaryotes and eukaryotes. These clusters are thought to be among the oldest structures found in biological cells. In chloroplasts, Fe-S clusters play a key role in photosynthetic electron transport as well as nitrogen and sulfur assimilation. The capacity of the Fe atom in Fe-S clusters to take up an electron reversibly provides the required electron carrier capacity in these pathways. Iron and sulfur limitation both affect plant primary production and growth. It has long been known that iron deficiency leads to defects in photosynthesis and bleaching in young leaves, phenomena that are closely linked to a defect in chloroplastic photosystem-I (PSI) accumulation, a major Fe-S containing protein complex in plants. Although the functional importance of Fe-S cluster proteins is evident and isolated chloroplasts have been shown to be able to synthesize their own Fe-S clusters,

Genetic Engineering, Volume 27, Edited by J. K. Setlow

much is yet to be learned about the biosynthesis of Fe-S proteins in plastids. The recent discovery of a NifS-like protein in plastids has hinted to the existence of an assembly machinery related to bacterial Fe-S assembly systems. This chapter aims to summarize what we presently know about the assembly of Fe-S clusters in plants with an emphasis on green plastids.

INTRODUCTION: IRON-SULFUR CLUSTERS, FUNCTIONS, EVOLUTION, AND FORMATION

Many Fe-S clusters are redox active due to the capacity of Fe to take up an electron reversibly. This property is used, for instance, in components of the photosynthetic electron transport chain and in the respiratory electron transport chain of mitochondria and bacteria. Next to these electron transport roles in energy-transducing systems, Fe-S proteins can have redox roles in enzymes, for instance, those involved in nitrogen and sulfur reduction. Fe-S clusters in enzymes can also have a catalytic role other than redox activity, for example, in aconitase. Finally, some Fe-S proteins function in the regulation of gene expression, and as sensors for oxygen and Fe status within all living organisms, prokaryotes, and eukaryotes (1).

The Fe in Fe-S clusters is mostly found as Fe^{3+}, with the possibility for specific Fe atoms to be reduced to Fe^{2+}. Fe-S clusters contain S as sulfide (S^{2-}) (1). In Fe-S proteins, the inorganic and acid labile S of the cluster is typically bonded with iron and it is iron that is chelated directly to the protein side-chain residues. With these general rules for the architecture, many types of cluster are possible in biological systems. Fe-S clusters can differ in the number of Fe and S atoms and the way in which they are chelated to a protein, properties that in turn can affect cluster redox potential and biological function. Furthermore, some clusters contain additional metal ions such as Ni and Mo. The most common types of clusters are 4Fe-4S clusters and the 2Fe-2S clusters. Typically, Fe-S clusters are chelated to the protein by cysteines, but other residues may contribute to chelation of the cluster. For instance, the 2Fe-2S cluster of ferredoxin-type proteins is chelated by four cysteines, with the thiol S of the protein bonding to the Fe atoms, whereas, in the Rieske-type proteins, the 2Fe-2S cluster is coordinated to two protein cysteines and two histidines (2).

The Fe-S clusters are thought to be among the oldest structures found in biological cells (3). Indeed, Fe and S may have been abundantly available in the environment in which life first evolved and the conditions at that time probably favored spontaneous Fe-S cluster formation. This availability as well as the utility of Fe-S clusters in catalysis and electron transfer may have contributed to an early "addiction" of life to Fe, particularly as Fe-S (3). Somewhat ironically, oxygenic photosynthesis, made possible by the use of Fe in electron transport systems, greatly reduced the availability of Fe due to reactivity of oxygen to Fe and the insolubility of the resulting iron oxides. Thus, present-day organisms are in fierce competition for limited available iron, despite the abundance of Fe in the earth's outer crust. Sulfur, perhaps available abundantly in reduced form in the early atmosphere, may have gradually become oxidized and is now available to plants mostly as sulfate.

Although 2Fe-2S and 4Fe-4S clusters have been assembled *in vitro* in some model proteins with ferrous iron and sulfide, it is now clear that the process is not spontaneous *in vivo* and Fe-S assembly proteins have been shown to be required for the biological formation of these [for reviews, see (4, 5)]. We aim to describe what is presently known about the Fe-S cluster assembly pathways in plants, particularly in chloroplasts.

Fe-S Proteins in Plastids

Fe-S clusters are essential components for photosynthesis, the process unique for plants and algae that drives life on earth. The Fe in Fe-S clusters plays a pivotal role in electron transfer from water to NADPH, which is used to reduce CO_2 to form sugars (6), as well as for N and S reduction and assimilation, for example, by nitrite reductase and sulfite reductase (7, 8).

The photosynthetic electron transport chain contains three major complexes, photosystem-II (PS-II), the cytochrome b_6/f complex, and PSI. In addition, the NDH complex, which is similar to the mitochondrial NADH reductase, functions in cyclic electron transport (9). Iron is present in all of these complexes, and as such Fe is the most important redox-active metal ion for photosynthetic electron transport, both quantitatively and qualitatively (10). Iron in PS-II is present in heme and non-heme iron. In the cytochrome b_6/f complex, Fe is present as heme and as 2Fe-2S Rieske-type clusters. PSI contains three 4Fe-4S clusters (11) and the electron carrier ferredoxin (Fd) contains one 2Fe-2S cluster (6). Some other plastidic Fe-S proteins are Tic55, a Rieske-type protein of the plastid protein import machinery (2Fe-2S), ferredoxin-thioredoxin reductase (4Fe-4S), sulfite reductase (4Fe-4S), nitrite reductase (4Fe-4S), and glutamate synthase (3Fe-4S) (2, 12).

The Biogenesis of Fe-S Proteins

In eukaryotes, an Fe-S assembly machinery is present in the mitochondria. Work in yeast suggested that Fe-S cluster formation is the only essential function of mitochondria. For a review see (4). Furthermore, cytosolic Fe-S clusters depend on the mitochondrial Isc machinery involving homologues of the genes encoded by the *nif/isc* clusters of bacteria and an ABC-type transporter in the mitochondrial inner membrane, which may serve to export intermediates in Fe-S assembly (13). A similar mitochondrial machinery may be present in plants (14). Mutations in the Arabidopsis Starik gene encoding a mitochondrial protein that is a functional homologue of the yeast ABC exporter required for cytosolic Fe-S cluster formation produce plants with severe growth and developmental phenotypes (14). More recently, there is also evidence for a cytosolic Fe-S cluster formation machinery in yeast (15)

Most chloroplast proteins are nuclear encoded and synthesized with a cleavable N-terminal transit sequence required for translocation into the organelle (16). Nuclear-encoded chloroplast metallo-proteins like ferredoxin (Fd) acquire their cofactors after import into the organelle (17–20). Indeed, chloroplasts appear to have their own Fe-S biosynthetic machinery: Fe-S cluster assembly in Fd was observed in isolated chloroplasts with cysteine as the sulfur donor,

a reaction that further requires light or ATP and NADPH (21, 22). Fe-S cluster assembly into radiolabeled, freshly imported ferredoxin precursor obtained by *in vitro* translation was demonstrated using isolated intact chloroplasts (17). The reaction proceeds in the absence of cytosol (19). Together, these experiments indicate the presence of an Fe-S cluster formation machinery in chloroplasts.

Following import into the organelle, the maturation of Fe-S proteins depends on a number of processes, described in more detail in the following sections. Firstly, iron must be taken up and mobilized. Secondly, sulfur must be taken up, reduced, and assimilated. Finally, Fe-S clusters must be assembled from available components and inserted into apo-proteins. Because cysteine was identified as a source for Fe-S formation in chloroplasts (21, 22), a protein with cysteine desulfurase activity is likely involved in this process. CpNifS, the first characterized NifS-like protein from plants, is the only plastid protein with this activity that has been identified (23, 24) and a CpNifS-dependent machinery likely is responsible for plastid Fe-S cluster formation.

Iron Uptake and Storage in Plants

A first step required for Fe-S formation is the uptake of iron. Although Fe is abundant in the earth's crust, it is mainly present as insoluble iron-oxide in the soil which is not bioavailable. As a consequence, Fe is one of the three most limiting nutrients to plant growth (10). Plants have developed two strategies to take up iron [for reviews, see (25, 26)]. Grasses secrete phyto-siderophores that complex iron to make it soluble and available for uptake in the root by a specialized transporter (27). Other plants like *Arabidopsis thaliana* use a ferric reductase (28) to reduce Fe(III) to Fe(II), which is more soluble and can be taken up by the IRT transporter at the root surface (29, 30). Plants also make Fe more bioavailable by pumping protons into their rhizosphere using ATPases; these protons can replace Fe and other cations at negatively charged groups on the soil surface.

Much is yet to be learned about how iron is distributed throughout the plant and inside plant cells. Fe may be chelated by nicotianamine or organic acids during long-distance transport. Iron import in the mesophyll cells may involve the activity of a ferric reductase and the action of metal transporter of the NRAMP or YSL families, but the exact mechanism is not yet clear (25). It is estimated that up to 90% of the iron in green tissues is in chloroplasts. Fe(II) transport activity has been identified for chloroplast envelopes (31); however, the molecular machinery involved is not yet identified. In leaf chloroplasts much of the Fe is used in photosynthesis, particularly in PSI, whereas the remaining or excess Fe is chelated and stored by the chloroplast protein ferritin (32). The *Arabidopis* genome encodes four different plastid ferritins, which are differentially expressed. It is very likely that the Fe used for Fe-S clusters is recruited from ferritin-bound stores.

S Assimilation and Reduction

Sulfur (S) is an essential macronutrient for plants, and present at 0.1–1% of plant dry weight depending on the plant family and soil type (10). Sulfur is generally less limiting for plant growth than other macronutrients such as N or P,

but nevertheless positive responses to S fertilization have been reported from many areas in the world including most agricultural areas. Sulfur deficiency manifests itself as chlorosis of younger leaves and stunted growth (10). The role of S in molecules is very diverse; this is because S can exist in multiple oxidation states (+6, +4, 0, −2) with different chemical properties (33). Next to its role in Fe-S clusters, sulfur is an essential element for plant primary metabolism as a structural component of proteins and lipids, antioxidants, regulatory molecules, metal-binding molecules, and co-factors/co-enzymes.

The flow of S in plants can be summarized as follows. Most S is taken up as sulfate, which is first activated and then reduced to sulfite and finally sulfide, which is subsequently incorporated into cysteine. The main form of S in soils and thus the form taken up by plants is sulfate. This is the most oxidized form of S (valence state +6), and the predominant bioavailable form in most soils. The form of S present in biomolecules is mostly reduced S, although S also occurs in its oxidized form in sulfolipids and various sulfated compounds [for a review, see (34)]. Cysteine is the first organic form of S after sulfate reduction.

The assimilation of sulfate into cysteine takes place mainly in the chloroplast. On its way from the soil to the chloroplast, sulfate enters the plant by group 1 high-affinity sulfate transporters in the plasma membrane (35-37). Translocation of sulfate to the shoot by way of the xylem appears to be facilitated by sulfate transporters from groups 4, 3, and 2 in Arabidopsis roots, involved in vacuolar efflux and xylem loading (38, 39). Sulfate is taken up from the xylem into leaf mesophyll cells, perhaps by the combined action of group 2 and 3 sulfate transporters (40, 41). From the cytosol, sulfate is transported to the chloroplasts. There may be an H^+-sulfate co-transporter in the chloroplast envelope, but so far none has been identified.

Sulfate is activated by reaction with ATP to form adenosine-5-phosphosulfate (APS). This reaction is catalyzed by ATP sulfurylase. The predominant isoform of this enzyme is located in the plastids, but there is also a minor cytosolic form; the two isoforms are regulated differently (42). The further reduction of sulfite to sulfide is mediated by sulfite reductase, a plastidic enzyme (43). The six electrons needed for this step are thought to come from ferredoxin (Fd). Sulfide is incorporated into cysteine (Cys) by coupling to O-acetylserine (OAS). This reaction is mediated by the enzyme OAS thiol lyase, also called cysteine synthase; the OAS needed for this reaction is produced by serine acetyltransferase (SAT). Because only plastidic forms of APS reductase and sulfite reductase have been found, reduction of sulfate to sulfide is thought to occur exclusively in plastids. Because of the higher reducing power in the photosynthetic chloroplasts, most of sulfate reduction probably happens in chloroplasts, although non-green plastids also perform sulfate reduction. After formation, Cys is rapidly converted to other compounds in the chloroplast or other compartments. Therefore, the Cys concentration in the cell is quite low (in the micromolar range).

Much of Cys is incorporated into proteins, either in the plastids or in the cytosol. Cysteine residues in proteins often serve an important role in protein structure and function. The structural importance is due to the capacity of two Cys thiol groups to form a disulfide bond, which can contribute to protein tertiary and quaternary structure. In intracellular proteins, thiols are mostly in a reduced

state. The reducing power of these thiol groups can be used to reduce other cell components. For instance, in chloroplasts the redox capability of Cys in thioredoxin is crucial for the regulation of photosynthetic enzymes (12). The thiol group of Cys also has metal-binding properties and is responsible for the metal-binding capacity of many metal-binding proteins including Fe-S clusters but also other proteins such as metallothioneins (44) and metal transporters such as P-type ATPases of which there are eight in Arabidopsis (45).

Cysteine holds a central position in S metabolism and is used for the biosynthesis of a variety of other reduced S compounds including methionine, S-adenosylmethionine (SAM), glutathione (GSH), and phytochelatins (PCs), the coenzymes thiamine, biotin, lipoic acid, and co-enzyme-A, the molybdenum co-factor and Fe-S clusters. About 2% of the organic reduced S in the plant is present in the form of nonprotein thiols, and around 90% of this fraction is glutathione (γ-Glu-Cys-Gly, GSH). Glutathione is synthesized enzymatically in both the plastids and the cytosol (67).

Cysteine can be converted to alanine and sulfide by Cys desulfurases (CysD). These are NifS-like proteins, that is, related in structure to the NifS protein from *Azotobacter vinelandii* (46). In Arabidopsis, one NifS-like enzyme has been reported in plastids (23, 24), whereas a second form may be present in mitochondria (14). CysD enzymes function to provide reduced S for the production of Fe-S clusters (see below) as well as several coenzymes (47).

MICROBIAL IRON-SULFUR CLUSTER BIOSYNTHETIC MACHINERIES

The study of Fe-S assembly has progressed most rapidly in microbial systems and because these studies provided very useful insights into the Fe-S machinery in plastids we provide a brief overview here. Fe-S cluster assembly in microbes can be divided into three steps: mobilization of S from cysteine and Fe from cellular stores, cluster assembly, and finally insertion in apo-proteins (5).

The first Fe-S assembly machinery studied was the *nif* system of *Azotobacter vinelandii*, which is responsible for the formation of Fe-S clusters for nitrogenase, required under nitrogen fixation conditions (46). The *A. vinelandii nif* gene cluster includes a cysteine desulfurase (CysD) encoding gene, *NifS*, as well as the other genes *nifU, nifA, NifV,* and *cysE*, all thought to be involved in Fe-S cluster formation. NifS-like proteins are pyridoxal 5'-phosphate (PLP)-dependent, enzymes that produce elemental sulfur or selenium from (seleno)cysteine, leaving alanine [(48); for a review on cysteine desulfurases, see (47)]. A second NifS-like protein occurs in *A. vinelandii*, IscS, which has a housekeeping function in the formation of other cellular Fe-S proteins (49). *IscS* is present in a gene cluster that contains paralogs of some of the *nif* genes (*iscU*, similar to the N-terminus of *nifU,* and *iscA*); thus, the *nif* and *isc* clusters have a similar organization (49). The NifU- and NifA-like proteins are thought to serve a scaffold function for the Fe-S cluster during its synthesis and before its transfer to the target protein and conserved cysteines play a pivotal role in this process (50, 51). The Isc gene cluster also includes an Hsp70 and Hsp40 and a ferredoxin-type protein. Homologues of the *nif*/*isc* genes have been discovered in several

other bacteria including *E. coli* (49) and are also present in the mitochondria of eukaryotes (4). Next to IscU- and IscA-type proteins the mitochondria have an Nfu protein, which is similar to the C-terminus of NifU (4). In yeast mitochondria, the Hsp70/40 machinery is required for the utilization of Fe-S clusters assembled on IscU (52).

In *E. coli* and *Erwinia chrysanthemi,* a third gene cluster involved in Fe-S cluster formation is the Suf operon, which also includes a NifS-like cysteine desulfurase called SufS/CsdB in *E. coli*. (53). A major function of the Suf operon may be in protecting the cell from oxidative stress and iron starvation (54, 55). Figure 1 summarizes the structures of the three gene clusters implied in Fe-S formation in bacteria. A comparison of the sequences of NifS-like proteins from various organisms reveals two classes of these proteins (48). The Isc-type cysteine desulfurases fall into class I, whereas the Suf operon encoded NifS-like protein (SufS/CsdB) falls into class II, more related to enzymes implied in selenium metabolism (Figure 2). Besides a NifS-like protein the Suf operon contains SufA, SufB, SufC, SufD, and SufE. SufA is related to NifA and IscA and may have a scaffold function (56), whereas SufE was shown to activate SufS (57, 58). SufC is a nonintrinsic cytosolic member of the ABC domain transporter super-family. It forms a complex with Suf B and D, but the biochemical role of this complex is not yet clear (54, 58).

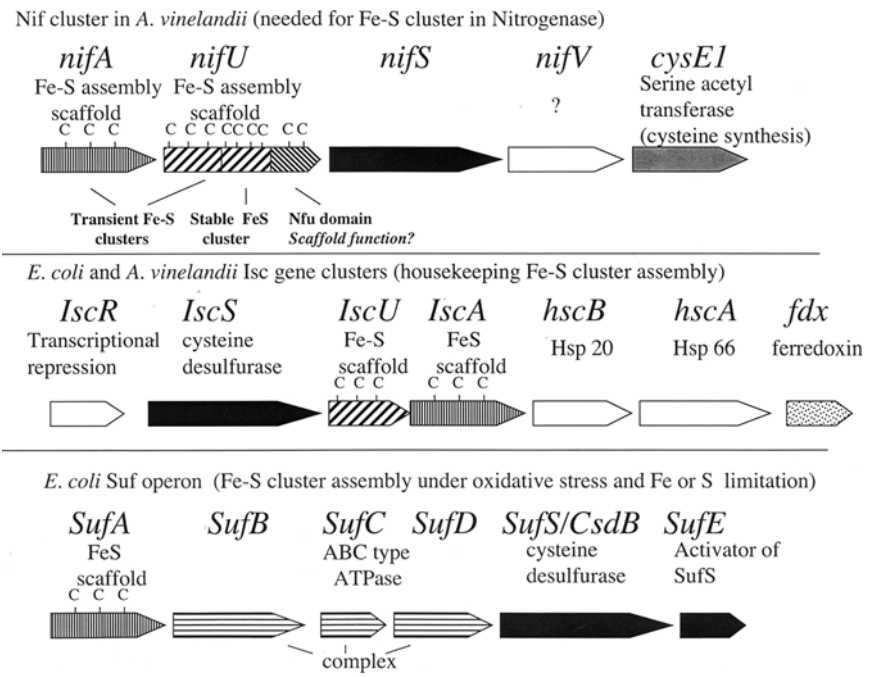


Figure 1. Overview of components of bacterial Fe-S assembly systems. Gene clusters are indicated with genes encoding proteins with similar structure indicated in a similar shading. Cysteines in scaffold proteins are indicated as c. Adapted with modification from (47).

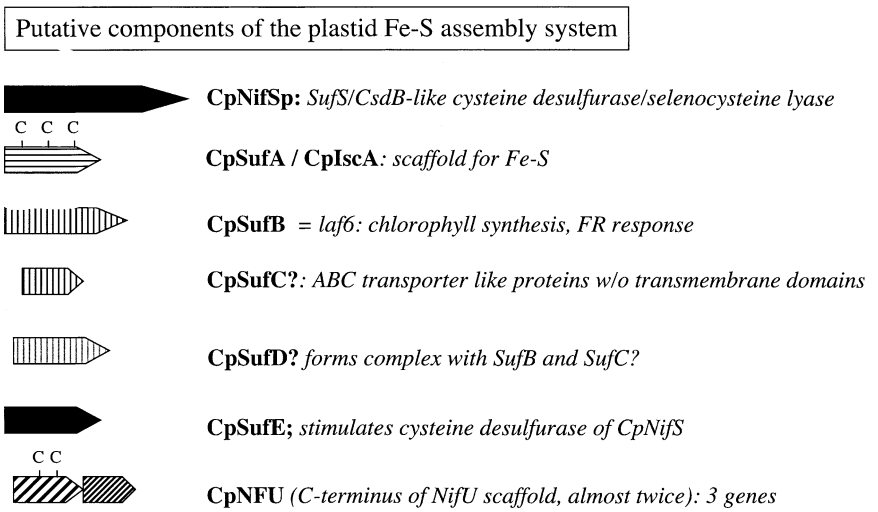


Figure 2. Schematic structure of chloroplast-localized homologues of bacterial Fe-S assembly proteins.

Chloroplasts are thought to be derived from a cyanobacterial ancestor. Even though still much is to be learned about the assembly of Fe-S clusters in cyanobacteria, it is of interest to know that the genome of non-nitrogen fixing cyanobacteria, which are perhaps the most close to plant plastids, encode homologues of IscS and Nfu (C-terminus of NifU) as well as homologues of the *E. coli* Suf operon including SufS, but proteins corresponding to IscU (or the N-terminus of NifU) are absent.

The Fe-S Assembly Machinery in Plastids

CpNifS

Inasmuch as cysteine was the sulfur source for Fe-S formation in ferredoxin, a plastidic NifS-like protein or a similar enzyme should be involved in Fe-S formation in chloroplasts (20). Two genes encoding NifS-like proteins have been identified in the *Arabidopsis* genome. One of the encoded proteins is present in mitochondria (14) and the other one, called CpNifS, is located in plastids (23, 24). The discovery of a NifS-like protein in plastids has prompted database searches for possible NifS-dependent protein factors that may function in Fe-S cluster assembly in chloroplasts. Putative Fe-S assembly factors with chloroplast

transit sequences are indeed encoded in the *Arabidopsis* genome (Table 1, Figure 2). CpNifS is most similar to a cyanobacterial NifS-like protein, and among the *E.coli* homologues is most similar to SufS/CsdB, a group II NifS-like protein (Figure 3). CpNifS was found to be able to use both Cys and SeCys as substrates, with a 300-fold lower cysteine desulfurase activity compared with its selenocysteine lyase activity (23).

In microbes, NifS-like proteins have also been implied to function in aspects of S metabolism other than Fe-S cluster formation, namely the biosynthesis of biotin, thiamine and molybdenum co-factor, MoCo (47). NifS-like proteins may play similar roles in plants. In bacteria and mammals, essential Se metabolism also involves NifS-like proteins, which are needed for the incorporation of Se into selenoproteins and seleno-tRNAs (47). A summary of the various possible roles of NifS-like proteins in S and Se metabolism is given in Figure 4.

It has now been shown that Se is an essential element for bacteria and animals—a requirement not yet shown for plants. On the other hand, higher Se concentrations are toxic to all organisms. Thus, organisms must prevent Se toxicity and at the same time many organisms need Se for their metabolism; NifS-like proteins may play a role in both aspects. Indeed, Arabidopsis plants that overexpress CpNifS show increased tolerance to selenate. Furthermore, transcript profiling experiments in Arabidopsis showed that a group of genes that are up-regulated in S deficiency are also up-regulated by selenate treatment, but this up-regulation is less pronounced in plants that overexpress CpNifS (Van Hoewyk et al., unpublished data). Together these results suggest that CpNifS can help reduce Se stress by avoiding Se-induced S deficiency.

The role of CpNifS in Fe-S formation was first addressed directly by Ye et al. (59). To test whether CpNifS is involved in Fe-S cluster formation for

Table 1. Putative components of a chloroplast Fe-S machinery in *Arabidopsis thaliana*.

Protein names	AGI#	Length	Proposed function	Reference
CpNifS/CpSufS	At1g08480	463	Cys-desulfurase	23, 24
CpNfu1	At4g01940	230	Scaffold	60, 61
CpNfu2	At5g49940	235	Scaffold	60, 61
CpNfu3	At4g25910	236	Scaffold	60, 61
CpSufA/CpNFA	At1g10500	180	Scaffold	Own results, Unpublished
CpSufB/Laf6	At4g04770	557	Far-red signaling?	63
CpSufC	At3g10670	338	ATPase/embryogenesis	64
CpSufD	At1g32500	475	Unclear	TAIR
CpSufE	At4g26500	371	Activator of CpNifS?	Own results, Unpublished
HCF101 (NifH-like)	At3g24430	532	4Fe-4S insertion	6

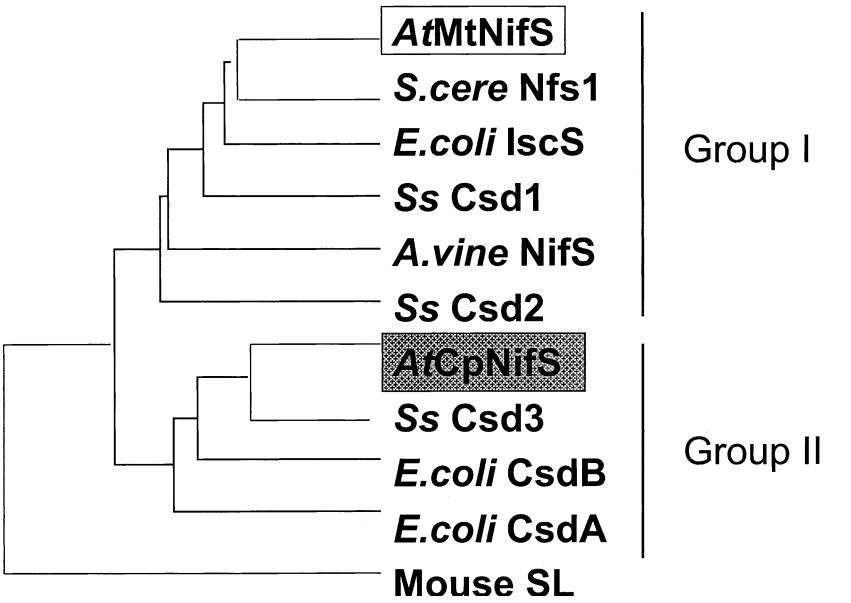


Figure 3. Grouping of NifS-like proteins based on sequence similarity. NifS-like proteins are pyridoxal-5 phosphate (PLP) dependent enzymes with both cysteine desulfurase and selenocysteine lyase activities that can be placed into two groups based on sequence similarity. In bacteria and in yeast mitochondria, cysteine desulfurases of group I with structural similarity to the NifS enzyme from *Azotobacter vinelandii* provide sulfur for Fe-S formation. Bacterial NifS-like proteins of group I such as IscS from *Escherichia coli* and *Azotobacter vinelandii* have been implied as housekeeping enzymes in Fe-S formation and are present in operons together with scaffolding proteins such as IscU and IscA . The physiological role of group II Nifs-like proteins such as *E. coli* SufS/CsdB in Fe-S synthesis is less evident, but work with double mutants indicates a partially overlapping function of IscS and SufS/CsdB in *E. coli*. AtMtNifS and AtCpNifS, the *Arabidopsis* mitochondrial and chloroplast NifS; *S. cere* NFS1, yeast mitochondrial NifS; SsCsd1, SsCsd2, and SsCsd3, synechocystis NifS-like proteins; *A. vine* NifS, *Azotobacter vinelandii* NifS; *E. coli* CsdB, SufS protein; *E. coli* CsdA, CsdA; Mouse SL, selenocysteine lyase from mouse, used as an outgroup in this phylogenetic tree (modified from 23).

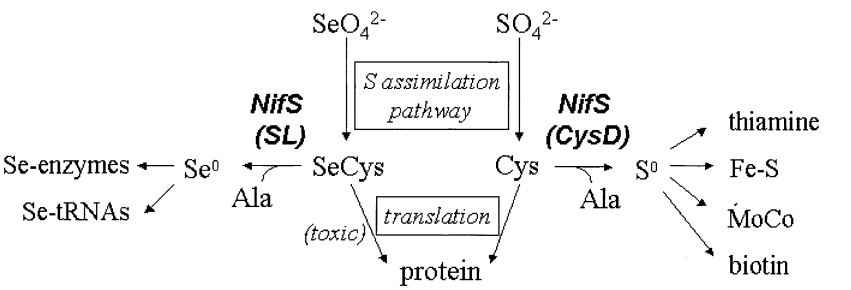


Figure 4. Possible functions of NifS-like enzymes in S and Se metabolism. Next to the function in Fe-S formation, NifS-like proteins of either group may be involved in the biosynthesis of thiamine, biotin, molybdenum cofactor, and seleno-protein and Se-tRNA synthesis.

photosynthetic proteins, an *in vitro* reconstitution assay was developed for ferredoxin. In this assay, apo-fd is reconstituted to the holo-form by acquiring an Fe-S cluster, which was synthesized *in vitro* from cysteine sulfur and a ferrous iron salt. Holo-fd was separated from apo-fd and other proteins and quantified by HPLC using an ion exchange column. Purified CpNifS was active by itself in stimulating holo-fd formation in this assay. The amount of reconstituted ferredoxin was dependent on the CpNifS concentration. It was calculated that, under the assay conditions, 16 molecules of apo-Fd were reconstituted per CpNifS monomer. Thus, CpNifS has a catalytic role in iron-sulfur cluster formation in ferredoxin *in vitro*. The activity requires an intact PLP-cofactor, and CpNifS protein with a mutation of the conserved active site cysteine (Cys_{418}-Ser) is inactive, indicating that ferredoxin reconstitution involves the cysteine desulfurase activity of CpNifS.

Stromal proteins at 300 μg/mL showed activity comparable to 10 μg/mL CpNifS. Based on a quantification by means of Western blotting, we calculated that CpNifS constitutes 0.06 ± 0.02% of total stromal protein. Thus, the apparent reconstitution activity of stroma was 50-80 times more than that of pure CpNifS protein and stromal components activate CpNifS. To investigate whether Fe-S cluster reconstitution activity of stroma was dependent on CpNifS, an affinity column was used to deplete stroma of CpNifS, the removal of which was confirmed by immunoblot. Both the original stroma and the antibody-treated stroma were examined for ferredoxin reconstitution activity. The activity of the antibody-treated stroma was decreased to background levels, suggesting that the reconstitution activity of stroma was entirely dependent on CpNifS. Importantly, adding back pure CpNifS to depleted stroma to its original concentration restored the reconstitution activity. Stroma that had been treated with preimmune serum did not lose its Fe-S reconstitution activity.

To investigate whether CpNifS may be complexed to other stromal proteins, a gel filtration experiment was performed using a high-resolution column, and the elution of CpNifS followed using immunoblotting. Purified CpNifS eluted from the column in a single peak with a retention time expected for the dimer, as was found before (23). Interestingly, the CpNifS present in stroma eluted in two peaks approximately 90% eluted as a CpNifS dimer of 86 kDa, as was observed earlier using pure CpNifS. An additional, smaller amount of CpNifS eluted at a high molecular weight of ~600 kDa. This result indicates that CpNifS interacts with other proteins *in vivo* and may form a transient complex with them (59).

CpNfu as a Possible Fe-S Assembly Scaffold Protein in Plastids

Database searches (TAIR: www.arabidopsis.org) indicate that plastids do not have an IscU homologue, or a protein similar to the N-terminal domain of NifU, but several other potential members of a plastid Fe-S cluster formation machinery were identified (see Table 1 for a listing). The three *CpNFU* genes (*CpNfu* 1-3) encode chloroplast proteins that are differentially expressed but closely related in sequence to each other and similar to cyanobacterial Nfu and the C-terminus of NifU (60, 61). The domain structure of CpNfu proteins

is of interest. The three chloroplast Nfu proteins have a domain with high similarity to cyanobacterial Nfu, including the conserved cysteine that is implied in transient cluster binding. In addition, a second Nfu-like domain that lacks the cysteine residues is present at the C-terminus of the CpNfu proteins. CpNfu2 forms a transient cluster (60) that can be passed on to apo-ferredoxin *in vitro* (60, 61). Insertion mutants in one of the *CpNfu* genes (*CpNfu2*, *At5g49940*) have a dwarf phenotype and are deficient in some but not all plastid Fe-S proteins (61, 62). In the mutant lines the accumulation of both 2Fe-S and 4Fe-S proteins (PSI and sulfite reductase) is diminished and the organization of PSI is affected. Interestingly though, the KO-is viable and Fe-S protein levels were only diminished *in vitro*. Furthermore, the Rieske type 2Fe-2S of the B/F complex and the 3Fe-4S cluster of glutamate synthase were not affected (61, 62). It is possible that those clusters would require the action of any of the other two CpNfu gene products. However, the expression levels and sequence similarities of these genes may suggest that a different scaffold may be required for these substrate proteins. Because CpNfu proteins can carry a transient Fe-S cluster that can be transferred to Fd, the observed effect on Fe-S assembly in the CpNfu2 mutant is likely a direct one. However, in view of the observed effect of the CpNfu mutation on two types of clusters, it would be of interest to verify the mRNA expression levels of Fd and PSI encoding genes to rule out indirect effects of the mutations. So far, a direct link between CpNifS and CpNfu has not been established at the biochemical level, but this may only be a matter of time.

CpIscA as an Alternative Scaffold

Another potential candidate for interaction with CpNifS is CpSufA (Abdel-Ghany, Ye, Pilon-Smits and Pilon, unpublished). A T-DNA insertion line for this gene was obtained. Thus far, only plants that are heterozygous for the insertion were found. A preliminary analysis indicates that when these heterozygotes are sown on media with sucrose, one-quarter of the seedlings show a visible growth phenotype. Thus, the homozygous CpSufA knockout may be lethal in plants grown on soil, perhaps due to impaired photosynthesis. However, further analyses will be required before firm conclusions on the *in vivo* role of CpSufA can be made. CpSufA was shown to be plastidic by GFP-fusion studies. We have purified CpSufA and studied its effects on CpNifS-dependent reconstitution of Fd *in vitro*. Pre-incubation of pure CpNifS and pure CpSufA in the presence of cysteine and a ferrous iron salt was shown to give a two-fold stimulation of apo-Fd reconstitution compared with CpNifS alone. Gel filtration experiments indicated purified CpSufA is a tetramer. However, upon incubation with CpNifS, purified CpSufA acquires a transient Fe-S cluster as indicated by the absorption spectrum of CpSufA and direct measurement of Fe and it becomes a dimer. The cluster in dimeric CpSufA can subsequently be transferred to apo-Fd to form holo-Fd. Thus, CpSufA can function as an assembly scaffold for Fe-S clusters.

Other Suf-Type System Components and Hcf101

Other potential candidate proteins that may assist CpNifS in Fe-S cluster formation in plastids are the Arabidopsis homologues of SufA, B, C, D, and E and Hcf101 protein (Table 1). The putative SufB and SufC homologues are confirmed to be in the chloroplast and mutants have phenotypes that indicate a role in development (63, 64). Expression profiling indicated that the potential *SufB* gene is regulated by Fe-deficiency in Arabidopsis (65), but, so far, a link of CpSufBCD or E with CpNifS or Fe-S clusters has not been made in plants. Bacterial SufE protein is required to stimulate the low endogenous cysteine desulfurase activity of SufS/CsdB. Our laboratory has localized the plant SufE protein to the chloroplast and we subsequently labeled the protein CpSufE. CpSufE is expressed in all major tissues, like CpNIFS, and it is feasible that CpSufE and CpNifS interact. Preliminary experiments in our lab indicate that CpSufE can indeed stimulate the cysteine desulfurase activity of CpNifS (Ye et al., unpublished). It will be interesting to see what the exact physiological role of the CpSufE is. The function of the bacterial SufB, C, and D proteins is still unclear; they appear to form a complex and may be involved in providing ferrous iron, or in transferring the Fe-S cluster from the scaffold protein to the target protein. In view of the function of the bacterial Suf operon in protection from oxidative stress, Suf homologues should make suitable members of the plastidic Fe-S cluster machinery, since the chloroplast is an oxygenic compartment due to its photosynthetic oxygen production.

Another interesting putative component of the plastid Fe-S machinery is HCF101. HCF101 (high chlorophyll fluorescence 101) encodes a NifH-related P-loop ATPase that seems to be required for 4Fe-4S but not 2Fe-2S in chloroplasts (66). The mechanism of action of the protein is so far not clear.

FUTURE PROSPECTS

The mechanisms of Fe-S assembly in plastids are complex and we are far from a complete understanding of this fascinating process. Figure 5 shows a working model for the Fe-S cluster formation machinery in plastids that includes components that have been characterized biochemically. The exact role of scaffold proteins in the biosynthesis of specific Fe-S proteins requires the development of sophisticated *in vitro* systems that can measure not only 2Fe-2S insertion in ferredoxin-type proteins, but also insertion in Rieske-type proteins, 4Fe-4S and 3Fe-3S proteins. This requires novel model proteins and assays that take into account the observed need for NADPH and ATP in plastid Fe-S assembly. Furthermore, the analysis of double mutants will help reveal possible functional overlap. To assess whether effects of mutations on the accumulation of proteins is a direct effect, expression at the mRNA level should be studied.

Thus far, very little is known about the molecular details of Fe uptake in plastids. Furthermore, the regulation of Fe-storage and recruitment for Fe-S assembly is unclear. How do plants coordinate the need for Fe in photosynthesis with S metabolism and Fe uptake and mobilization? This question is not trivial since both free Fe and S are considered toxic. The newly available genetic and

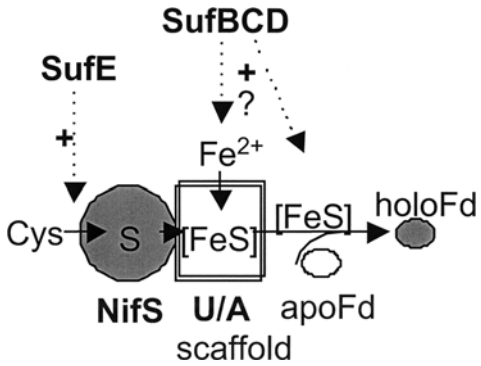


Figure 5. A working model for Fe-S formation in plastids.

genomics tools will help reveal novel elements of the Fe-S biosynthetic machinery and the regulation of the machinery as a whole in response to developmental cues, the need for photosynthesis, and nutrient status.

ACKNOWLEDGMENT

The work in the authors' laboratory is supported by USDA-NRI Grant No. 2003-35318-13758.

REFERENCES

1 Beinert, H., Holm, R.H. and Münck, E. (1997) Science 277, 653-659.
2 Imsande, J. (1999) Plant Physiol. Biochem. 37, 87-97.
3 Vorburger Mielcsarek, E. and Bertch McGrayne, S. (2000) Iron, Nature's Universal Element. Rutgers University Press, New Brunswick, N.J. pp. 1-204.
4 Lill, R. and Kispal, G. (2000) Trends Biochem. Sci. 25, 352-356.
5 Frazzon, J., Fick, J.R. and Dean, D.R. (2002) Biochem. Soc. Trans. 30, 680-685.
6 Raven, J.A., Evans, M.C. and Korb, R.E. (1999) Photosynthesis Res. 60, 111-149.
7 Lancaster, J.R., Vega, J.M., Kamin, H., Orme-Johnson, N.R., Orme-Johnson, W.H., Krueger, W.H. and Siegel, L.M. (1979) J. Biol. Chem. 254, 1268-1272.
8 Krueger, R.J. and Siegel, L.M. (1982) Biochemistry 21, 2892-2904.
9 Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K., Endo, T., Tasaka, M. and Shikanai, T. (2004) Nature 429, 579-582.
10 Marschner, H. (1995) Mineral Nutrition of Higher Plants, Academic Press, London. pp. 3-384.

11 Chitnis, P.R. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 593-626.
12 Buchanan, B.B., Gruissem, W.G. and Jones, R.L. (2000) Biochemistry and Molecular Biology of Plants. Amer. Soc. Plant Biol., Rockville, MD. pp. 2-1367.
13 Kispal, G., Csere, P., Prohl, C. and Lill, R. (1999) EMBO J. 18, 3981-3989.
14 Kushnir, S., Babiychuk, E., Storozhenko, S., Davey, M.W., Papenbrock, J., De Rycke, R., Engler, G., Stephan, U.W., Lange, H., Kispal, G., Lill, R. and Van Montagu, M. (2001) Plant Cell 13, 89-100.
15 Balk, J., Pierik, A.J., Netz, D.J., Muhlenhoff ,U. and Lill, R (2004) EMBO J. 23, 2105-2115.
16 Keegstra, K. and Cline, K. (1999) Plant Cell 11, 557-570.
17 Li, H-M., Theg, S.M., Bauerle, C.M. and Keegstra, K. (1990) Proc. Nat. Acad. Sci. U.S.A. 8, 6748-6752.
18 Pilon, M., de Kruijff, B. and Weisbeek, P.J. (1992) J. Biol. Chem. 267, 2548-2556.
19 Pilon, M., America, T., van't Hof, R., de Kruijff, B. and Weisbeek, P. (1995) Advances in Molecular and Cell Biology (Rothman, S.S. ed.) Membrane Protein Transport. JAI Press, Greenwich, CT. Vol. 4, pp. 229-255.
20 Merchant, S. and Dreyfuss B.W. (1998) Ann .Rev. Plant Physiol. Plant Mol. Biol. 49, 25-51.
21 Takahashi, Y., Mitsui, A., Hase, T. and Matsubara, H. (1986) Proc. Nat. Acad. Sci. U.S.A. 83, 2434-2437.
22 Takahashi, Y., Mitsui, A. and Matsubara, H. (1990) Plant Physiol. 95, 97-103.
23 Pilon-Smits, E.A.H., Garifullina, G., Abdel-Ghany, S., Kato, S., Mihara, H., Hale, K.L., Burkhead, J., Esaki, N., Kurihara, T. and Pilon, M. (2002) Plant Physiol. 130, 1309-1318.
24 Leon, S., Touraine, B., Briat, J-F. and Lobreaux, S. (2002) Biochem. J. 366, 557-564.
25 Curie, C. and Briat, J.F. (2003) Ann. Rev. Plant Physiol. Plant Mol. Biol. 54, 183-206.
26 Hell, R. and Stephan, U.W. (2003) Planta 216, 541-551.
27 Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S.L., Briat, J.F. and Walker, E.L. (2001) Nature 409, 346-349.
28 Robinson, N.J., Procterm, C.M., Connolly, E.L. and Guerinot M.L. (1999) Nature 397, 694-697.
29 Connolly, E.L., Fett, J.P. and Guerinot, M.L. (2002) Plant Cell 14, 1347-1357.
30 Vert, G., Grotz, N., Dedaldechamp, F., Gaymard, F., Guerinot, M.L., Briat, J.F. and Curie, C. (2002) Plant Cell 14, 1223-1233.
31 Shingles, R., North, M. and McCarty, R.E. (2002) Plant Physiol. 128, 1022-1030.
32 Petit, J-M., Briat, J-F. and Lobreaux, S. (2001) Biochem. J. 359, 575-582.
33 Beinert, H. (2000). Eur. J. Biochem. 267, 5657-5664.

34 Leustek, T., Martin, M.N., Bick, J-A. and Davies, J.P. (2000) Ann. Rev. Plant Physiol. Plant Mol. Biol. 5, 141-165.
35 Smith, F.W., Ealing, P.M., Hawkesford, M.J. and Clarkson, D.T. (1995) Proc. Nat. Acad. Sci. U.S.A. 92, 9373-9377.
36 Shibagaki, N., Rose, A., McDermott, J., Fujiwara, T., Hayashi, H., Yoneyama, T. and Davies, J.P. (2002) Plant J. 29, 475-486.
37 Yoshimoto, N., Takahashi, H., Smith, F.W., Yamaya, T. and Saito, K. (2002) Plant J. 29, 465-473.
38 Takahashi, H., Yamazaki, M., Sasakura, N., Watanabe, A., Leustek, T., de Almeida Engler, J., van Montagu, M. and Saito, K. (1997) Proc. Nat. Acad. Sci. U.S.A. 94, 1102-11107.
39 Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., Hawkesford, M.J. and Saito, K. (2000) Plant J. 23, 171-182.
40 Takahashi, H., Sasakura, N., Kimura, A., Watanabe, A. and Saito, K. (1999) Plant Physiol. 121, 686.
41 Grossman, A.R. and Takahashi, H. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 163-210.
42 Rotte, C. and Leustek, T. (2000) Plant Physiol. 124, 715-724.
43 Bork, C., Schwenn, J.D. and Hell, R. (1998) Gene 212, 147-153.
44 Zhou, J. and Goldsbrough, P.B. (1994) Plant Cell 6, 875-884.
45 Axelsen, K.B. and Palmgren, M.G. (2001) Plant Physiol. 126, 696-706.
46 Zheng, L., White, R.H., Cash, V.L., Jack, R.F. and Dean, D.R. (1993) Proc. Nat. Acad. Sci. U.S.A. 90, 2754-2758.
47 Mihara, H. and Esaki, N. (2002) Appl. Microbiol. Biotechnol. 60, 12-23.
48 Mihara, H., Kurihara, T., Yoshimura, T., Soda, K. and Esaki, N. (1997) J. Biol. Chem. 272, 22417-22424.
49 Zheng, L., Cash, V.L., Flint, D.H. and Dean, D.R .(1998) J. Biol. Chem. 273, 13264-13272.
50 Agar, J.N., Krebs, C., Frazzon, J., Huynh, B.H., Dean, D.R. and Johnson, M.K. (2000) Biochemistry 39, 7856-7862.
51 Krebs, C., Agar, J.N., Smith, A.D., Frazzon, J., Dean, D.R., Huynh, B.H. and Johnson M.K. (2001) Biochemistry 40, 14069-14080.
52 Muhlenhoff, U., Gerber, J., Richhardt, N. and Lill, R. (2003) EMBO J. 22, 4815-4825.
53 Takahashi, Y. and Tokumoto, U. (2002) J. Biol. Chem. 277, 28380-28383.
54 Nachin, L., Loiseau, L., Expert, D. and Barras, F. (2003) EMBO J. 22, 427-437.
55 Outten, F.W., Djaman, O. and Storz, G. (2004) Mol. Microbiol. 52, 861-872.
56 Ollagnier-de Choudens, S., Nachin, L., Sanakis, Y., Loiseau, L., Barras, F. and Fontecave, M. (2003) J. Biol. Chem. 278, 17993-18001.
57 Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. and Barras, F. (2003) J. Biol. Chem. 278, 38352-38359.
58 Outten, F.W., Wood, M.J., Munoz, F.M. and Storz, G. (2003) J. Biol. Chem. 278, 45713-45719.
59 Ye, H., Garifullina, G.F., Abdel-Ghany, S., Zhang, L., Pilon-Smits, E.A.H. and Pilon, M. (2004) Planta 220, 602-608 (2005).

60 Leon, S., Touraine, B., Ribot, C., Briat, J.F. and Lobreaux, S. (2003) Biochem. J. 371, 823-830.
61 Yabe, T., Morimoto, K., Kikuchi, S., Nishio, K., Terashima, I. and Nakai, M. (2004) Plant Cell 16, 993-1007.
62 Touraine, B., Boutin, J-P., Marion Poll, A., Briat, J-F., Peltier, G. and Lobreaux S. (2004) Plant J., 40, 101-111
63 Møller, G.M., Kunkel, T. and Chua, N.-H. (2001) Genes Dev. 15, 90-103.
64 Xu, X.M. and Møller, S.G. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 9143-9148.
65 Wintz, H., Fox. T., Wu, Y.Y., Feng, V., Chen, W., Chang, H.S., Zhu, T. and Vulpe, C. (2003) J. Biol. Chem. 278, 47644-47653.
66 Lezhneva, L., Amann, K. and Meurer, J. (2004) Plant J. 37, 174-85.
67 Noctor, G., Arisi, A.C.M., Jouanin, L. and Foyer, C.H. (1998) Plant Physiol. 118, 471-482.

IRON TRANSPORT AND METABOLISM IN PLANTS

Loubna Kerkeb and Erin L. Connolly

Department of Biological Sciences
University of South Carolina
Columbia, South Carolina 29208

INTRODUCTION

Iron is an essential micronutrient for plants. It is a component of a variety of enzymes involved in a range of biochemical processes, including respiration, photosynthesis, and nitrogen fixation. Iron functions to accept and donate electrons, and is thus involved in numerous electron transfer reactions. However, iron is toxic in excess since both Fe(II) and Fe(III) can act as catalysts in the formation of hydroxyl radicals, which are potent oxidizing agents that may damage DNA, proteins and lipids (1). For these reasons, it is crucial that cells maintain precise homeostatic mechanisms to ensure that iron is present at adequate, but nontoxic levels; both iron uptake and storage are carefully regulated processes.

A second factor that influences iron uptake by plants is the limited solubility of iron in aerobic soils of neutral or basic pH. Although iron is the fourth most abundant element in the earth's crust, it tends to form insoluble oxyhydroxide polymers and the concentration of free Fe(III) in aerobic soils of neutral pH ($\sim 10^{-17}$M) is much lower than concentrations required for the optimal growth of

Genetic Engineering, Volume 27, Edited by J. K. Setlow

plants (~10^{-9} to 10^{-4}M) (2). Iron deficiency is most commonly a problem for plants grown on calcareous soils, which represent approximately 30% of soils worldwide.

Iron deficiency is an enormous problem in human health; the World Health Organization estimates that over 3 billion people worldwide suffer from iron deficiency (3). Women and children, in particular, are at risk for iron deficiency. In many locations throughout the world, dietary iron is obtained primarily from plant foods. In general, plant foods are a poor source of iron as plants often contain low levels of iron, and iron in plant foods is often not bioavailable. For these reasons, it is crucial that we develop a comprehensive understanding of the machinery and regulatory mechanisms that control iron uptake and distribution within plants. Such information should facilitate the production of crop plants with improved iron content and thus would help to alleviate iron deficiency worldwide. Here, we describe recent progress made in understanding iron uptake and homeostasis in plants. We emphasize insights that are particularly relevant to the creation of crops with enhanced iron content and crops that are able to tolerate growth on soils considered iron deficient.

STRATEGY I

All dicots and nongrass monocots utilize the Strategy I response to mobilize and take up iron from the soil (4). In many respects, the Strategy I response of higher plants is similar to the high-affinity iron uptake system utilized by the yeast *Saccharomyces cerevisiae* (5). In addition, it is thought that the unicellular green alga *Chlamydomonas reinhardtii* utilizes a Strategy I–like iron uptake system (6). The three major components of the Strategy I response are: 1) secretion of protons to acidify the rhizosphere and enhance solubility of Fe(III); 2) reduction of Fe(III) to Fe(II); and 3) uptake of Fe(II) (Figure 1). In addition, in response to iron deficiency, some Strategy I species display morphological changes, including the development of root hairs and transfer cells, whereas others secrete flavins and phenolic compounds (7-9). A plasma membrane H^+-ATPase likely serves to pump protons across the plasma membrane of epidermal cells into the rhizosphere (10, 11). Although the precise identity of the H^+ATPase involved in the iron deficiency response is not known, recent work has resulted in the identification of both the Fe(III) chelate reductase and the Fe(II) transporter (1217).

Fe(III) Reduction

Reduction of Fe(III) to Fe(II) is carried out by a root, low iron–inducible, plasma membrane–bound, ferric chelate reductase (18). It is thought that reduction of Fe(III) to Fe(II), rather than uptake of Fe(II), is the rate-limiting step of iron uptake in Strategy I plants (19, 20). Two complementary approaches allowed the identification of the gene that encodes the ferric chelate reductase responsible for reduction of iron at the root surface. Yi and Guerinot identified an Arabidopsis mutant (*frd1,* for *f*erric *r*eductase *d*efective) that was unable to induce root-surface ferric chelate reductase activity in response to iron deficiency (13). A polymerase chain reaction (PCR)-based approach allowed the identification of a

family of genes in Arabidopsis (*FRO,* for *f*erric *r*eductase *o*xidase) that showed homology to yeast ferric chelate reductase and to the human neutrophil respiratory burst NADPH oxidase, gp91^{phox} (21). Gp91^{phox} is involved in the generation of active oxygen species in response to pathogen attack (22, 23). The *FRO2* gene was shown to map to the same location as the *frd1* mutation. Subsequently, transformation of the *frd1* mutant with the *FRO2* gene rescued the *frd1* phenotype and proved that *FRO2* encodes the root ferric chelate reductase (12).

FRO2 is predicted to contain eight transmembrane domains and contains four histidine residues that are thought to coordinate two heme groups that are located in the membrane (12). FRO2, like the yeast ferric chelate reductases and gp91phox, contains FAD and NADPH-binding domains. It is thought that FRO2 functions to accept cytosolic electrons, pass them through the heme groups across the membrane to Fe(III), which, in turn, is converted to Fe(II). Arabidopsis *FRO2* is expressed predominantly in the epidermal cells of iron-starved roots (20). The Arabidopsis genome contains seven additional *FRO* genes (12, 21); no other Arabidopsis *FRO* has been characterized functionally. The pea *FRO1* gene was identified and expression of *PsFRO1* in yeast showed that the enzyme functions to reduce Fe(III) to Fe(II) (24). *PsFRO1* is expressed in many locations throughout the plant, including the roots, nodules, and leaves, suggesting that PsFRO1 may function in iron uptake from the soil and in iron distribution within the plant (24).

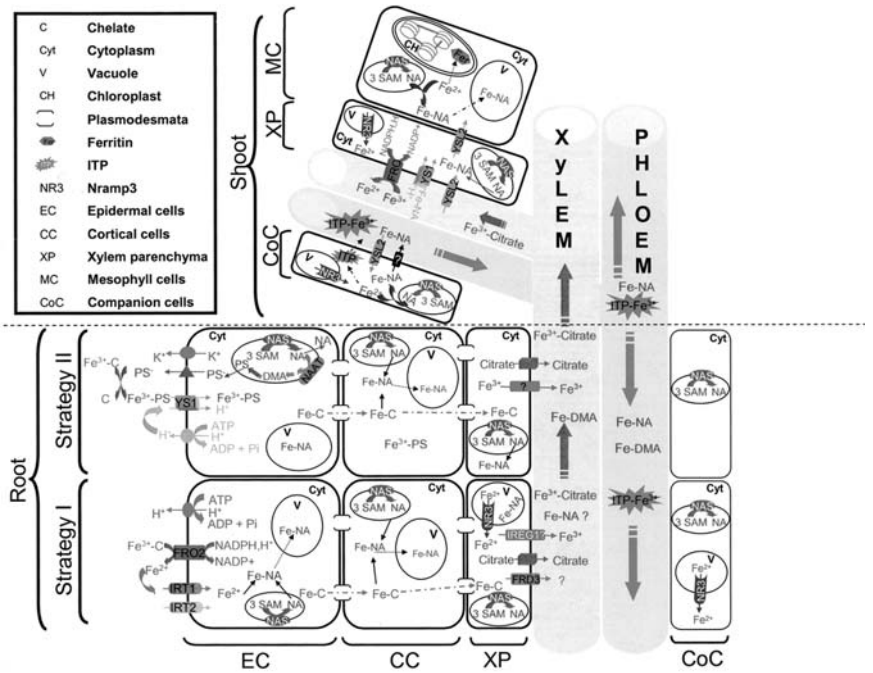


Figure 1. Overview of iron uptake and distribution in Strategy I and Strategy II plants. (See color insert.)

Fe(II) Transport

A yeast complementation approach allowed the identification of *IRT1* (*I*ron *R*egulated *T*ransporter) (14), which encodes the major high-affinity transporter responsible for iron uptake from the soil (15-17). In addition to iron, IRT1 is known to transport zinc, manganese, cadmium, and cobalt (14, 15, 25). Interestingly, mutation of particular amino acid residues in IRT1 alters the specificity of the transporter (26). *IRT1* is expressed in the epidermal cells of roots and in flowers (15) and expression of *IRT1* in roots is induced by iron starvation (14, 27). In addition, IRT1 is known to localize to the plasma membrane (15). Loss of IRT1 activity leads to reduced accumulation of iron, altered chloroplast morphology, severe chlorosis, and is lethal, as an *IRT1* knockout is unable to produce seed unless it is supplemented with high concentrations of iron (15-17). Presumably, application of iron rescues the *IRT1* knockout line through the activity of low-affinity iron transporters present in the outer cell layers of the root.

IRT1 is a founding member of the ZIP (*Z*RT1 *I*RT1-like *P*rotein) family of metal transporters (28-30). In Arabidopsis, the ZIP family contains 15 members (31). Of these, only *IRT1* and *IRT2* have been characterized. Like *IRT1*, *IRT2* is expressed in the outer layers of iron-deficient roots (32). However, the precise role of IRT2 is not yet known as an *IRT2* knockout line has no observable phenotype and overexpression of *IRT2* in the *IRT1* knockout line does not rescue the *IRT1* knockout phenotype (15, 16). *IRT1* homologues have been identified in a number of Strategy I species, including pea (33, 34) and tomato (35). Interestingly, *IRT1* homologues also have been identified in rice, a Strategy II species (36).

ZIP family members are characterized by a conserved topology, including eight predicted transmembrane domains and an intracellular loop domain between transmembrane domains 3 and 4. In most ZIP proteins, the intracellular loop is the site of a histidine motif. In IRT1, the motif consists of a series of alternating histidine and glycine residues (HGHGHGH) (14). Although it has been hypothesized that the histidine motif may be involved in sensing metal status within the cell, a role for the histidine motif has not yet been identified (14, 37).

Other metal transporters are implicated in iron transport in roots. In particular, members of the Nramp (*N*atural *R*esistance *A*ssociated *M*acrophage *P*rotein) family of proteins may have roles in iron uptake from the soil. The human DMT1 (Nramp2/DCT1) is responsible for iron uptake from the intestinal lumen (38). Six *Nramp* genes are present in the Arabidopsis genome (31, 39, 40). *Nramp* genes also have been identified in tomato (41), rice (42, 43), and soybean (44). Studies have shown that AtNramp1, AtNramp3, and AtNramp4 can transport iron when expressed in yeast and expression of each gene is induced by iron deficiency. Overexpression of *AtNramp1* leads to resistance to iron toxicity, suggesting that it may reside on an intracellular membrane and play a role in iron distribution within the cell, rather than iron uptake from the soil (39). AtNramp3 is located on the tonoplast of cells in the vasculature; analysis of *AtNramp3* knockout and overexpression lines suggests that AtNramp3 functions in efflux of metals from the vacuole (45). Studies of the expression of tomato *Nramp1* sug-

gest that, although it is expressed in roots and is up-regulated by iron deficiency, it may play a role in iron distribution within the plant (41). Thus, data collected to date implicate Nramp family members in iron transport within plants, rather than in uptake from the soil (see below).

STRATEGY II

Like most microorganisms, graminaceous plants utilize siderophores as a strategy to mobilize and take up iron when it is not readily available. Phytosiderophores (PS) are nonproteinogenic amino acids belonging to the mugineic acid (MA) family of compounds (46). PSs are synthesized and then secreted into the rhizosphere where they bind Fe(III) with high affinity. Fe(III)-PSs are then transported into root cells by a plasma membrane transport system (Figure 1) (47). In grasses, PS biosynthesis and secretion, as well as Fe(III)-PS uptake are induced by iron starvation. Different graminaceous species produce different types and quantities of PSs. The amount of PSs released into the soil correlates with the plant's ability to tolerate iron starvation (7, 48). Oat and barley are relatively resistant to iron limitation, whereas maize and rice (which secrete fewer PSs) are sensitive to iron limitation. In addition, plants use microbial siderophores for iron acquisition; however, little is known about this mechanism (49). It is thought that the Strategy II response is more efficient than the Strategy I response as grasses can survive on calcareous soils that do not support growth of nongrass species.

Phytosiderophore Biosynthesis

Plant PSs were first identified in oat and rice (50). Since then, the biochemical pathway for PS synthesis has been elucidated and many of the essential genes have been cloned from barley and rice (51-57). PS biosynthesis begins when three molecules of S-adenosylmethionine (SAM) are combined to form one molecule of nicotianamine (NA); this occurs in a single enzymatic step catalyzed by nicotianamine synthase (NAS).

Higuchi et al. were first to isolate seven *HvNAS* genes from barley (51); subsequently, *NAS* genes have been identified in barley (58), tomato (59), Arabidopsis (60), rice (61), and maize (62). *In vitro* NA production has been demonstrated for HvNAS1, AtNAS1, AtNAS2, AtNAS3, ZmNAS1, OsNAS1, OsNAS2, and OsNAS3 (60, 61, 63). In addition, several EST clones corresponding to putative *NAS* genes are present in the public databases for other grass species including wheat and winter rye. *NAS* expression and activity in roots is regulated by iron availability. *HvNAS1* is not expressed in iron-sufficient roots but is expressed at high levels in iron-deficient roots (51). In maize, transcripts of *ZmNAS1* and *ZmNAS2* were detected in roots but not leaves of iron-sufficient plants; *ZmNAS1* and *ZmNAS2* expression was induced in iron-deficient roots (62). *OsNAS1* and *OsNAS2* transcripts and protein accumulate in response to iron starvation (63). Promoter-*GUS* analysis showed that, under iron-sufficient conditions, *OsNAS1* and *OsNAS2* are expressed in root companion cells and pericycle cells adjacent to the xylem, whereas under iron starvation, the expression of

these genes extended to all root cells and corresponded to an increase in secretion of PSs.

Although the first steps of PS production up to NA synthesis are shared by monocots and dicots, the subsequent reactions of PS synthesis are exclusive to grasses. Deamination of NA is carried out by nicotianamine aminotransferase (NAAT) (64). The product 2'-deoxymugineic acid (DMA) is the precursor of all other MAs. Mori's group showed that introducing the barley *NAAT* genes into rice confers an increase in PS secretion and consequently an enhanced tolerance to low iron availability; this result suggests that the conversion of NA to DMA is the rate-limiting step in MA production in rice (65). DMA can be hydroxylated by a dioxygenase encoded by *IDS3* to form MA. MA and DMA can undergo additional hydroxylation catalyzed by the IDS2 protein to form further MA derivatives such as epiHMa and epiHDMA (46, 66). The additional hydroxyl groups are thought to increase the stability of the Fe(III)-chelate complexes (67).

Phytosiderophore Secretion

The molecular mechanism of MA secretion is still unclear. In barley, MA secretion has been shown to follow a distinct diurnal rhythm (53). A peak in secretion of MAs occurs just after initial illumination and secretion ceases 2-3 h after daybreak. In addition, the diurnal rhythm of MA secretion is correlated with high equimolar potassium release from cortex cells. Sakaguchi et al. showed that MA secretion in iron-deficient roots of barley is inhibited by a number of anion channel blockers and by valinomycin, which disrupts potassium gradients. These results suggest that MAs are secreted as monovalent anions *via* anion channels using the potassium gradient. However, at present, neither the genes nor the protein(s) involved in MA efflux have been identified (68).

Moreover, it has been suggested that secretion of MAs might require vesicular transport; iron-deficient barley roots display an increase in the size and number of vesicles in cortical cells, which correlates with the initiation of MA release. The authors speculated that such vesicles might be the site of MA synthesis (69). It is likely that these vesicles are derived from rough endoplasmic reticulum (rER), as they contained ribosomes on their cytoplasmic face (66). Localization studies using ZmNAS1-sGFP and ZmNAS2-sGFP fusion proteins suggest that ZmNAS1 and ZmNAS2 are located at the membrane of vesicles derived from the ER where MA synthesis likely takes place (62). A microarray study showed that two barley genes encoding proteins from the Rab GTPase and ARF (*A*DP-*R*ibosylation *F*actor) families, which are thought to participate in directional intracellular transport of vesicles or organelles, showed both induction in response to iron limitation and diurnal regulation. Expression of both genes peaked just prior to the onset of MA secretion. The authors speculated that these proteins may be involved in vesicle-mediated secretion of MAs (66).

The *yellow-stripe 3* (*ys3*) mutant of maize displays interveinal chlorosis, a typical iron deficiency trait. This phenotype can be rescued by co-cultivation with wild-type plants or by exogenous application of MAs suggesting that the interveinal chlorosis phenotype is due to disruption of MA secretion (70). Identification of the *YS3* gene should provide insight into the mechanism of MA secretion.

Fe(III)-Phytosiderophore Uptake

Our understanding of Fe(III)-PS complex uptake was significantly facilitated by analysis of the maize *yellow-stripe1* (*ys1*) mutant. *ys1* makes normal amounts of PSs but lacks the ability to use Fe(III)-PS complexes efficiently (71). Recently, Curie et al. cloned the maize *YS1* gene and showed that it encodes the root Fe(III)-PS transporter. ZmYS1 is a 682 amino acid protein that is predicted to contain 12 transmembrane-spanning domains (47). ZmYS1 belongs to the oligopeptide transporter (OPT) family; because MAs are modified trimers of SAM, it is logical that they would be transported by a member of the OPT family. Yeast functional complementation experiments showed that expression of *ZmYS1* restored the iron-limited growth defect of the *fet3fet4* yeast strain when iron was supplied as Fe(III)-DMA, but not when it was supplied as Fe(III)-citrate (47). This result supports the hypothesis that YS1 is involved in transport of Fe-MA complexes in grasses. Gene expression analyses revealed that *ZmYS1* mRNA is detectable in roots but not in leaves of Fe-sufficient seedlings and expression is induced in roots and leaves when plants are grown in the absence of iron. ZmYS1 protein levels correlated with *ZmYS1* transcript levels because only Fe-starved tissues contained high levels of ZmYS1 protein (72). The expression of *ZmYS1* in the leaves suggests that ZmYS1 may have multiple functions in the plant; thus, in addition to functioning in iron uptake by roots, ZmYS1 may have a role in iron distribution in the aerial portions of the plant (47). Yeast functional complementation, expression in *Xenopus* oocytes and radioactive isotope uptake assays indicated that ZmYS1 is able to transport Fe(III)-PS, Fe(II)-NA, and possibly Fe(III)-NA, which is consistent with the idea that ZmYS1 may function in the shoot tissues as a transporter for Fe-NA complexes, Fe-PS complexes or both when plants are iron-limited (72, 73).

Roberts et al. examined the role of ZmYS1 in possible transport of metals other than iron. ZmYS1 was shown to mediate transport of Cu-MA complexes. However, *ZmYS1* mRNA and protein levels did not change in response to Cu deficiency, suggesting that ZmYS1-mediated transport of copper many not be physiologically significant (72). Competition with ^{55}Fe uptake assays revealed that Co-MA is a potential substrate for transport by ZmYS1 (72). Schaaf et al. showed that although exhibiting a high affinity for iron, ZmYS1 is able to transport Zn-DMA, Cu-DMA, Ni-DMA, Ni-NA, and, with a lower affinity, Mn-DMA and Cd-DMA. Metal-DMA, as compared with metal-NA, is the favored substrate for ZmYS1. Fe(III)-DMA uptake by ZmYS1 is improved with decreasing external pH and is severely inhibited by CCCP, a proton uncoupler, suggesting that ZmYS1 functions as a proton-coupled symporter for metal-PS and metal-NA complexes (73). This result is surprising since the release of phytosiderophores by grasses increases under iron limitation, which occurs in calcareous soil marked by a high pH. These data suggest that a plasma membrane proton-ATPase may generate a local pH gradient that allows Fe(III)-DMA/proton co-transport (73).

Following the identification of *ZmYS1*, a large number of *YS1* orthologs were noted in existing public databases; the rice, Arabidopsis, and maize genomes each contain a number of *YS1*-like genes (47). The presence of *YSL* (*y*ellow *s*tripe-*l*ike) genes in Arabidopsis is intriguing because this is a Strategy I plant

that is able to synthesize NA, but cannot synthesize or utilize MAs. Besides Fe(II) and Fe(III), NA is known to bind manganese, zinc, copper, and other micronutrients *in vivo* and *in vitro* (58, 74–76). NA is important for iron homeostasis and appears to play a role in the internal transport of Fe(II) and Fe(III) in Strategy I plants (59). Thus, it has been proposed that YSL proteins function to transport metal-NA complexes in Strategy I plants (47).

LONG DISTANCE TRANSPORT OF IRON

Following uptake of iron across the plasma membrane of root epidermal cells, iron must move laterally to the vasculature. Iron is loaded into the xylem and moves to the aerial portions of the plant *via* the transpiration stream. Subsequently, iron may leave the xylem and cross the plasma membrane of leaf cells. Iron also may be loaded into the phloem for transport to the root and shoot apices and developing seeds (Figure 1).

Radial Transport Across the Root

Long-distance iron transport requires radial transport from the root epidermal cells to the xylem through either a symplastic or apoplastic pathway. Iron that moves to the xylem *via* the symplastic pathway is chelated or sequestered in a nonactive form to prevent precipitation and generation of oxygen radicals (77). Although several different candidate chelators have been proposed, including organic and amino acids, NA emerges as the most likely candidate for a number of reasons. NA is found in all higher plants (78), NA forms stable complexes with Fe(II) and Fe(III) (79), and iron-NA is less reactive than free iron (74). Moreover, the *chloronerva* mutant of tomato, which is unable to synthesize NA due to disruption of the single gene in tomato encoding NAS, displays iron-deficiency symptoms such as interveinal chlorosis in young leaves and up-regulation of components of the root iron acquisition system (51, 58, 59). Thus, it is very likely that NA may act to chelate iron and facilitate its symplastic translocation to the xylem parenchyma (79, 80). Presumably, once iron enters the symplastic pathway, it may move *via* plasmodesmata all the way to the xylem parenchyma cells. In addition, it is noteworthy that much of the iron associated with roots is part of the apoplastic pool of iron (81).

Iron Loading to Xylem

Loading of iron into the xylem requires efflux from xylem parenchyma cells resulting in iron transfer to the apoplast. Very little is known about iron release into the stele. DeBoer et al. showed that a respiration dependent proton pump at the plasma membrane of xylem parenchyma cells pumps protons into the apoplast and lowers the pH of the xylem sap creating a driving force for potential cation/H^+ antiport (82). On the other hand, potential measurements related to fluxes of cations and anions at the plasma membrane suggested that the release of ions into the xylem sap occurs in a passive manner through ion channels (83). Iron efflux proteins have not yet been identified in plants. A mam-

malian iron transporter, IREG1 (also known as ferroportin), is involved in iron efflux by intestinal enterocytes and is induced by iron deficiency (84). Three IREG/ferroportin genes are found in the Arabidopsis genome. It is tempting to speculate that these genes may encode transporters involved in iron efflux to the xylem.

Besides being the principle chelator of free iron in cells and a substrate in PS biosynthesis in Strategy II plants, NA might have an additional role in iron loading into the xylem. Analysis of *OsNAS1-GUS* and *OsNAS2-GUS* transgenic lines showed that *OsNAS1* and *OsNAS2* are expressed in the pericycle cells adjacent to the xylem of both Fe-sufficient and Fe-deficient roots, implying that *OsNAS1* and *OsNAS2* contribute to the biosynthesis of NA or DMA necessary for xylem loading (63). Large amounts of DMA have been detected in the xylem sap of Fe-sufficient plants and Fe-deficient plants (85, 86); NA also has been detected in xylem exudates (75).

Iron Movement through Xylem

Iron is translocated through the xylem as a Fe(III)-citrate complex (87), which suggests that the Fe(II) taken up by Strategy I plants must be oxidized to Fe(III) at some point during the lateral movement of iron within the root. A positive correlation between iron deficiency and increased concentrations of organic acids in the roots, xylem exudates, and leaves have been observed in both Strategy I and Strategy II species (88). Pich et al. also noted that the *chloronerva* (*chln*) mutant of tomato, which is defective in NA synthesis, has increased levels of citrate in stem exudates (89). In several species, the concentration of organic acids in roots is higher in iron-efficient genotypes than iron-inefficient ones (88).

Iron Uptake by Leaf Cells

The machinery involved in movement of iron from the xylem into the leaf symplast is still unclear (80, 88, 90). According to Marschner, delivery of iron to different organs starts with release of free Fe(II) within the apoplast. Fe(III) could be reduced through photo-reduction of the Fe(III)-citrate complex or enzymatically by a plasma membrane reductase protein (91). Apoplastic ascorbate also could function in the generation of free Fe(II) (92-94). In addition, Fe(II) may be re-oxidized and precipitate in the cell wall space, possibly as Fe-hydroxide or Fe-phosphate species.

Fe(III) reductase activity has been detected in leaves of sunflower (95) and *Vigna unguiculata* (92). In pea, expression of *FRO1* was induced in leaves in response to iron limitation and *in situ* hybridization showed that *PsFRO1* is expressed in the mesophyll cells of leaves (24). In addition, expression of *AtFRO3* was shown to be induced by iron-deficiency in leaves (21, 96). Thimm et al. used a microarray approach to show that expression of Arabidopsis P-type ATPases is induced by iron deficiency in leaves, suggesting that acidification of the apoplast might be required for iron uptake by the leaf cells (97). Acidification of the apoplast would facilitate reduction of Fe(III) and might contribute to the regulation of iron transport across the plasma membrane (88, 98).

Iron may be transported across the plasma membrane of leaf cells as a complex with NA, and YSL family members may be involved in transport of Fe-NA complexes into leaf cells. AtYSL2 has been shown to localize to the plasma membrane and analysis of *AtYSL2-GUS* transgenic lines showed that *AtYSL2* is expressed in the vasculature of leaves. Available data support the hypothesis that AtYSL2 transports Fe-NA across the plasma membrane of leaf cells, and thus is involved in lateral movement of iron away from the xylem (99). Other members of the YSL, ZIP, and/or Nramp families may be involved in uptake of iron by leaf cells; functional characterization of loss-of-function single- and double-mutant lines should help to definitively identify the transporters involved in iron transport into leaf cells.

Recent work showed that *FRD3*, which encodes a member of the multidrug and toxin efflux (MATE) transporter family (100), is expressed in the root pericycle and vasculature. FRD3 localizes to the plasma membrane, and iron is mislocalized in the *frd3* mutant. Green and Rogers speculate that FRD3 might be required for the efflux of a low-molecular-weight organic compound into the xylem in the roots that is necessary for unloading of iron from the xylem in the shoot. The substrate for FRD3 remains to be identified (101).

Phloem Transport of Iron

Fe(II) absorbed by the leaf cells may be distributed to various intracellular compartments and may serve as a co-factor for various enzymes, assist in chlorophyll synthesis, and be stored within the chloroplastic iron storage protein ferritin for future use; or it may be exported to developing sinks and growing roots *via* the phloem pathway. To be soluble and mobile in the phloem sap (pH 7.2-8.5), iron must be chelated. Studies with the iron-hyperaccumulating pea mutants *bronze* (*brz*) and *degenerative leaflets* (*dgl*) confirmed that iron is loaded into the phloem in a chelated form and that overexpression of iron chelators leads to increased loading of iron into the phloem (102, 103). Fe(III) can be chelated by ITP1 (*I*ron *T*ransport *P*rotein), an 11 kDa protein belonging to the LEA (*L*ate *E*mbryogenesis *A*bundant) family for transfer in the phloem. ITP1 was identified in *Ricinus communis,* although a putative ortholog exists in the Arabidopsis genome (104).

Studies suggest that NA is required for distribution of iron through the phloem (75, 105). As described above, Inoue and collaborators showed that *OsNAS1* and *OsNAS2* are expressed in the companion cells of Fe-deficient roots and to lesser extent in the companion cells of Fe-sufficient roots, implying that NAS activity is required in companion cells for phloem loading and/or unloading and that the requirement for NA is higher in companion cells of Fe-deficient roots (63). Moreover, *OsNAS3* expression was restricted to the pericycle and companion cells of the roots and to the companion cells of leaves irrespective of Fe status, suggesting a role in iron loading and/or unloading in the xylem and phloem. Furthermore, *OsYSL2*, a plasma membrane Fe-NA transporter, has an expression pattern in the phloem companion cells that is similar to that of *OsNAS3* (106). Together, these results support the idea that Fe is transported in the phloem as a complex with NA.

The expression of *OsYSL2* in the vascular bundles of reproductive organs suggests that translocation of iron into developing seeds is mediated by NA (106). *AtYSL2* also is expressed in the vasculature of flowers and siliques, suggesting that NA might contribute to iron translocation into flowers and developing siliques in dicots as well as monocots (99). In tobacco, northern analysis showed that *NtNAS* is expressed in petals, filaments, pistils and anthers, indicating that NA is produced in the inflorescence and may be required for proper flower development by supplying iron to developing pollen (107). Furthermore, expression of HvNAAT in tobacco (a Strategy I species that lacks NAAT) causes reduced NA levels; this, in turn, results in altered flower development, indicating that NA is required for proper distribution of iron to developing flowers (107). Arabidopsis *IRT1* and *FRO2* also are expressed in the anther filaments, suggesting that AtIRT1 and AtFRO2 may function in iron loading of pollen, in addition to iron uptake from the soil (15, 20). Other Arabidopsis *FRO* family members also are expressed in the flowers and in siliques; thus, movement of iron from the phloem to the seed may require reduction of Fe(III) to Fe(II) (Mukerjee, Campbell and Connolly, submitted).

INTRACELLULAR DISTRIBUTION AND STORAGE OF IRON

The molecular mechanisms that control distribution of iron to the various organelles in plants are largely unknown. It is likely that some members of the ZIP, NRAMP, YSL, IREG, and/or other families may be involved in transport of iron into cellular compartments. In addition, it is possible that ferric chelate reductase activity is required for iron uptake by particular organelles.

Chloroplast

In leaves, 90% of the iron in cells is located in the chloroplast (108). Iron absorption by isolated chloroplasts is regulated by light and depends upon electron transport in thylakoid membranes or the ATP generated by these membranes (109). Fe(II) can be transported across the chloroplast inner membrane (110). The Arabidopsis genome contains eight *FRO* genes; only *FRO2* has been characterized to date (12, 20). Although it is likely that FRO2 resides at the plasma membrane, it is possible that other AtFROs could be involved in iron distribution to various subcellular compartments. AtFRO7 and possibly AtFRO6 are predicted to localize to the chloroplast. Thus, they might be involved in Fe translocation into the chloroplast (Mukerjee, Campbell and Connolly, submitted).

Storage of iron in plant cells is largely accomplished by complexation with ferritin. In contrast to the situation in animal cells, in plants, ferritin is found in the chloroplast (111). Ferritin forms a complex with 24 subunits that is able to store up to 4,500 atoms of iron in a central cavity. In plants, expression of ferritin is controlled at the transcriptional and post-transcriptional levels, whereas in animals, expression of ferritin is regulated at the level of translation. Four ferritin genes have been identified in the Arabidopsis genome and each encodes a protein that is predicted to contain a transit peptide for delivery to the plastid. The

expression of *AtFer1* and *AtFer3* is up-regulated in the root and shoot in response to iron overload and H_2O_2, implying that chelation of iron by ferritins serves to protect cells from oxidative stress caused by iron overload. *AtFer4* expression is induced by iron overload but not H_2O_2. *AtFer2* gene expression is specific to mature siliques and dry seeds suggesting a role in iron storage in seeds (112).

Vacuole

Aside from storage of Fe-ferritin in the chloroplast, little is known about intracellular distribution of iron in plant cells. NA might play an important role in iron intracellular distribution because NA is localized in the cytoplasm of tomato leaf and root cells under low iron conditions, whereas NA is thought to be mainly found in the vacuole when plants are grown under high iron conditions (113). These data suggest that a member of the YSL family might be involved in NA translocation into the vacuole.

In Arabidopsis, an *AtNRAMP3-GFP* fusion construct revealed that AtNRAMP3 is localized to the vacuolar membrane in onion cells and Arabidopsis protoplasts; this, together with additional data, suggested that AtNRAMP3 might be responsible for iron efflux from the vacuole into the cytosol (45). AtNRAMP1 might be localized either to the plasma membrane, plastids or the tonoplast (39), whereas the iron transporters LeNRAMP1 and LeNRAMP3 are localized in intracellular vesicles and/or vacuole depending on the iron status of the plant (41). These data imply that in plants, NRAMP proteins might generally function in intracellular distribution/redistribution of iron.

Mitochondria

Iron-sulfur (Fe-S) cluster-containing proteins play crucial roles in many important cellular processes. In eukaryotes, maturation of cellular Fe-S proteins is carried out mainly in the mitochondria and defects in the biogenesis of Fe-S cluster-containing proteins disturb iron metabolism of mitochondria (114). Recent work suggests that, in plants, Fe-S cluster biogenesis occurs both in chloroplasts and mitochondria (115-117). Thus, iron influx to the mitochondria is required for proper Fe-S protein biosynthesis. To date, iron transporters involved in uptake of iron by mitochondria have not been identified. Atm1p, a yeast ATP-binding cassette (ABC) transporter located at the mitochondrial inner membrane in yeast, has been proposed to be an effluxer of Fe-S proteins from mitochondria to the cytosol. The *AtSTA1* gene encodes a protein homologous to Atm1p. A T-DNA insertion in *AtSTA1* causes chlorosis and analysis of the iron levels of isolated mitochondria showed that mitochondria isolated from *sta1* cells contain 50-80% more free iron than mitochondria isolated from wild-type cells (118).

IRON REGULATION, SENSING, AND SIGNALING

Because iron is both essential and potentially toxic, cells maintain precise regulatory mechanisms to control levels and forms of iron within cells. Iron

uptake, compartmentalization, distribution, complexation, and storage are carefully balanced to ensure that cells receive an adequate supply of iron for metabolic processes, while avoiding iron overload.

Transcriptional Regulation

Many studies have reported genes whose expression is regulated by iron status. For example, expression of *AtIRT1* and *AtFRO2* is carefully controlled such that transcripts accumulate following the imposition of iron-deficiency and transcripts disappear upon iron re-supply (20, 119). Wintz et al. used microarray analysis to study the regulation of gene expression by iron, zinc, and copper deficiency (96). They confirmed previous reports showing that *AtFRO2*, *AtFRO3* and *AtIRT1* are up-regulated by iron deficiency (12, 14, 21) and that ferritin transcription (*AtFer1* and *AtFer4*) is down-regulated by iron deficiency (120). In addition, expression of two *AtNAS* genes was shown to be elevated in response to iron deficiency (96). Interestingly, this study showed that expression of several other putative transport proteins is regulated by iron deficiency; for example, two members of the Arabidopsis oligopeptide transporter family, *AtOPT2* and *AtOPT3*, show dramatic increases in transcript abundance in response to iron deficiency. These results, then, implicate a number of previously uncharacterized transport proteins in iron metabolism. Functional characterization of these genes will determine what role, if any, the corresponding proteins play in iron homeostasis in plants.

As mentioned, expression of ferritin in plants is mainly regulated at the level of transcription and ferritin levels increase in response to iron excess and decrease in response to iron deficiency in leaves as well as in roots in several plants species (111). Petit et al. identified a *cis*-regulatory element in the promoter region of *ZmFer1* (120). This iron-dependent regulatory sequence (IDRS) is responsible for transcriptional repression of *ZmFer1* and *AtFer1* under low iron supply conditions. In soybean, an 86 bp sequence named the iron-regulatory element (IRE) was identified upstream of the ferritin gene and was shown to be involved in iron-mediated derepression of the ferritin gene (121).

As mentioned above, the barley *IDS2* (*I*ron *D*eficiency *S*pecific clone #2) gene encodes an enzyme involved in synthesis of phytosiderophores (122). Kobayashi et al. have shown that the promoter of *IDS2* contains two homologous sequences (IDE1 and IDE2) that are important for iron-deficiency-inducible expression in roots (123). Similar elements were found in many iron-deficiency-inducible promoters, including *HvIDS2*, *HvNAAT-A*, *HvNAAT-b*, *HvNAS1*, *HvIDS3*, *OsNAS1*, *OsNAS2*, *OsIRT1*, *AtIRT1,* and *AtFRO2*, suggesting the conservation of *cis*-acting elements in various genes among Strategy I and Strategy II plants.

A microarray study using 16,128 clones corresponding to at least 6,000 Arabidopsis genes revealed induction of expression of many genes involved in mobilization and export of carbon, glycolysis, the citric acid cycle, the oxidative pentose phosphate pathway, and fermentation in response to iron deficiency. These results suggest that the response to iron deficiency requires an overall increase in respiration and an increase in carbon import and anaerobic respiration

in the roots (97). Another microarray study examined iron-deficiency-induced changes in expression of barley genes using a rice microarray that contained approximately 8,987 EST clones (66). Approximately 200 genes were found to be induced by iron limitation, including genes that encode enzymes involved in methionine biosynthesis (phytosiderophores are synthesized from methionine) and the Yang cycle, which plays a key role in methionine recycling (66).

Post-translational Regulation

Recent evidence suggests that, in addition to transcriptional control of gene expression, a post-translational mechanism may be important for proper maintenance of iron homeostasis. Connolly et al. reported that despite high levels of *IRT1* mRNA in the shoots and roots of iron-deficient and iron-sufficient *35S-IRT1* plants, IRT1 protein is found only in iron-deficient roots (27). In addition, FRO2 also is regulated post-transcriptionally (20). ZRT1 (*Z*inc-*R*egulated *T*ransporter), a ZIP family member, is the high affinity zinc transporter in yeast; it is known that ZRT1 is subject to ubiquitination and protein degradation in the vacuole when zinc is present at high levels (37, 124, 125). Inasmuch as IRT1 is related to ZRT1, the mechanism of post-transcriptional regulation of IRT1 may be similar to that of ZRT1. Indeed, our recent experiments support this hypothesis (Kerkeb et al., unpublished data).

Sensing and Signaling

Plants must correctly sense the iron status of the shoot and send a signal to the root to maintain appropriate levels of iron. The initial information acquired regarding the signaling pathway was obtained from studies done with mutants with alterations in iron nutrition. The phenotypes of the pea mutants *brz* and *dgl* strongly suggest that the iron stress responses are under long-distance control by the shoot. *brz* and *dgl* are nonallelic recessive mutations that lead to constitutive iron acquisition. Reciprocal grafting experiments suggest that these mutants constitutively express a shoot element that triggers up-regulation of ferric chelate reductase and H^+-ATPase activities in the roots (126). The identity of this signal is still unknown, but it is most likely transmitted by the phloem (127).

More recent experiments suggest that, in addition to the shoot-derived long-distance signal, roots are able to sense local fluctuations in the availability of external iron. Root hair formation requires the presence of iron in the medium, indicating that a cell-specific response is responsible for the change in epidermal cell patterning (98, 128). Split-roots experiments revealed that a higher number of transfer cells and an increased H^+-ATPase density in the root epidermis were mainly formed on the roots of tomato plants exposed to iron-deficient medium. Thus, the development of transfer cells is primarily dependent on the local surroundings of the root and is less affected by the global nutrient status of the plant (128). Split-root experiments also showed that the expression of *AtIRT1* and *AtFRO2* is controlled both locally, by the root iron pool, as well as systemically, by a shoot-generated signal (119).

As mentioned above, the Arabidopsis *frd3*/*man1* mutant accumulates high levels of iron and shows constitutive expression of the iron-deficiency responses

(100, 129). *FRD3* is expressed in the roots under iron-replete and iron-deficient conditions and is up-regulated by low iron supply. FRD3 has been proposed to be involved in either the perception of a signal communicating the iron status of the shoot to the roots or in the transport of a signal molecule to an intracellular receptor (100). Recent data point to a role for FRD3 in transporting a compound in roots that plays a key role in iron delivery to the shoot in a usable form (101).

The expression of iron-regulated genes seems to be regulated by different pathways in shoots and roots. PsFRO1 is thought to be involved in iron reduction in root epidermal cells and in the shoot, and its mRNA accumulates under iron deficiency (24). However, whereas *PsFRO1* expression is constitutive in the roots of *dgl* and *brz* mutants, its pattern of expression is unchanged in the shoots of both mutants, indicating that expression of *PsFRO1* is affected by different signals in the shoots and the roots (24).

The *chln* mutant of tomato displays iron-deficiency symptoms such as interveinal chlorosis in young leaves; as in *dgl* and *brz*, components of the root iron uptake system, including proton extrusion and ferric reductase activity, are up-regulated in *chln*, indicating a defect in the iron sensing system (130). This mutant is unable to synthesize NA due to disruption of the single gene in tomato encoding NAS (51, 58, 59). Because NA binds iron, it is possible that it prevents iron precipitation and therefore maintains a pool of soluble iron or protects cells from iron toxicity. Moreover, Fe(II)-NA could, either directly or indirectly, act as a sensor for the iron status in the plant.

The tomato *fer* mutant is unable to turn on several root iron-deficiency responses and can be rescued by growth with high levels of Fe-HEDTA (131). Reciprocal grafting studies showed that FER is required in the roots and not in the shoots. *FER* has been shown to encode a protein containing a basic helix-loop-helix (bHLH) domain (132). FER seems to be required for sensing iron availability in the root and subsequently regulating the appropriate responses (41). The direct target genes of FER have not yet been identified (132). Studies of gene expression in *fer*, *chln,* and *fer*/*chln* double mutants indicated that expression of *LeIRT1* and *LeNRAMP3* is dependent on *FER*, that NA is required for proper regulation of these genes and that *FER* acts prior to *NAS* in the same pathway (41). In addition, *LeFRO1* expression was disrupted in the tomato *chln* and *fer* mutants, indicating that *FER* and *NAS* are involved in the regulation of *LeFRO1* expression in the roots (133). Recently, Colangelo and Guerinot showed that the *AtFIT1* gene of Arabidopsis encodes a bHLH transcription factor that regulates expression of *AtFRO2* and *AtIRT1* (134). *AtFRO2* mRNA is not detected in the *fit1* mutant, indicating that FIT1 directly or indirectly functions in the pathway that leads to expression of elevated levels of *AtFRO2* in iron-deficient roots. Interestingly, whereas *AtIRT1* mRNA levels are normal in the *fit1* mutant, IRT1 protein is undetectable in the *fit1* mutant, indicating that AtFIT1 regulates expression of *AtFRO2* and *AtIRT1* in different ways. Although tomato FER and Arabidopsis FIT1 appear to be functional homologues, expression of *LeFER* is constitutive, while expression of *AtFIT1* is induced by iron deficiency.

A microarray study in tomato reported that several genes induced by iron deficiency in the root could potentially encode proteins involved in iron signal transduction pathways (135). These include a leucine zipper transcription factor, a MAP kinase, a MAP kinase kinase, and a 14-3-3 protein. Induction of the

expression of these genes was seen in response to potassium, phosphate, and iron deficiencies, suggesting that a common signal transduction pathway may function to signal changes in plant nutrient status. Some of the same genes also show increased expression in response to iron deficiency in barley, suggesting that regulation of iron deficiency responses may be similar in Strategy I and Strategy II plants (66).

One possibility for a signaling molecule involved in the iron-deficiency response is nitric oxide (NO). Recently, NO-mediated ferritin regulation has been reported in Arabidopsis (136). NO was shown to act downstream of iron through the iron-dependent regulatory sequence (IDRS) of the *AtFer1* promotor (120, 136), suggesting that NO plays an important role in the regulation of iron homeostasis in plants. In addition, NO improves the availability of iron within plants (137). NO application rescues the phenotype of iron-inefficient mutants in plant species belonging to both strategies, suggesting that NO might have a general role in iron homeostasis (137).

The plant hormone ethylene is thought to participate in the induction of downstream morphological responses in Fe-deficient plants. The development of root hair and rhizodermal transfer cells is induced by exogenous application of ethylene or the ethylene precursor (ACC) (8, 138, 139). In addition, treatments with ethylene biosynthesis inhibitors cause a decrease in Fe(III) chelate reductase activity in iron-starved cucumber roots while ACC enhances such activity (140). These data suggest that ethylene might play a role in the iron-deficiency-induced signaling pathway. Nevertheless, studies using Arabidopsis mutants with altered hormone metabolism or signaling indicated that although ethylene and possibly auxin may be required for root hair formation in response to iron deficiency, induction of Fe(III) chelate reductase activity does not appear to be mediated by a signaling pathway that includes ethylene (9, 141). Finally, it is known that abscisic acid (ABA) is involved in a pathway leading to ferritin accumulation in response to iron (142).

POTENTIAL APPLICATIONS AND FUTURE DIRECTIONS

As described above, iron deficiency represents an enormous problem in human health. Thus, great interest surrounds the production of transgenic crop plants that contain elevated levels of bioavailable iron to help eliminate this problem (90). To create such plants, we must develop a complete understanding of iron metabolism in plants, including iron uptake, distribution, storage, and regulation. Recent work suggests that the development of crops with enhanced iron content is feasible.

Expression of the soybean ferritin gene in developing rice seeds leads to enhanced iron content in seeds (143). A second study showed that expression of alfalfa ferritin in the vegetative tissues of tobacco results in enhanced tolerance to oxidative stress and pathogens (144). Third, overexpression of soybean ferritin in tobacco resulted in induction of root iron-deficiency responses (145). Together, these studies suggest that manipulation of ferritin levels may prove to be an important component of strategies aimed at altering iron accumulation and homeostasis in plants. Traditionally, it has been thought that iron in the form of ferritin in plant foods is largely unavailable for absorption. However, recent work suggests that soybean ferritin may be much more available for absorption than

previously thought (146). Another group has used the combined approach of overexpressing ferritin, as well as phytase and a metallothionein-like protein (to increase bioavailability of iron) in rice grains (147). The resulting transgenic rice has a higher iron content; future experiments should determine whether or not the iron in these transgenic plants is, in fact, more bioavailable.

Recent work also suggests that it should be possible to create crop plants that are better able to be grown on iron-deficient soils. *AtFRO2* is subject to post-transcriptional regulation as described above; thus plants that overexpress *AtFRO2* from the CaMV 35S promoter do not accumulate elevated levels of iron (20). However, Fe-deficient *35S-FRO2* plants have enhanced levels of root Fe(III) chelate reductase activity relative to wild-type plants. As a result, *35S-FRO2* transgenic plants are remarkably tolerant of growth on low-Fe medium. This result suggests that overexpression of *FRO2* may yield plants capable of thriving on low-Fe soils. In addition, expression of barley NAAT genes in rice led to enhanced tolerance of low iron availability (65). These results are particularly exciting as approximately one-third of the world's soils are considered Fe deficient.

ACKNOWLEDGMENTS

We thank Mary Lou Guerinot for helpful comments. Work in the Connolly laboratory is supported by grants from the U.S. Department of Agriculture (9900598 and 0188925) and the National Science Foundation (0344305).

REFERENCES

1 Halliwell, B. and Gutteridge, J.M.C. (1992) FEBS Lett. 307, 108-112.
2 Guerinot, M.L. and Yi, Y. (1994) Plant Physiol. 104, 815-820.
3 World Health Organization. (2003) Available at: (http://www.who.int/nut/malnutrition_worldwide.htm#ida).
4 Römheld, V. and Marschner, H. (1986) Plant Physiol. 80, 175-180.
5 Eide, D.J. (1998) Ann. Rev. Nutr. 18, 441-69.
6 Eckhardt, U. and Buckhout, T.J. (1998) J. Exp. Bot. 9, 1219-1226.
7 Marschner, H., Kissel, M. and Römheld, V. (1986) J. Plant Nutr. 9, 695-713.
8 Schmidt, W. and Bartels, M. (1996) Plant Physiol. 110, 217-225.
9 Schmidt, W., Tittel, J. and Schikora, A. (2000) Plant Physiol. 122, 1109-1118.
10 Dell'Orto, M., Santi, S., De Nisi, P., Cesco, S., Varanini, Z., Zocchi, G. and Pinton, R. (2000) J. Exp. Bot. 51, 695-701.
11 Palmgren, M.G. (2001) Ann. Rev. Plant Physiol. Plant Mol. Biol. 52, 817-845.
12 Robinson, N.J., Proctor, C.M., Connolly, E.L. and Guerinot, M.L. (1999) Nature 397, 694-697.
13 Yi, Y. and Guerinot, M.L. (1996) Plant J. 10, 835-844.
14 Eide, D.J., Broderius, M., Fett, J. and Guerinot, M.L. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 5624-5628.

15 Vert, G., Grotz, N., Dedaldechamp, F., Gaymard, F., Guerinot, M.L., Briat, J.-F. and Curie, C. (2002) Plant Cell 14, 1223-1233.
16 Varotto, C., Maiwald, D., Pesaresi, P., Jahns, P., Francesco, S. and Leister, D. (2002) Plant J. 31, 589-599.
17 Henriques, R., Jasik, J., Klein, M., Martinoia, E., Feller, U., Schell, J., Pais, M.S. and Koncz, C. (2002) Plant Mol. Biol. 50, 587-597.
18 Buckhout, T.J., Bell, P.F., Luster, D.G. and Chaney, R.L. (1989) Plant Physiol. 90, 151-156.
19 Grusak, M.A., Welch, R.M. and Kochian, L.V. (1990) Plant Physiol. 94, 1353-1357.
20 Connolly, E.L., Campbell, N., Grotz, N., Prichard, C.L. and Guerinot, M.L. (2003) Plant Physiol. 133, 1102-1110.
21 Robinson, N.J., Sadjuga and Groom, Q.J. (1997) Plant Soil 196, 245-248.
22 Chanock, S.J., El Benna, J., Smith, R.M. and Babior, B.M. (1994) J. Biol. Chem. 269, 24519-24522.
23 Henderson, L.M. and Chappell, J.B. (1996) Biochim. Biophys. Acta 1273, 87-107.
24 Waters, B.M., Blevins, D.G. and Eide, D.J. (2002) Plant Physiol. 129, 85-94.
25 Korshunova, Y.O., Eide, D.J., Clark, W.G., Guerinot, M.L. and Pakrasi, H.B. (1999) Plant Mol. Biol. 40, 37-44.
26 Rogers, E.E., Eide, D.J. and Guerinot, M.L. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 12356-12360.
27 Connolly, E.L., Fett, J.P. and Guerinot, M.L. (2002) Plant Cell 14, 1347-1357.
28 Grotz, N., Fox, T., Connolly, E.L., Park, W., Guerinot, M.L. and Eide, D.J. (1998) Proc. Nat. Acad. Sci. U.S.A. 95, 7220-7224.
29 Guerinot, M.L. (2000) Biochim. Biophys. Acta 1465, 190-198.
30 Guerinot, M.L. and Eide, D.J. (1999) Curr. Opin. Plant Biol. 2, 244-249.
31 Maser, P., Thomine, S., Schroeder, J.I., Ward, J.M., Hirschi, K., Sze, H., Talke, I.N., Amtmann, A., Maathuis, F.J.M., Saunders, D., Harper, J.F., Tchieu, J., Gribskov, M., Persans, M.W., Salt, D.E., Kim, S.A. and Guerinot, M.L. (2001) Plant Physiol. 126, 1646-1667.
32 Vert, G., Briat, J.F. and Curie, C. (2001) Plant J. 26, 181-189.
33 Cohen, C.K., Garvin, D.F. and Kochian, L.V. (2004) Planta 218, 784-792.
34 Cohen, C.K., Fox, T.C., Garvin, D.F. and Kochian, L.V. (1998) Plant Physiol. 116, 1063-1072.
35 Eckhardt, U., Marques, A.M. and Buckhout, T.J. (2001) Plant Mol. Biol. 45, 437-448.
36 Bughio, N., Yamaguchi, H., Nishizawa, N., Nakanishi, H. and Mori, S. (2002) J. Exp. Bot. 53, 1677-1682.
37 Gitan, R.S., Shababi, M., Kramer, M. and Eide, D.J. (2003) J. Biol. Chem. 278, 39558-39564.
38 Andrews, N.C. (1999) Int. J. Biochem. Cell Biol 31, 991-994.
39 Curie, C., Alonso, J.M., Le Jean, M., Ecker, J.R. and Briat, J.F. (2000) Biochem. J. 347, 749-755.

40 Thomine, S., Wang, R., Ward, J.M., Crawford, N.M. and Schroeder, J.I. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 4991-4996.
41 Bercezky, Z., Wang, H.Y., Schiubert, V., Ganal, M. and Bauer, P. (2003) J. Biol. Chem. 278, 24697-24704.
42 Belouchi, A., Kwan, T. and Gros, P. (1997) Plant Mol. Biol. 33, 1085-1092.
43 Belouchi, A., Cellier, M., Kwan, T., Saini, H.S., Leroux, G. and Gros, P. (1995) Plant Mol. Biol. 29, 1181-1196.
44 Kaiser, B.N., Moreau, S., Castelli, J., Thomson, R., Lambert, A., Bogliolo, S., Puppo, A. and Day, D.A. (2003) Plant J. 35, 295-304.
45 Thomine, S., Lelievre, F., Debarbieux, E., Schroeder, J.I. and Barbier-Bygoo, H. (2003) Plant J. 34, 685-695.
46 Mori, S. (1999) Curr. Opin. Plant Biol. 2, 250-253.
47 Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S.L., Briat, J.F. and Walker, E.L. (2001) Nature 409, 346-349.
48 Takagi, S.I., Nomoto, K. and Takemoto, T. (1984) J. Plant Nutr. 7, 1-5.
49 Yehuda, Z., Shenker, M., Romheld, V., Marschner, H., Hadar, Y. and Chen, Y. (1996) Plant Physiol. 112, 1273-1280.
50 Takagi, S.I. (1976) Soil Sci. Plant Nutr. 22, 423-433.
51 Higuchi, K., Suzuki, K., Nakanishi, H., Yamaguchi, H., Nishizawa, N.K. and Mori, S. (1999) Plant Physiol. 119, 471-479.
52 Higuchi, K., Nishizawa, N., Romheld, V., Marschner, H. and Mori, S. (1996) J. Plant Nutr. 19, 1235-1239
53 Takagi, S., Nomoto, K. and Takemoto, T. (1984) J. Plant Nutr. 7, 469-477.
54 Takahashi, M., Yamaguchi, H., Nakanishi, H., Shioiri, T., Nishizawa, N.K. and Mori, S. (1999) Plant Physiol. 121, 947-956.
55 Yamaguchi, H., Nakanishi, H., Nishizawa, N.K. and Mori, S. (2000) J. Exp. Bot. 51, 2001-2007.
56 Nakanishi, H., Yamaguchi, H., Sasakuma, T., Nishizawa, N.K. and Mori, S. (2000) Plant Mol. Biol. 44, 199-207.
57 Kobayashi, T., Nakanishi, H., Takahashi, M., Kawasaki, S., Nishizawa, N.K. and Mori, S. (2001) Planta 212, 864-871.
58 Herbik, A., Koch, G., Mock, H.-P., Dushkov, D., Czihal, A., Thielmann, J., Stephan, U.W. and Baumlein, H. (1999) Eur. J. Biochem. 265, 231-239.
59 Ling, H.Q., Koch, G., Baulein, H. and Ganal, M.W. (1999) Proc. Nat. Acad. Sci. U.S.A. 96, 7098-7103.
60 Suzuki, K., Higuchi, K., Nakanishi, H., Nishizawa, N.K. and Mori, S. (1999) Soil Sci. Plant Nutr. 45, 993-1002.
61 Higuchi, K., Watanabe, Y., Takahashi, M., Kawasaki, S., Nakanishi, H., Nishizawa, N.K. and Mori, S. (2001) Plant J. 25, 159-167.
62 Mizuno, D., Higuchi, K., Sakamoto, T., Nakanishi, H., Mori, S. and Nishizawa, N.K. (2003) Plant Physiol. 132, 1989-1997.
63 Inoue, H., Higuchi, K., Takahashi, M., Nakanishi, H., Mori, S. and Nishizawa, N.K. (2003) Plant J. 36, 366-381.
64 Kanazawa, K., Higuchi, K., Nishizawa, N.K., Fushiya, S., Chino, M. and Mori, S. (1994) J. Exp. Bot. 45, 1903-1906.

65 Takahashi, M.T., Nakanishi, H., Kawasaki, S., Nishizawa, N.K. and Mori, S. (2001) Nat. Biotechnol. 19, 466-469.
66 Negishi, T., Nakanishi, H., Yazaki, J., Kishimoto, N., Fujii, F., Shimbo, K., Yamamoto, K., Sakata, K., Sasaki, T., Kikuchi, S., Mori, S. and Nishizawa, N. (2002) Plant J. 30, 83-94.
67 Von Wiren, N., Khodr, H. and Hider, R.C. (2000) Plant Physiol. 124, 1149-1157.
68 Sakaguchi, T., Nishizawa, N.K., Nakanishi, H., Yoshimura, E., and Mori, S. (1999) Plant Soil 215, 221-227.
69 Nishizawa, N. and Mori, S. (1987) J. Plant Nutr. 10, 1012-1020.
70 Basso, B., Bagnaresi, P., Bracale, M. and Soave, C. (1994) Maydica 39, 97-105.
71 Von Wirén, N., Mori, S., Marschner, H. and Römheld, V. (1994) Plant Physiol. 106, 71-77.
72 Roberts, L.A., Pierson, A.J., Panaviene, Z. and Walker, E.L. (2004) Plant Physiol. 135, 112-120.
73 Schaaf, G., Ludewig, U., Erenoglu, B.E., Mori, S., Kitahara, T. and Von Wiren, N. (2004) J. Biol. Chem. 279, 9091-9096.
74 Von Wiren, N., Klair, S., Bansal, S., Briat, J.F., Khodr, H., Shioiri, T., Leigh, R.A. and Hider, R.C. (1999) Plant Physiol. 119, 1107-1114
75 Pich, A. and Scholz, G. (1996) J. Exp. Bot. 294, 41-47.
76 Schmidke, I., Kruger, C., Frommichen, R., Scholz, G. and Stephan, U.W. (1999) Physiol. Plant. 106, 82-89.
77 Stephan, U.W. (2002) Plant and Soil 241, 19-25.
78 Scholz, G., Becker, R., Pich, A. and Stephan, U.W. (1992) J. Plant Nutr. 15, 1647-1665.
79 Stephan, U.W., Schmidke, I., V.W., S. and Scholz, G. (1996) BioMetals 9, 84-90.
80 Hell, R. and Stephan, U.W. (2003) Planta 216, 541-551.
81 Becker, R., Fritz, E. and Manteuffel, R. (1995) Plant Physiol. 108, 269-275.
82 DeBoer, A.H., Prius, H.B.A. and Zanstra, P.E. (1983) Planta 157, 259-266.
83 Kohler, B. and Raschke, K. (2000) Plant Physiol. 122, 243-254.
84 McKie, A.T., Marciani, P., Rolfs, A., Brennan, K., Wehr, K., Barrow, D., Miret, S., Bomford, A., Peters, T.J., Farzaneh, F., Hediger, M.A., Hentze, M.W. and Simpson, R.J. (2000) Mol. Cell. 5, 299-309.
85 Kawai, S., Kamei, S., Matsuda, Y., Ando, R., Kondo, S., Ishizawa, A. and Alam, S. (2001) Soil Sci. Plant Nutr. 47, 265-272.
86 Mori, S. and Nishizawa, N. (1987) Plant Cell Physiol. 28, 1081-1092.
87 Tiffin, L.O. (1966) Plant Physiol. 45, 280-283.
88 Abadia, J., Lopez-Millan, A.F., Rombolà, A.D. and Abadia, A. (2002) Plant and Soil 241, 75-86.
89 Pich, A., Scholz, G. and Stephan, U.W. (1994) Plant Soil 165, 189-196.
90 Grusak, M.A. and DellaPenna, D. (1999) Ann. Rev. Plant Physiol. Plant Mol. Biol. 50, 133-161.
91 Marschner, H. (1995) in Mineral Nutrition of Higher Plants, Academic Press, Boston, pp. 1-889.

92 Brüggemann, W., Maas-Kantel, K. and Moog, P.R. (1993) Planta 190, 151-155.
93 Luwe, M., Takahama, U. and Heber, U. (1993) Plant Physiol. 101, 969-976.
94 Brown, J.C., Cathey, H.M., Bennett, J.H. and Thimijan, R.W. (1979) Agron. J. 71, 1015-1021.
95 De la Guardia, M.D. and Alcantara, E. (1996) J. Exp. Bot. 47, 669-675.
96 Wintz, H., Fox, T., Wu, Y.–Y., Feng, V., Chen, W., Chang, H.–S., Zhu, T. and Vulpe, C. (2003) J. Biol. Chem. 278, 47644-47653.
97 Thimm, O., Essigmann, B., Kloska, S., Altmann, T and Buckhout, T.J. (2001) Plant Physiol. 127, 1030-1043.
98 Schmidt, W. (2003) Trends Plant Sci. 8, 188-193.
99 DiDonato, R.J., Roberts, L.A., Sanderson, T., Bosler Eisley, R. and Walker, E.L. (2004) Plant J. 39, 403-414.
100 Rogers, E.E. and Guerinot, M.L. (2002) Plant Cell 14, 1787-1799.
101 Green, L.S. and Rogers, E.E. (2004) Plant Physiol. 136, 2523-2531.
102 Grusak, M.A. (1994) Plant Physiol. 104, 649-655.
103 Marentes, E. and Grusak, M.A. (1998) Seed Sci. Res. 8, 367-375.
104 Kruger, C., Berkowitz, O., Stephan, U.W. and Hell, R. (2002) J. Biol. Chem. 277, 25062-25069.
105 Stephan, U.W. and Scholz, G. (1993) Physiol. Plant. 88, 522-529.
106 Koike, S., Inoue, H., Mizuno, D., Takahashi, M., Nakanishi, H., Mori, S. and Nishizawa, N.K. (2004) Plant J. 39, 415-424.
107 Takahashi, M., Terada, Y., Nakai, I., Nakanishi, H., Yoshimura, E., Mori, S. and Nishizawa, N.K. (2003) Plant Cell 15, 1263-1280.
108 Terry, N. and Abadia, J. (1986) J. Plant Nutr. 9, 609-646.
109 Bughio, N., Takahashi, M., Yoshimura, E., Nishizawa, N.K. and Mori, S. (1997) Plant Cell Physiol. 38, 101-105.
110 Shingles, R., North, M., and McCarty, R.E. (2002) Plant Physiol. 128, 1022-1030.
111 Briat, J.F., Lobréaux, S., Grignon, N. and Vansuyt, G. (1999) Cell. Mol. Life Sci. 56, 155-166.
112 Petit, J.M., Briat, J.F. and Lobréaux, S. (2001) Biochem. J. 359, 575-582.
113 Pich, A., Manteuffel, R., Hillmer, S., Scholz, G., and Schmidt, W. (2001) Planta 213, 967-976.
114 Lill, R. and Kispal, G. (2000) Trends Biochem. Sci. 25, 352-356.
115 Leon, S., Touraine, B., Ribot, C., Briat, J.F. and Lobreaux, S. (2003) Biochem. J. 371, 823-830.
116 Leon, S., Touraine, B., Briat, J.F. and Lobreaux, S. (2002) Biochem. J. 366, 557-564.
117 Touraine, B., Boutin, J.P., Marion-Poll, A., Briat, J.F., Peltier, G. and Lobréaux, S. (2004) Plant J. 40, 101-111.
118 Kushnir, S., Babiychuk, E., Storozhenko, S., Davey, M.W., Papenbrock, J., De Rucke, R., Engler, G., Stephan, U.W., Lange, H., Kispai, G., Lill, R. and Van Montagu, M. (2001) Plant Cell 13, 89-100.

119 Vert, G., Briat, J.F. and Curie, C. (2003) Plant Physiol 132, 796-804.
120 Petit, J.M., Van Wuytswinkel, O., Briat, J.F. and Lobréaux, S. (2001) J. Biol. Chem. 276, 5584-5590.
121 Wei, J. and Theil, E.C. (2000) J. Biol. Chem. 275, 17488-17493.
122 Okumura, N., Nishizawa, N.K., Umehara, Y., Ohata, T., Nakanishi, H., Yamaguchi, T., Chino, M. and Mori, S. (1994) Plant Mol. Biol. 25, 705-719.
123 Kobayashi, T., Nakayama, Y., Itai, R.N., Nakanishi, H., Yoshihara, T., Mori, S. and Nishizawa, N.K. (2003) Plant J. 36, 780-793.
124 Gitan, R.S. and Eide, D.J. (2000) Biochem. J. 346, 329-336.
125 Gitan, R.S., Luo, H., Rodgers, J., Broderius, M., and Eide, D.J. (1998) J. Biol. Chem. 273, 28617-28624.
126 Grusak, M.A. and Pezeshgi, S. (1996) Plant Physiol. 110, 329-334.
127 Maas, F.M., van de Wetering, D.A.M., van Beusichem, M.L. and Bienfait, H.F. (1988) Plant Physiol. 87, 167-171.
128 Schikora, A. and Schmidt, W. (2001) Protoplasma 218, 67-75.
129 Delhaize, E. (1996) Plant Physiol. 111, 849-855.
130 Stephan, U.W. and Grün, M. (1989) Biochem. Physiol. Pfanzen 185, 189-200.
131 Brown, J.C., Chaney, R.L. and Ambler, J.E. (1971) Physiol. Plant. 25, 48-53.
132 Ling, H.Q., Bauer, P., Bereczky, Z., Keller, B. and Ganal, M. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 13938-13943.
133 Li, L., Cheng, X. and Ling, H.-Q. (2004) Plant Mol. Biol. 54, 125-136.
134 Colangelo, E.P. and Guerinot M.L. (2004) Plant Cell. 16, 3400-3412.
135 Wang, Y.-H., Garvin, D.F. and Kochian, L.V. (2002) Plant Physiol. 130, 1361-1370.
136 Murgia, I., Delledonne, M. and Soave, C. (2002) Plant J. 30, 521-528.
137 Graziano, M., Beligni, M.V. and Lamattina, L. (2002) Plant Physiol. 130, 1852-1859.
138 Landsberg, E.C. (1996) Protoplasma 194, 69-80.
139 Schikora, A. and Schmidt, W. (2002) Plant Physiol. 125, 1679-1687.
140 Romera, F.J. and Alcántara, E. (1994) Plant Physiol. 105, 1133-1138.
141 Schmidt, W. and Schikora, A. (2001) Plant Physiol. 125, 2078-2084.
142 Briat, J.-F. and Lobréaux, S. (1997) Trends Plant Sci. 2, 187-193.
143 Goto, F., Yoshihara, T., Shigemoto, N., Toki, S. and Takaiwa, F. (1999) Nat. Biotechnol. 17, 282-286.
144 Deak, M., Horvath, G.V., Davletova, S., Torok, K., Sass, L., Vass, I., Barna, B., Kiraly, Z. and Dudits, D. (1999) Nat. Biotechnol. 17, 192-196.
145 Van Wuytswinkel, O., Vansuyt, G., Grignon, N., Fourcroy, P. and Briat, J.-F. (1999) Plant J. 17, 93-97.
146 Murray-Kolb, L.E., Welch, R., Theil, E.C. and Beard, J.L. (2003) Am. J. Clin. Nutr. 77, 180-184.
147 Lucca, P., Hurrell, R. and Potrykus, I. (2002) J. Am. Coll. Nutr. 21, 184S-190S.

SALT STRESS SIGNALING AND MECHANISMS OF PLANT SALT TOLERANCE

Viswanathan Chinnusamy,[1] Jianhua Zhu,[2] and Jian-Kang Zhu[2]

[1]Water Technology Centre
Indian Agricultural Research Institute
New Delhi 110012
India
[2]Institute for Integrative Genome Biology and Department of Botany and Plant Sciences
University of California
Riverside, California 92521

INTRODUCTION

Soil salinity of agricultural land has led to the breakdown of ancient civilizations. Even today it threatens agricultural productivity in 77 million hectares (mha) of agricultural land, of which 45 mha (20% of irrigated area) is irrigated and 32 mha (2.1% of dry land) is unirrigated (1). Salinization is further spreading in irrigated land because of improper management of irrigation and drainage. Rain, cyclones, and wind also add NaCl into the coastal agricultural land. Soil salinity often leads to the development of other problems in soils such as soil sodicity and alkalinity. Soil sodicity is the result of the binding of Na^+ to the negatively charged clay particles, which leads to clay swelling and dispersal. Hydrolysis of the Na-clay complex results in soil alkalinity. Thus, soil salinity is a major factor limiting sustainable agriculture.

The U.S. Department of Agriculture's (USDA) salinity laboratory defines saline soil as having electrical conductivity of the saturated paste extract (EC_e) of 4 dS m^{-1} (1 dS m^{-1} is approximately equal to 10 mM NaCl) or more. High

Genetic Engineering, Volume 27, Edited by J. K. Setlow

concentrations of soluble salts such as chlorides of sodium, calcium, and magnesium contribute to the high electrical conductivity of saline soils. NaCl contributes to most of the soluble salts in saline soil.

The development of salinity-tolerant crops is needed to sustain agricultural production. Conventional breeding programs aimed at improving crop tolerance to salinity have had limited success because of the complexity of the trait (2). Slow progress in breeding for salt-tolerant crops can be attributed to the poor understanding of the molecular mechanisms of salt tolerance. Understanding the molecular basis of plant salt tolerance will also help improve drought and extreme-temperature-stress tolerance, inasmuch as osmotic and oxidative stresses are common to these abiotic stresses. The salt-tolerant mechanisms of plants can be broadly described as ion homeostasis, osmotic homeostasis, stress damage control and repair, and growth regulation (3). This chapter reviews recent progress in understanding salt-stress signaling and breeding/genetic engineering for salt-tolerant crops.

EFFECT OF SALINITY ON PLANT DEVELOPMENT

Salinity affects almost all aspects of plant development, including germination, vegetative growth, and reproductive development. Soil salinity imposes ion toxicity, osmotic stress, nutrient (N, Ca, K, P, Fe, Zn) deficiency, and oxidative stress on plants. Salinity also indirectly limits plant productivity through its adverse effects on the growth of beneficial and symbiotic microbes. High salt concentrations in soil impose osmotic stress and thus limit water uptake from soil. Sodium accumulation in cell walls can rapidly lead to osmotic stress and cell death (1). Ion toxicity is the result of replacement of K^+ by Na^+ in biochemical reactions, and Na^+- and Cl^--induced conformational changes in proteins. For several enzymes, K^+ acts as cofactor and cannot be substituted by Na^+. High K^+ concentration is also required for binding tRNA to ribosomes and thus protein synthesis (3, 4). Ion toxicity and osmotic stress cause metabolic imbalance, which in turn leads to oxidative stress (5).

In terms of plant tolerance to salinity, plants are classified as halophytes, which can grow and reproduce under high salinity (> 400 mM NaCl), and glycophytes, which cannot survive high salinity. Most of the grain crops and vegetables are natrophobic (glycophytes) and are highly susceptible to soil salinity, even when the soil EC_e is < 4 dS m^{-1}. Crops such as bean (*Phaseolus vulgaris*), eggplant (*Solanum melongena*), corn (*Zea mays*), potato (*Solanum tuberosum*), and sugarcane (*Saccharum officinarum*) are highly susceptible, with a threshold EC_e of < 2 dS m^{-1}, whereas sugar beet (*Beta vulgaris*) and barley (*Hordeum vulgare*) can tolerate an EC_e of up to 7 dS m^{-1}. Sugar beets and barley are highly sensitive to salinity during germination but are highly tolerant during the later phases of crop development (6; http://www.ussl.ars.usda.gov/saltoler.htm). Soil type (particularly Ca^{2+} and clay content), rate of transpiration (which determines the amount of salt transported to the shoot for any given rate of salt uptake and loading to the xylem by roots), and radiation may further alter the salt tolerance of crops.

Salinity affects photosynthesis mainly through a reduction in leaf area, chlorophyll content and stomatal conductance, and to a lesser extent through a

decrease in photosystem II efficiency (7). The adverse effects of salinity on plant development are more profound during the reproductive phase. Figure 1 shows the adverse effect of salinity on vegetative and reproductive development and the differential sensitivity of yield components to different intensities of salt stress in rice (8). Wheat plants stressed at 100-175 mM NaCl showed a significant reduction in spikelets per spike, delayed spike emergence, and reduced fertility, which results in poor grain yield. However, Na^+ and Cl^- concentrations in the shoot apex of these wheat plants were below 50 and 30 mM, respectively, which is too low to limit metabolic reactions (9). Hence, the adverse effects of salinity may be attributed to the salt-stress effect on the cell cycle and differentiation.

Salinity arrests the cell cycle transiently by reducing the expression and activity of cyclins and cyclin-dependent kinases that results in fewer cells in the meristem, thus limiting growth (10). In *Arabidopsis*, the reduction in root meristem size and root growth during salt stress is correlated with the down-regulation of *CDC2a* (cyclin-dependent kinase), *CycA2;1,* and *CycB1;1* (mitotic cyclins) (11). The activity of cyclin-dependent kinase is diminished also by post-translational inhibition during salt stress (10). Salt stress-induced abscisic acid (ABA) may also mediate cell cycle regulation. ABA up-regulates the expression of the inhibitor of cyclin-dependent kinase *ICK1*, which is a negative regulator of CDC2a (12). Salinity adversely affects reproductive development by inhibiting microsporogenesis and stamen filament elongation, enhancing programmed cell death in some tissue types, ovule abortion, and senescence of fertilized embryos. In *Arabidopsis*, 200 mM NaCl stress causes as high as 90% ovule abortion (13).

GeneChip microarray transcriptome analysis of salt-stressed (100 mM NaCl stress for 3 hours) *Arabidopsis* plants revealed approximately 424 and 128 genes were upregulated (> 2 fold) in roots and leaves, respectively (14). In

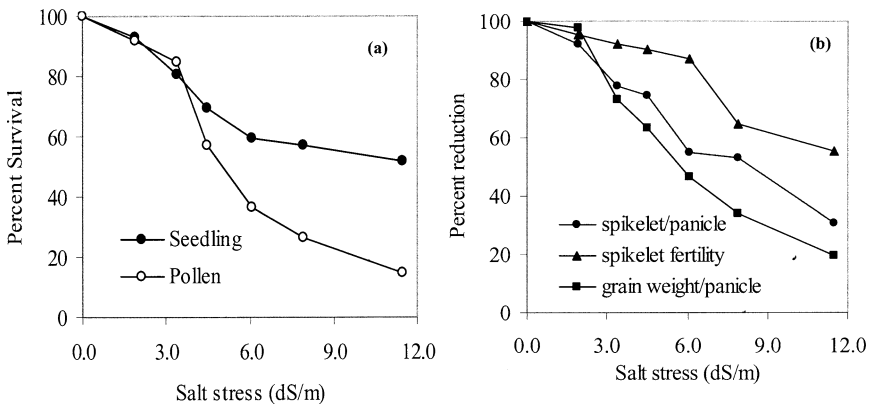


Figure 1. Effect of salinity stress on reproductive development in rice cv. M202. a) Seedling stage is more tolerant to salinity than reproductive stage; b) spikelet number is more sensitive to salinity than spikelet fertility. (This graph was drawn by using the data from Zeng and Shannon, 2000; one dS m^{-1} is approximately equal to 10 mM NaCl.)

Arabidopsis, cDNA microarray analysis showed that about 194 genes were up-regulated and about 89 were down-regulated by salt stress (15). In rice, of 1,700 cDNAs analyzed, approximately 57 genes were up-regulated by NaCl stress (16). Many of the NaCl up-regulated genes were also up-regulated by dehydration, cold, and ABA (14-16), which suggests that some of the stress responses are common to all these abiotic stresses. These results show that plant responses to salt stress are controlled by several genes and salt tolerance is a complex phenomenon.

PERCEPTION OF SALT STRESS

The ability of the plant to combat environmental stress is determined by the efficiency of the plant to sense the environmental stress and activate its defense machinery. Salt stress is perceived by plants as ionic and osmotic stresses. Excess Na^+- and Cl^--induced conformational changes in protein structure and membrane depolarization can lead to the perception of ion toxicity. Plasma membrane proteins, ion transporters, and/or Na^+-sensitive enzymes have been hypothesized as sensors of toxic Na^+ concentrations in extracellular and intracellular sites. Many transporters with long cytoplasmic tails similar to that of SOS1 (*S*alt *O*verly *S*ensitive 1, the plasma membrane Na^+/H^+ antiporter) have been implicated as being sensors of the molecule transported by that transporter. Similar to the sugar permease BglF in *E. coli* and the ammonium transporter Mep2p in yeast, SOS1 has been proposed to be one of the potential sensors of Na^+ ions in plants (17). Another potential candidate sensor is a Na^+-K^+ co-transporter of *Eucalyptus camaldulensis*. This transporter showed increased ion uptake under hypo-osmotic conditions when expressed in *Xenopus laevis* oocytes (18).

Salinity-imposed osmotic stress leads to cell turgor loss and cell volume change. Hence, the potential sensors of osmotic stress include membrane-associated stretch-activated channels, cytoskeleton (microtubules and microfilaments), and transmembrane protein kinases, such as two-component histidine kinases. One of the putative sensors of osmotic stress in *Arabidopsis* is the hybrid two-component histidine kinase *ATHK1* (19). The proposed role of AtHK1 in salt tolerance is discussed later in this chapter.

SECOND MESSENGERS

The ameliorative effects of Ca^{2+} in maintaining plant growth under salinity (20) and Ca^{2+}-induced ion channel discrimination against Na^+ (21) have been well known for a long time. In addition to its effect on preventing Na^+ entry into cells, Ca^{2+} acts as a signaling molecule in salt stress signaling (22, 23). Cytosolic Ca^{2+} oscillations during salt stress are regulated through the activities of mechanosensitive and ligand-gated Ca^{2+} channels on the plasma membrane, endoplasmic reticulum and vacuole (3, 17). Excess Na^+-induced membrane depolarization may activate mechanosensitive Ca^{2+} channels to generate Ca^{2+} signature under salt stress (4, 17). Pharmacological studies and genetic analysis have shown the involvement of inositol 1,4,5-trisphosphate (IP_3)-gated Ca^{2+} channels in the regulation of Ca^{2+} signature during salt stress (24-26). The *FRY1* locus of *Arabidopsis* encodes an inositol polyphosphate 1-phosphatase, which catabolizes

IP_3. The *Arabidopsis fry1* mutant is impaired in inositol polyphosphate 1-phosphatase and thus exhibits impaired ABA-induced IP_3 transients. The *fry1* mutation leads to sustained accumulation of IP_3 and hypersensitivity to ABA, cold, and salt stress. Thus, IP_3 plays a crucial role in cytosolic Ca^{2+} oscillations during ABA, salt, and cold stress signaling (26). Salinity stress also leads to synthesis of the plant stress hormone ABA (27-29) and accumulation of reactive oxygen species (ROS) (5). Calcium and/or H_2O_2 act as second messengers of ABA-induced stomatal closure and gene expression under abiotic stresses (30, 31). Transient expression analysis revealed that IP_3 and cyclic ADP ribose (cADPR)-gated calcium channels are involved in ABA-induced cytosolic Ca^{2+} oscillations (32). ABA induces the expression and activity of ADP-ribosyl (ADPR) cyclase, which synthesizes cADPR (33). Involvement of a heterotrimeric GTP-binding (G)-protein has been demonstrated in ABA signal transduction during guard cell regulation (34). Since ABA synthesis is induced under salinity, the G-protein-associated receptors may also elicit Ca^{2+} signatures during salinity stress. Salt stress-induced Ca^{2+} signatures are then sensed and transduced by calcium sensor proteins, namely SOS3 and SOS3-like calcium binding proteins (SCaBPs), calcium-dependent protein kinases (CDPKs), and calmodulins (CaMs).

ION HOMEOSTASIS

Plants achieve ion homeostasis by restricting the uptake of toxic ions, maintaining the uptake of essential ions and compartmentalization of toxic ions into the vacuole of specific tissue types. In most crop plants, Na^+ is the primary cause of ion toxicity, and hence, management of cellular Na^+ concentration is critical for salt tolerance (4). Sodium ions can be kept below the toxic level in the cytosol by 1) restricting Na^+ entry at the root cortex cells, 2) excreting Na^+ from root cells into soil, 3) retrieving Na^+ from the transpirational xylem stream to recirculate it to the roots, 4) storing Na^+ in the vacuole of mature cells, and 5) excreting Na^+ through salt glands (3). Among these mechanisms, Na^+ excretion through salt glands is important only in halophytes. Biochemical, electrophysiological, and molecular genetic evidence show that the SOS pathway plays a crucial role in the regulation of cellular and whole plant ion homeostasis (Figure 2) (17).

SODIUM UPTAKE

Restricting Na^+ entry into the root cells and then into the transpirational stream is critical to prevent a buildup of toxic levels of salt in the shoot. Both glycophytes and halophytes must exclude about 97% of the Na^+ present in the soil at the root surface to prevent toxic levels of Na^+ accumulation in the shoots (35). Sodium entry into the transpirational stream depends upon the amount of Na^+ uptake by Na^+ and nonspecific cation transporters and the proportion of water entry in the apoplastic/bypass pathway into the xylem. Na^+ from the soil gains the initial entry into the cells of the root epidermis and cortex. The Casparian strip in the endodermis plays a crucial role in preventing apoplastic Na^+ influx into the root stele. Compared with *Arabidopsis*, halophytes such as salt cress (*Thellungiella halophila*) develop both an extra endodermis and a cortex cell layer

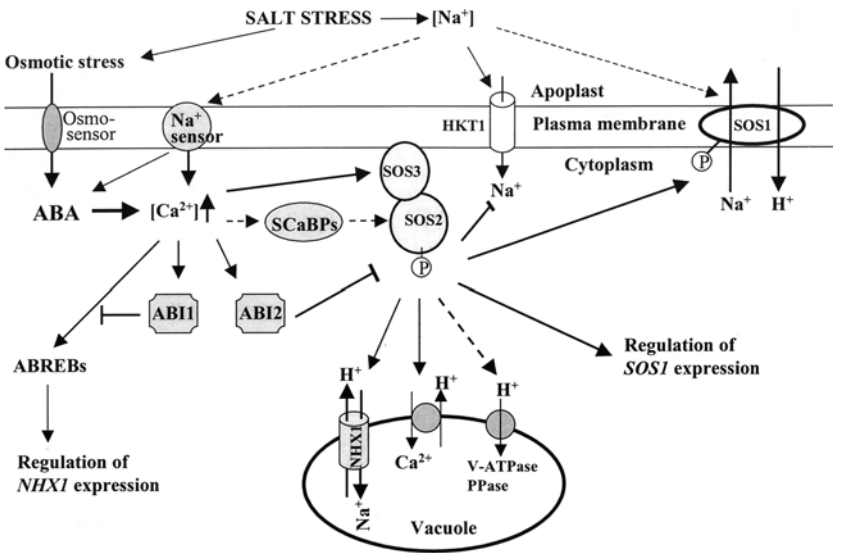


Figure 2. SOS signaling pathway regulates ion homeostasis during salt stress in *Arabidopsis*. The *S*alt *O*verly *S*ensitive 3 (SOS3) perceive the salt stress–induced Ca2+ signals and activate SOS2 kinase. Activated SOS2 kinase phosphorylates SOS1, a plasma membrane Na+/H+ antiporter. Phosphorylated SOS1 transports Na+ out of cytosol. The *SOS1* transcript level and perhaps Na+ transport through Na+ transporter HKT1 are also regulated by SOS3-dependent SOS2 kinase. The SOS2 kinase also activates tonoplast Na+/H+ antiporter (NHX1) that sequesters Na+ into the vacuole and vacuolar H+/Ca2+ antiporter (VCX1). Activation of NHX1 and VCX1 by SOS2 are SOS3-independent and probably regulated through SOS3-like Ca2+ Binding Proteins (SCaBPs). ABI1 regulate the gene expression of NHX1 through ABFs (ABA responsive element Binding Factors). ABI2 interact with SOS2 and negatively regulate ion homeostasis either by inhibiting SOS2 kinase activity or the activities of SOS2 targets.

in roots (36). In maize seedlings stressed at 200 mM NaCl, the Casparian strip radial width was increased by 47% compared with control seedlings (37). This feature may help to reduce the Na^+ entry into the transpirational stream. In crops such as rice, water entry into the xylem through the bypass pathway accounts for all the Na^+ buildup in the shoots, whereas in crops such as wheat, transport protein-mediated Na^+ uptake accounts for most of the Na^+ buildup in the shoots (38). Silica deposition and polymerization of silicate in the endodermis and rhizodermis blocks Na^+ influx through the apoplastic pathway in the roots of rice (39). Regulation of these anatomical and morphological changes in root development during salt stress needs further understanding.

Sodium uptake is mediated by both voltage-dependent and -independent cation channels. The role of voltage-independent cation channels in Na^+ uptake is poorly understood. Voltage-dependent cation channels such as K^+ inward rectifiers (HKT, HAK, and KUP) mediate Na^+ uptake into root cells. Sodium competes with K^+ uptake through Na^+-K^+ co-transporters and may also block the K^+-specific transporters of root cells (17). Expression studies in yeast cells revealed that high-affinity K^+-uptake activity of both *Arabidopsis* AtKUP1 and

barley HvHAK1 is inhibited by millimolar concentrations of Na^+ (40, 41). Cellular K^+ concentration can be maintained by the activity/expression of inward-rectifying K^+-specific transporters under high salinity. Salt stress, as well as K^+ starvation, up-regulates the expression of *Mesembryanthemum crystallinum* (common ice plant) high-affinity K^+ transporter genes (*McHAK*s). McHAKs specifically mediate K^+ uptake and show high discrimination for Na^+ at high salinity (42). In contrast, high-affinity K^+ transporters (HKTs) of wheat (43, 44), *Arabidopsis* (45), and *Eucalyptus* (18) act as low-affinity Na^+ transporters when expressed in *Xenopus* oocytes. HKT transporters of *Eucalyptus camaldulensis* (evergreen tree) and wheat possess Na^+:K^+ symport activity but mediate mainly Na^+ transport under high salinity (18, 43). The expression of *OsHKT1* is significantly down-regulated in salt-tolerant rice cv* Pokkali as compared with salt-sensitive rice cv IR29 during 150 mM NaCl stress (46). Transgenic wheat plants expressing antisense wheat *HKT1* showed significantly less $^{22}Na^+$ uptake and enhanced growth under high salinity as compared with control plants (47). This evidence suggests that HKT1 homologues contribute to Na^+ influx during salt stress and down-regulation of *HKT1* may help limit Na^+ influx to roots.

In yeast, *HAL1* and *HAL3* regulate the expression of P-type ATPase, Na^+ efflux, and K^+ uptake. Transgenic overexpression of the yeast *HAL1* gene enhanced salt tolerance of melon shoots *in vitro* (48), tomato (49, 50), and watermelons (51). Transgenic tomato plants overexpressing yeast *HAL1* showed increased K^+ accumulation. Irrigation with 35 mM NaCl to plants till maturity decreased the control plant fruit yield by 57.5%, whereas transgenic plants showed 24–42% decreased fruit yield. However, under normal growing conditions, the transgenic lines were less productive than the wild type (50). Overexpression of the *Arabidopsis HAL3a* gene also enhanced the salt tolerance of transgenic *Arabidopsis* (52).

Electrophysiological evidence suggests that cyclic nucleotides (cAMP and cGMP) may minimize Na^+ influx into the cell by down-regulating voltage-independent cation channels in *Arabidopsis* (53). Exposure of *Arabidopsis* plants to salt and osmotic stress results in increased cytosolic cGMP concentration within 5 seconds (54). Pyridoxal-5-phosphate is a co-factor for transaminases involved in the biosynthesis of aminoacids that are precursors for nucleotide biosynthesis. The *Arabidopsis sos4* mutant defective in a pyridoxal kinase gene showed hypersensitive root growth under NaCl and KCl stress and accumulated more Na^+ but less K^+ than the wild type. Pyridoxal-5-phosphate and its derivatives act as ligands for P2X receptor ion channels in animals (55). Pyridoxal-5-phosphate may regulate Na^+ efflux by SOS1, because SOS1 contains a putative pyridoxal-5-phosphate binding domain (3). Thus, regulation of K^+ and Na^+ uptake by pyridoxal-5-phosphate and cyclic nucleotides may help in plant salt tolerance. Signaling pathways that regulate Na^+ and K^+ uptake by higher plants during salinity need further study.

* cv: cultivar.

SODIUM EFFLUX

Sodium efflux from root cells is a frontline defense that prevents the accumulation of toxic levels of Na^+ in the cytosol and Na^+ transport to the shoot. Plasma membrane Na^+/H^+ antiporters pump out Na^+ from root cells. In *Arabidopsis*, the plasma membrane Na^+/H^+ antiporter SOS1 mediates Na^+ efflux, and its activity is regulated by the SOS3-SOS2 kinase complex during salt stress (Figure 2) (17). Salt stress-induced Ca^{2+} signatures are sensed by SOS3. SOS3 has three calcium-binding EF hands and an N-myristoylation motif and shows sequence similarity to the calcineurin B subunit of yeast and neuronal Ca^{2+} sensors of animals (56, 57). Calcineurin is a protein phosphatase (PP2B) that regulates salt tolerance in yeast. SOS3 and SOS3-like calcium-binding proteins (SCaBPs) identified in *Arabidopsis* differ from yeast calcineurin structurally and functionally. SCaBPs do not have a calcineurin A subunit catalytic domain. Unlike calcineurin, which activates protein phosphatases, SOS3 activates the Ser/Thr protein kinase during salt stress. Thus, SOS3 and SCaBPs are a new class of Ca^{2+} sensor proteins in higher plants. Mutations that disrupt either Ca^{2+} binding (*sos3-1*) or myristoylation (G2A) of SOS3 cause salt stress hypersensitivity in *Arabidopsis* (57). SOS3 binds Ca^{2+} with low affinity as compared with other Ca^{2+}-binding proteins such as caltractin and calmodulin (57). The differences in the affinity of these Ca^{2+} sensors may be employed by cells to distinguish various Ca^{2+} signals. SOS3 transduces the salt stress signal by activating SOS2, a Ser/Thr protein kinase with an N-terminal kinase catalytic domain that is similar to that of yeast sucrose nonfermenting 1 (SNF1) and animal AMP-activated kinase (AMPK), and has a unique C-terminal regulatory domain. The C-terminal regulatory domain of SOS2 consists of an autoinhibitory FISL motif (58). Under normal cellular conditions, the catalytic and regulatory domains of SOS2 interact with each other, likely preventing substrate phosphorylation by blocking substrate access. Yeast two-hybrid and *in vitro* binding assay have shown that in the presence of Ca^{2+}, SOS3 binds to and activates the SOS2 kinase (59). The FISL motif in the regulatory domain of SOS2 is necessary and sufficient for interacting with SOS3, and deletion of this FISL motif constitutively activates SOS2. Replacing Thr^{168} in the kinase domain by Asp also results in a constitutively active SOS2 kinase (60).

Molecular genetic analyses led to the identification of targets of the SOS3-SOS2 regulatory pathway. One of the targets of the SOS pathway is SOS1. SOS1 has significant protein sequence homology and conserved domains similar to that of the plasma membrane Na^+/H^+ antiporter from bacteria, fungi, and animals. The expression of *SOS1* is ubiquitous but stronger in epidermal cells surrounding the root tip and in parenchyma cells bordering the xylem. The expression of the SOS1::GFP fusion protein and anti-SOS1 antibody confirmed that SOS1 is localized in the plasma membrane of root and leaf cells (61-63). *sos1* mutant plants show hypersensitivity to salt stress (100 mM NaCl) and accumulate more Na^+ in shoots than do wild-type plants (61). Isolated plasma membrane vesicles from *sos1* mutants showed significantly less inherent as well as salt stress-induced Na^+/H^+ antiporter activity than did vesicles from the wild type (64). This evidence shows that SOS1 functions as a Na^+/H^+ antiporter on the plasma mem-

brane and plays a crucial role in sodium efflux from the root cells. Indeed, transgenic *Arabidopsis* plants overexpressing *SOS1* exhibited lower levels of Na^+ in the xylem transpirational stream and in the shoot than wild-type plants and enhanced salt tolerance. Transgenic plants grew, bolted and flowered with increasing concentrations of salt stress (50-200 mM NaCl), whereas control plants become necrotic and did not bolt (65). The expression level of *SOS1* is also significantly higher in salt cress (*T. halophila*) than in *Arabidopsis*, even in the absence of salt stress (66).

The Na^+/H^+ exchange activity of SOS1 is regulated by the SOS3-SOS2 complex under salt stress. Isolated plasma membrane vesicles from *sos3* and *sos2* mutants showed significantly less Na^+/H^+ exchange activity than that of wild-type plants. Consistent with this finding, these mutants also accumulate higher levels of Na^+, similar to those accumulated by the *sos1* mutant. However, the addition of activated SOS2 is sufficient to rescue the Na^+/H^+ exchange activity of plasma membrane vesicles from *sos3* and *sos2* mutants (64, 67). The SOS3-SOS2 kinase complex phosphorylates the SOS1 protein and activates SOS1 Na^+/H^+ antiporter activity (67). *SOS1* up-regulation during salt stress is also under the regulatory control of the SOS pathway, as shown by the impaired expression of *SOS1* in salt-stressed *sos2* and *sos3* mutants (60). Overexpression of the active form (Thr168 to Asp mutation) of *SOS2* under the control of the CaMV 35S promoter (*35S::T/DSOS2*) rescued the *sos2* and *sos3* mutants under salinity conditions. Transgenic *Arabidopsis* expressing *35S::T/DSOS2* showed enhanced SOS1 transporter activity and better vegetative and reproductive growth than wild-type plants when grown in soil irrigated with 200 mM NaCl (68). Coexpression of *SOS1*, *SOS2,* and *SOS3* rescued yeast cells deficient in Na^+ exchangers. Coexpression of *SOS2* and *SOS3* significantly increased SOS1-dependent Na^+ tolerance in the yeast mutant (67). This evidence demonstrates that SOS3 senses the salt-stress induced Ca^{2+} signals and activates SOS2 kinase, which in turn regulates the Na^+/H^+ exchange activity and expression of *SOS1* (Figure 2) (17).

SODIUM COMPARTMENTATION

Soil salinity decreases soil water potential, which leads to osmotic stress. To maintain water uptake during osmotic stress, plants have evolved a mechanism known as osmotic adjustment. Osmotic adjustment is active accumulation of solutes such as inorganic ions (Na^+ and K^+) and organic solutes (proline, betaine, polyols, and soluble sugars). Vacuolar sequestration of Na^+ is an important and cost-effective strategy for osmotic adjustment and at the same time reduces the cytosolic Na^+ concentration during salinity. Vacuolar Na^+/H^+ antiporters use the proton gradient generated by vacuolar H^+-adenosine triphosphatase (H^+-ATPase) and H^+-inorganic pyrophosphatase (H^+-PPase) for Na^+ sequestration into the vacuole. Hence, coordinated regulation of the Na^+/H^+ antiporters, H^+-ATPase and H^+-PPase is crucial for salt tolerance. Salt stress induces tonoplast H^+-ATPase and H^+-PPase activities (69). Transgenic *Arabidopsis* plants overexpressing *AVP1* (H^+-PPase) showed enhanced sequestration of Na^+ into the vacuole and maintained higher relative leaf water content and enhanced salt and drought stress tolerance as compared with the wild type

(70). NO-mediated signaling is implicated in the activation of plasma membrane H^+-ATPase (71), but the regulators of tonoplast H^+-ATPases and H^+-PPase are yet to be identified.

Vacuolar Na^+ sequestration is further regulated at the level of expression and activity of tonoplast Na^+/H^+ antiporters (NHXs). Expression of *NHX1* is induced by salinity and ABA in *Arabidopsis* (72, 73), rice (74), and cotton (75). The expression level of *NHX1* is correlated with genotypic differences in salt tolerance in cotton (75). Complementation studies showed that *AtNHX1* (72) and *OsNHX1* (69) could complement the yeast *nhx1* mutant. Transgenic *Arabidopsis* plants overexpressing *AtNHX1* showed significantly higher salt (200 mM NaCl) tolerance than wild-type plants (76). Transgenic tomato plants overexpressing *AtNHX1* were able to grow and produce fruits in the presence of very high salt concentrations (200 mM NaCl) at which wild-type plants did not survive. The yield and fruit quality of transgenic tomato plants under salt stress were equivalent to that of control plants under non-stress conditions (77). Similar results were obtained with transgenic canola (*Brassica napus*) overexpressing *AtNHX1* (78). These tomato and canola plants accumulated high concentrations of Na^+ in older leaves but not in reproductive parts (77, 78). Inspired by these results, transgenic rice plants overexpressing *Atriplex gmelini NHX1* (79), rice overexpressing *OsNHX1* (69), and tobacco overexpressing *Gossipium hirsutum NHX1* (75) were engineered. These transgenic plants showed better salt tolerance than control plants in the vegetative stage. However, transgenic rice plants overexpressing *OsNHX1* did not show a K^+-to-Na^+ ratio significantly different from that of control plants (69). Analysis of salt tolerance of the *osnhx1* null mutant of rice (69) may shed further light on the role of NHX1 in salt tolerance.

The SOS pathway and ABA regulate *AtNHX1* gene expression and its antiporter activity under salt stress. The promoter of *AtNHX1* contains putative ABA responsive elements (ABRE) between –736 and –728 from the initiation codon. *AtNHX1* expression under salt stress depends in part on ABA biosynthesis and ABA signaling through ABI1, because the salt stress-induced upregulation was reduced in ABA-deficient mutants (*aba2-1* and *aba3-1*) and the ABA-insensitive mutant, *abi1-1* (73). In *G. hirsutum*, *GhHNX1* expression is induced by ABA and appears to be regulated by MYB/MYC-type transcription factors (75). Analysis of the tonoplast Na^+/H^+-exchange activity in wild-type and *sos* mutants (*sos1*, *sos2* and *sos3*) revealed that SOS2 also regulates tonoplast Na^+/H^+-exchange activity. The impaired tonoplast Na^+/H^+-exchange activity from isolated *sos2* tonoplasts could be restored to the wild-type level by the addition of activated SOS2 protein. Inasmuch as the Na^+/H^+-exchange activity is unaffected in the *sos3* mutant, regulation of tonoplast Na^+/H^+-exchange activity by SOS2 is independent of SOS3 (80). SOS2 has been found to interact with plant calcium sensor proteins such as SOS3, SCaBP1, SCaBP3, SCaBP5, and SCaBP6 (81). One of these SCaBPs may signal SOS2 to regulate tonoplast Na^+/H^+-exchange activity (Figure 2) (80). SOS2 also has an additional SOS3-independent role in regulating the vacuolar H^+/Ca^{2+} antiporter VCX1, which plays a crucial role in regulating the duration and amplitude of cytosolic Ca^{2+} oscillations (82).

SODIUM TRANSPORT FROM SHOOTS TO ROOTS

Many of the glycophytes have limited ability to sequester Na^+ in leaf vacuoles. Therefore, these plants recirculate excess Na^+ from the leaf to the root. Sodium transport from shoots to roots is probably mediated by SOS1 and HKT1 in *Arabidopsis*. Under salt stress (100 mM NaCl), Na^+ accumulation in shoots of *sos1* mutant plants was greater than that of the wild type. Strong expression of *SOS1* in cells bordering the xylem suggests that SOS1 mediates either Na^+ release into or Na^+ retrieval from the xylem stream, depending on salt stress intensity, and thus is critical for controlling long-distance Na^+ transport from roots to shoots (62). Comparison of the expression pattern of *HKT1* in wheat and *Arabidopsis* revealed that AtHKT and wheat HKT1 might have different functions. *AtHKT1* is mainly expressed in phloem tissues but not in root peripheral cells, whereas wheat HKT1 is localized to the root epidermis and leaf vasculature. The sodium overaccumulation in shoots *2-1* (*sas2-1*) mutant of *Arabidopsis* showed significantly higher shoot Na^+ content but lower root Na^+ content and Na^+ concentration in the phloem sap exuding from leaves than the wild type. *sas2-1* mutation impaired *AtHKT1* and thus its Na^+ transport activity in *Xenopus* oocytes (83). T-DNA mutation in the *AtHKT1* gene has resulted in higher shoot Na^+ content and lower root Na^+ content than that in the wild type (84). Moreover AtHKT1 does not show significant K^+ transport activity in *Xenopus* oocytes. A single-point mutation, Ser-68 to glycine, was sufficient to restore K^+ permeability to AtHKT1 (84). These results show that AtHKT1 probably mediates Na^+ loading into the phloem sap in shoots and unloading in roots and thus helps to maintain a low Na^+ concentration in shoots (83). The *Arabidopsis athkt1*Δ mutation suppresses the salt hypersensitivity and K^+-deficient phenotype of *sos3* (85). Hence, the SOS pathway may regulate and coordinate the activities of AtHKT1 and SOS1 to control Na^+ transport from shoots to roots.

Salt stress–induced ABA accumulation, in addition to cytosolic Ca^{2+}, may also regulate the SOS pathway through the ABI2 protein phosphatase 2C. ABI2 interacts with the protein phosphatase interaction motif of SOS2. This interaction is abolished by the *abi2-1* mutation, which enhances the tolerance of seedlings to salt shock (150 mM NaCl) and causes ABA insensitivity. Hence, the wild-type ABI2 may negatively regulate salt tolerance by inactivating SOS2 or the SOS2-regulated ion channels such as HKT1, Na^+/H^+ antiporters, SOS1 and NHX1 (Figure 2) (86).

Transgenic manipulations of ion homeostasis have demonstrated the possibilities of genetic engineering salt-tolerant crop plants. Although multiple genes govern salt-stress tolerance, significant increases in salt tolerance have been achieved by single-gene manipulations, as revealed by *SOS1*-(65) and *NHX1*-(76-78) overexpressing transgenics. These transgenics were able to grow and flower at a salt concentration of 200 mM NaCl (~20 dS m^{-1}), which is lethal to wild-type plants. Most crop plants are susceptible to this concentration of salinity (6). In addition, these transgenics do not produce any obvious growth abnormalities or change in the quality of the consumable product, as shown by

NHX1-overexpressed transgenic tomato and *Brassica* (77, 78). Hence, genetic engineering for ion homeostasis by tissue-specific overexpression of *SOS1*, *NHX1,* and their positive regulator, the active form of SOS2, will help improve the salt tolerance of crop plants.

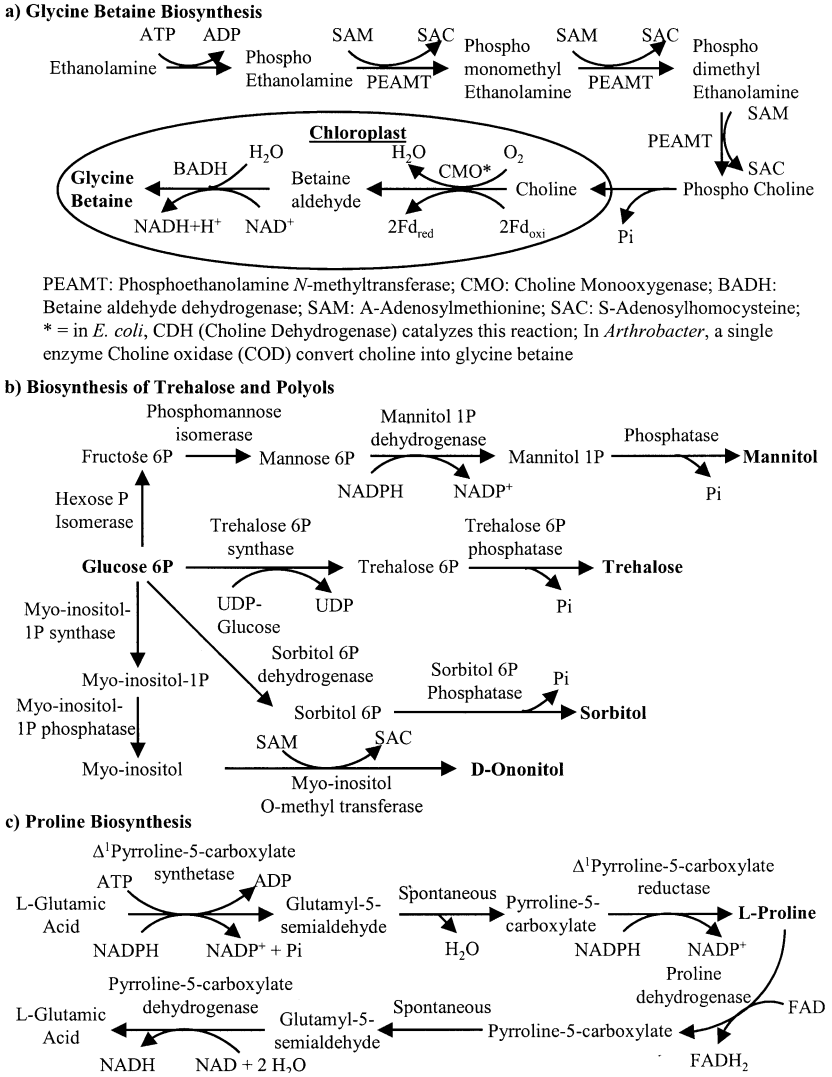
STRESS DAMAGE PREVENTION AND REPAIR

Stress damage prevention and repair pathways are necessary for cell survival at metabolically inhibitory levels of ionic or osmotic stresses. These strategies may include osmotic adjustment, osmoprotectant accumulation, oxidative stress management, induction of stress proteins (LEA-type proteins, chaperonin, etc.), and other physiological adaptations such as modifications in root and shoot growth and transpiration.

OSMOPROTECTANTS

Plants accumulate organic osmolytes such as proline, betaine, polyols, sugar alcohols, and soluble sugars to tolerate osmotic stress. These organic solutes protect plants from abiotic stress by 1) osmotic adjustment, which helps in turgor maintenance; 2) detoxification of reactive oxygen species; and 3) stabilization of the quaternary structure of proteins (87). Polyols and proline act as antioxidants (88). Proline also stabilizes subcellular structures (membranes and proteins) and buffers cellular redox potential under stress. Glycine betaine and trehalose stabilize the quaternary structures of proteins and highly ordered state of membranes. Glycine betaine also reduces lipid peroxidation during salinity stress. Hence, these organic osmolytes are known as osmoprotectants (3, 87-89). In addition to these organic osmoprotectants, polyamines also play a significant role in salt-stress tolerance. Mutations that impair arginine decarboxylase (ADC catalyzes the first committed step in polyamine biosynthesis) result in salt hypersensitivity (90, 91). Genes involved in osmoprotectant biosynthesis are upregulated under salt stress, and the concentrations of accumulated osmoprotectants correlate with osmotic stress tolerance (3, 89). Halophytes such as *T. halophila* accumulate significantly higher concentrations of proline than *Arabidopsis*, even under nonstress conditions (66). Genetic analysis of the *Arabidopsis t365* mutant impaired in the S-adenosyl-L-methionine:phosphoethanolamine N-methyltransferase (*PEAMT*) gene involved in glycine betaine biosynthesis (Figure 3) showed hypersensitivity to salt stress (92). Thus, glycine betaine accumulation is critical for salt tolerance. Several efforts have been made to engineer salt and other abiotic stress resistance in plants through genetic manipulation of osmoprotectant metabolism. The pathways of various osmoprotectant biosynthesis are shown in Figure 3. Genes of these pathways that are employed in genetic engineering for salt tolerance are briefly reviewed in Table 1.

Genetically engineered overproduction of compatible osmolytes in transgenic plants such as *Arabidopsis*, tobacco, rice, wheat, and *Brassica* has also been shown to enhance stress tolerance at the vegetative stage, as measured by germination, seedling growth, survival, recovery and photosystem II yield (Table 1).



PEAMT: Phosphoethanolamine *N*-methyltransferase; CMO: Choline Monooxygenase; BADH: Betaine aldehyde dehydrogenase; SAM: A-Adenosylmethionine; SAC: S-Adenosylhomocysteine; * = in *E. coli*, CDH (Choline Dehydrogenase) catalyzes this reaction; In *Arthrobacter*, a single enzyme Choline oxidase (COD) convert choline into glycine betaine

Figure 3. Osmoprotectant metabolism. Genes encoding for many of these enzymes have been employed for genetic engineering osmoprotectant accumulation in plants (Table 1).

Only in a few cases was salinity-stress tolerance of transgenics examined at the reproductive stage of the plant (94, 101, 108, 111, 112). In most cases, the contribution of the engineered osmoprotectant concentration to osmotic adjustment was not measured, or its contribution to osmotic adjustment was low. Abiotic stress tolerance of these transgenics was attributed to the osmoprotectant effect of these solutes (Table 1). Further, compartmentation of these osmoprotectants

Table 1. Metabolic engineering of osmoprotectant accumulation for salt-stress tolerance in plants.

Gene	Plant	Stress tolerance	Reference
Glycine betaine			
Arthrobacter globiformis choline oxidase (*CodA*)	*Arabidopsis*	Germination in 300 mM NaCl; seedling growth in 200 mM NaCl; retention of PSII activity at 400 mM NaCl.	93
A. globiformis CodA under CaMV 35S promoter	*Arabidopsis*	Exposure of 40-day-old plants to 100 mM NaCl stress for 3 days resulted in flower bud abortion and a decrease in number of seeds per silique in control plants. These adverse effects were less in transgenic plants.	94
A. globiformis CodA	Rice	Transgenic plants in which *CodA* is targeted to the chloroplasts were more tolerant to photoinhibition under 150 mM NaCl salt stress and cold stress than *CodA* expression in cytosol.	95
A. globiformis CodA	*Brassica juncea*	Better germination in 100-150 mM NaCl and seedling growth in 200 mM NaCl.	96
Arthrobacter pascens choline oxidase (*COX*)	Arabidopsis, *B.napus*, and tobacco	No significant differences in osmotic potential between transgenic and nontransgenic plants.	97
E. coli choline dehydrogenase (*betA*) and betaine aldehyde dehydrogenase (*betB*) genes	Tobacco	Biomass production of greenhouse-grown transgenic plants was greater than that of wild-type plants under salt stress; faster recovery from photoinhibition under high light, salt stress, and cold stress.	98
Atriplex hortensis BADH driven by maize ubiquitin promoter	*Triticum aestivum*	Seedling growth in 0.7 % NaCl.	99
Peroxisomal *BADH* of barley	Rice	Stability in chlorophyll fluorescence; accumulation of fewer Na^+ and Cl^- ions and more K^+ ions in shoots under 100 mM NaCl stress.	100
Proline			
Vigna aconitifolia L. P5CS Δ^1-pyrroline-5-carboxylate synthetase) gene	Tobacco	Better root growth and flower development under salt stress.	101

Gene	Plant	Stress tolerance	Reference
V. aconitifolia L. P5CS that lacks end product (proline) inhibition	Tobacco	Improved seedling tolerance and low free radical levels at 200 mM NaCl.	102
V. aconitifolia L. *P5CS* gene under barley HVA22 promoter	Rice	Faster recovery after a short period of salt stress to seedlings.	103
Antisense proline dehydrogenase gene	*Arabidopsis*	Tolerant to high salinity (600 mM NaCl); constitutive freezing tolerance (–7˚C) at vegetative stage.	104
Antisense Δ^1-pyrroline-5-carboxylate reductase gene under heat stress inducible promoter	Soybean	Antisense transgenic plants (vegetative stage) accumulate less proline and failed to survive 6 days of drought at 37˚C, whereas control plants survived.	105
Trehalose			
E. coli OstA (Trehalose 6P synthase) & *OstB* (Trehalose 6P Phosphatase) driven by ABA responsive promoter	Rice	Higher survival rate and K^+/Na^+ ratio, low Na^+ accumulation in the shoot, high PSII activity, high root and shoot growth under 100 mM NaCl stress at vegetative stage.	106
E. coli OstA & OstB driven by maize ubiquitin promoter	Rice	Enhanced seedling growth and PSII yield under salt, drought and cold stress.	107
Mannitol			
E. coli mt1D (Mannitol-1-phosphate dehydro-genase) driven by CaMV 35S promoter	Tobacco	Better fresh weight, plant height, and flowering under 250 mM NaCl for 30 days.	108
E. coli mt1D	*Arabidopsis*	Transgenic seeds were able to germinate in up to 400 mM NaCl, whereas control seeds ceased to germinate at 100 mM NaCl.	109

(*Cont.*)

Table 1. (*Cont.*)

Gene	Plant	Stress tolerance	Reference
E. coli mt1D	Tobacco	Better salt stress tolerance. Nonstressed transgenic plants were 20-25% smaller; Mannitol contributed only to 30-40% of the osmotic adjustment.	110
E. coli mt1D	*Triticum aestivum* L.	Only 8% biomass reduction as compared to 56% reduction in control plants under 150 mM NaCl stress. High level of mannitol accumulation causes stunted growth and sterility.	111
Celery mannose 6-P reductase driven by CaMV 35S promoter	*Arabidopsis*	Enhanced salt tolerance in terms of growth, flowering and seed production in soil irrigated with 300 mM NaCl.	112
D-Ononitol			
Ice plant Myo-inositol O-methyl transferase (*IMT1*)	Tobacco	Photosynthetic CO_2 fixation was slightly better under drought and salinity stress; faster recovery.	113
Sorbitol			
Apple Stpd1 (sorbitol-6-phosphate dehydrogenase) driven by CaMV 35S promoter	Japanese persimmon, *Diospyros kaki* Thunb.	Tolerance in Fv/Fm ratio under NaCl stress.	114

may also be required for enhanced tolerance. For example, transgenic rice plants that overexpress choline oxidase targeted to chloroplasts showed better tolerance to photoinhibition under salt and low-temperature stress than did plants overexpressing choline oxidase targeted to the cytosol (95). Often, engineered osmoprotectant overaccumulation results in impaired plant growth and development even under the nonstress environment. Transgenic tobacco plants overaccumulating mannitol (110), sorbitol (115), or trehalose (116) showed stunted growth. Often, engineered alterations in osmoprotectant accumulation results in infertility, depending on the concentration of osmoprotectant (111, 115). The use of a stress-inducible promoter to overexpress osmoprotectant biosynthesis helps in overcoming the growth defects while protecting the plants during osmotic stress (106). Although transgenic tobacco overexpressing the myo-inositol O-methyl transferase gene accumulated D-ononitol in the cytosol up to 600 mM during salt stress, D-ononitol did not enter the vacuole (113). Hence, further

understanding of the metabolic flux and compartmentation of osmoprotectants will help in precisely engineering osmoprotectant metabolism in plants for salt-stress tolerance.

REGULATION OF OSMOPROTECTANT METABOLISM

Evidence from genetic analysis, gene expression, and transgenic studies shows that osmoprotectant biosynthesis and accumulation in appropriate cellular organelles is critical for plant salt tolerance. However, the signaling cascades that regulate the osmoprotectant biosynthesis and catabolism during salt and other osmotic stress in higher plants are poorly understood. A signaling cascade similar to that of the yeast mitogen-activated protein kinase-high osmotic glycerol 1 (MAPK-HOG1) pathway may be involved in regulation of osmoprotectant biosynthesis (3, 19). *Arabidopsis AtHK1*, a putative osmosensory two-component hybrid histidine kinase, is implicated in osmosensing during salt stress. *AtHK1* expression is induced by salt stress, and it complements the yeast double mutant *sln1*Δ *sho1*Δ, which lacks osmosensors. Similar to the SLN1 osmosensor of yeast, AtHK1 is probably active at low osmolarity and may inactivate a response regulator by phosphorylation. High osmolarity caused by salt stress may inactivate AtHK1, which results in the accumulation of the active form of the nonphosphorylated response regulator and may activate osmolyte biosynthesis in plants by activating the MAPK pathways (19). Moreover, constitutive overexpression of a dominant-negative mutated form of *AtHK1* in transgenic *Arabidopsis* resulted in enhanced tolerance to salt and drought stress (117). Results from complementation analysis in yeast and transgenic *Arabidopsis* suggest that AtHK1 may act as an osmosensor in *Arabidopsis*. Determination of the *in vivo* role of higher plant putative sensory kinases and the identification of signaling intermediates and targets will shed more light on salt-stress signaling.

Genetic analysis of ABA-deficient mutants *los6/aba1* and *los5/aba3* of *Arabidopsis* revealed that proline biosynthesis during osmotic stress is regulated by ABA, because salt and other abiotic stress induction of *P5CS* gene expression is either diminished or blocked in these mutants (118, 119). In *Arabidopsis* and *Medicago truncatula*, of the two *P5CS* genes, the expression of only one gene is regulated by NaCl and osmotic stress (120, 121), which suggests that the promoters of these genes are differentially activated by developmental and osmotic stress cues. Biochemical analysis implicates phospholipase D (PLD) as a negative regulator of proline biosynthesis in *Arabidopsis* (122). Recent studies have shown that proline can act as signaling molecule to autoregulate the proline concentration and induce salt-stress-responsive proteins. In the desert plant *Pancratium maritimum* L., severe salt stress resulted in an inhibition of antioxidative enzymes such as catalase and peroxidase. Exogenous application of proline helped to maintain the activities of these enzymes and also upregulated several salt-stress-responsive dehydrin proteins (123). Microarray and RNA gel blot analyses have shown that 21 proline-inducible genes have the proline- or hypo-osmolarity-responsive element (PRE, ACTCAT) in their promoter (124, 125). Transient activation analysis of a PRE-containing promoter led to the identification of four bZIP transcription factors that may regulate proline dehydrogenase and other

proline- or hypo-osmolarity-responsive genes in *Arabidopsis* (126). Understanding the signaling events that regulate osmoprotectant metabolism during stress and recovery will be useful in improving salt and osmotic stress tolerance of crop plants.

LEA-TYPE PROTEINS

Late-embryogenesis-abundant (LEA) proteins are synthesized and stored in maturing seeds and are necessary for the desiccation tolerance of seeds. LEA-type proteins coding genes are called dehydrins, RD (responsive to dehydration), ERD (early responsive to dehydration), KIN (cold inducible), COR (cold regulated), and RAB (responsive to ABA) genes in different plant species (3, 127). LEA proteins are induced at higher levels by salt or ABA in salt-tolerant indica rice varieties than in salt-sensitive rice varieties (128). In higher plants, osmotic stress and ABA induce several LEA-type proteins in vegetative tissues. The expression levels of LEA proteins are correlated with desiccation tolerance in vegetative tissues, pollen, and seeds (129, 130). LEA proteins are rich in hydrophilic amino acids and are very stable. The proposed functions of LEA proteins under stress are to 1) protect the cellular structure by acting as a hydration buffer, 2) protect proteins and membranes, and 3) renature denatured proteins (129, 130). Genetically engineered rice plants constitutively overexpressing a barley *LEA* gene (*HVA1*) driven by rice actin-1 promoter showed better salt (200 mM NaCl) and drought stress tolerance and faster recovery once the stress was removed. Wilting, dying of old leaves, and necrosis of young leaves were delayed in transgenic rice as compared with control plants under both salt and water stress (131).

Transcriptional Regulation of *LEA*/*COR* Genes

ABA regulates several aspects of plant development, including seed development, desiccation tolerance of seeds, and seed dormancy, and plays a crucial role in abiotic and biotic stress tolerance of plants. Genetic analysis of ABA-deficient mutants established the essentiality of ABA signaling in stomatal control of transpiration (31). As discussed earlier, because the rate of transpiration determines the amount of salt transport into shoots, stomatal regulation by ABA is an important trait of plant salt tolerance. Salt and osmotic stress regulation of *LEA* gene expression is mediated by both ABA-dependent and -independent signaling pathways. Both the pathways appear to employ Ca^{2+} signaling, at least in part, to induce *LEA* gene expression during salinity and osmotic stress (3, 127). Northern analysis of *COR* gene expression in ABA-deficient mutants, namely *los5*/*aba3* and *los6*/*aba1* of *Arabidopsis*, showed that ABA plays a pivotal role in salt and osmotic stress–regulated gene expression. The expression of *RD29A*, *RD22*, *COR15A,* and *COR47* was severely reduced or completely blocked in the *los5* mutant (118), whereas in *los6*, the expression of *RD29A*, *RD19*, *COR15A*, *COR47,* and *KIN1* was lower than that in wild-type plants (119).

Promoters of *LEA*/*COR* genes contain dehydration-responsive elements/C-repeat (DRE/CRT), ABA-responsive elements (ABREs), MYC recognition sequence (MYCRS), and/or MYB recognition sequence (MYBRS) *cis*-elements. The regulation of gene expression through DRE/CRT *cis*-elements appears to be mainly ABA independent, whereas ABRE and MYB/MYC element-controlled gene expression is ABA dependent (127, 132). However, recent studies have shown that cross-talk exists between the ABA-dependent and -independent pathways. For example, *RD29A* expression depends on both *DRE* and *ABRE* elements (133), and ABA can also induce the expression of C-repeat binding proteins, CBF1-CBF3 (134). Salt-stress signaling through Ca^{2+} and ABA mediate the expression of *LEA* genes by transcription factors that activate *CRT*, *ABRE* and *MYC*/*MYB* *cis*-elements (Figure 4).

Calcium Sensor Proteins

An earlier section of this chapter described the role of ABA in regulating cytosolic Ca^{2+} signatures during salinity. Genetic and biochemical evidence show that ABA-mediated *COR* gene expression is regulated by Ca^{2+} signaling. In addition

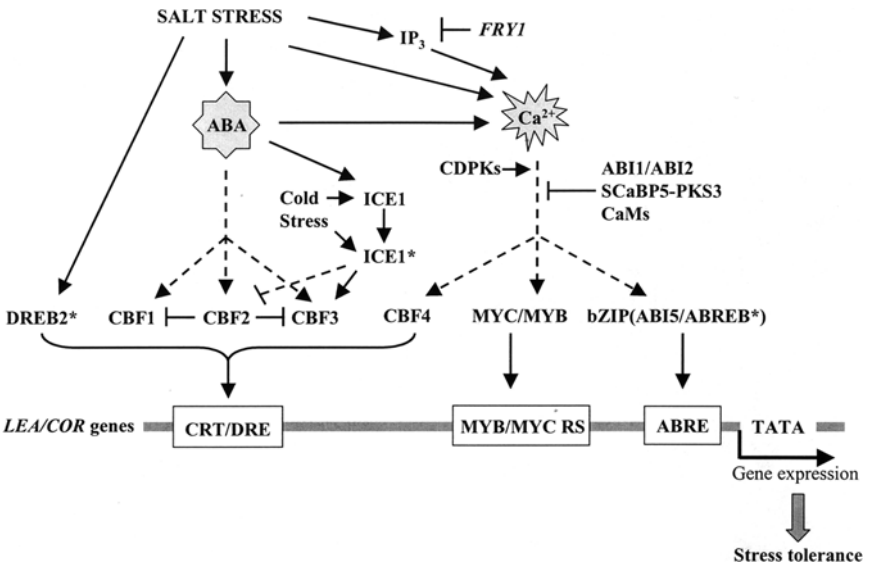


Figure 4. Transcriptional regulation of *LEA*/*COR* genes during salt stress. *LEA*/*COR* genes are activated MYC/MYB and bZIP type transcription factors mainly through ABA-dependent signaling. Salinity-induced ABA accumulation may mediate the expression of *CBF*s, which in turn induce expression of *LEA*/*COR* genes through DRE/CRT *cis*-elements during salinity. The ICE1, a myc-like bHLH transcription factor, regulates the transcription of CBFs during cold stress, while the upstream signaling events that regulate expression of CBFs under osmotic and ABA stresses are not known. Ca2+ signaling is positively regulated by CDPKs and negatively regulated by ABI1/2 protein phosphatase 2C, SCaBP5-PKS3 complex. and CaMs (* = indicates post-translation activation requirement).

to SOS3, salt stress–induced Ca^{2+} oscillations may also be perceived by Ca^{2+}-dependent protein kinases (CDPKs) and calmodulins (CaMs). *Arabidopsis AtCDPK1* and *AtCDPK2* are induced by salt and drought stress (135). In rice, salt, drought, and cold stress induce the expression of *OsCDPK7* (136). In *Mesembryanthemum crystallinum* (common ice plant), salinity and dehydration regulate myristoylation and localization of a CDPK (McCPK1) into the plasma membrane. Upon dehydration, McCPK1 changes its cellular localization from the plasma membrane to the nucleus, endoplasmic reticulum, and actin microfilaments (137). McCDPK1 phosphorylates the McCDPK1 substrate protein 1 (CSP1) *in vitro* in a Ca^{2+}-dependent manner, and salt stress induces co-localization of McCDPK1 and CSP1 in the nucleus of these ice plants (138). CDPKs transduce salt stress and ABA-induced Ca^{2+} signals to regulate the expression of LEA-type genes (136, 139). Transient expression analysis in maize protoplasts showed that an increase in cytosolic Ca^{2+} concentration activates CDPKs that induce the stress-responsive *HVA1* promoter, which is under the negative control of ABI1 protein phosphatase 2C (139). Overexpression analysis also confirmed the regulatory role of CDPKs in salinity-induced LEA gene expression. Transgenic rice overexpressing *OsCDPK7* showed enhanced induction of a *LEA*-type gene (*RAB16A*) and salt/drought tolerance, whereas antisense transgenic plants were hypersensitive to salt/drought stress (136).

CaMs may act as negative regulators of salt stress-induced Ca^{2+} signatures. Overexpression of *CaM3* in *Arabidopsis* repressed the expression of *COR* genes (*RD29A* and *COR6.6*) (140). The expression of *COR* genes is mediated by Ca^{2+} signals (141). Ca-ATPases mediate Ca^{2+} efflux from the cytoplasm and thus regulate the magnitude and duration of cytosolic Ca^{2+} oscillations. Endoplasmic reticulum Ca-ATPase (ACA2) has been shown to be activated by CaM and inhibited by CDPK (142). Salinity, dehydration, and cold stress–inducible *AtCaMBP25* (*Arabidopsis thaliana* calmodulin (*CaM*)-binding protein of 25 kDa) binds to a canonical CaM in a Ca^{2+}-dependent manner. Transgenic plants overexpressing *AtCaMBP25* showed hypersensitivity to salt and osmotic stress, whereas antisense *AtCaMBP25* transgenic plants were more tolerant to these stresses than the wild type. These results suggest that AtCaMBP25 may function as a negative effector of salt and osmotic stress signaling (143). The differences in affinity of SOS3, SCaBPs, CDPKs, and CaM for Ca^{2+} may determine the operation of specific signaling cascades and interactions. Thus, *LEA*/*COR* gene expression is regulated by the balance between the activities of CDPKs and CaMs. Ca-CaMs may also regulate cytoplasmic receptor-like kinases during salt and abiotic stress signaling. Salt-, cold-, and H_2O_2-inducible CaM binding cytoplasmic receptor-like kinase 1 (CRCK1) has been cloned from alfalfa (144). Transcriptome analyses showed the induction of receptor-like kinase genes in *Arabidopsis* under salt stress (14, 15). However, the roles of these proteins in salt-stress sensing and their targets are unknown.

Basic Leucine-Zipper-Family Transcription Factors

ABA-dependent expression of *COR* genes under osmotic stress is regulated by basic leucine-zipper (bZIP) (145) and MYB/MYC-type transcription fac-

tors (Figure 4) (146). Salt, drought, and ABA upregulate the expression of *Arabidopsis* bZIP transcription factors such as *ABREB1* (ABA-responsive element binding protein 1 = *ABF2*) and *ABREB2* (= *ABF4*) genes. These transcription factors have been shown to induce *RD29B* promoter-*GUS* in leaf protoplasts of wild-type *Arabidopsis* but not in *aba2* (ABA-deficient) and *abi1* (ABA-insensitive) mutants. The induction of *RD29B-GUS* by *ABREB*s is enhanced in an *era1* (enhanced response to *A*BA) mutant. This evidence suggests that ABA is necessary for the expression and activation of ABREB1 and ABREB2, which in turn regulate *COR* gene expression (145). Constitutive overexpression of *ABF3* and *ABREB2 (= ABF4)* in *Arabidopsis* enhanced the expression level of target *LEA* genes (*RAB18* and *RD29B*). These transgenic plants showed hypersensitivity to ABA, sugar, and salt stress during germination, but enhanced drought tolerance at the seedling stage (147).

MYB/MYC-Type Transcription Factors

MYB/MYC-type transcription factors such as AtMYC2 (=RD22BP1) and AtMYB2 regulate *LEA* gene expression in *Arabidopsis* during osmotic stress (Figure 4). Transgenic *Arabidopsis* plants overexpressing *AtMYC2* and *AtMYB2* showed constitutive expression of *RD22* and *AtADH*, and the expression levels were further increased with ABA treatment. The expression of *RD22* and *AtADH* genes is impaired in the *atmyc2* mutant. Transgenic *Arabidopsis* plants overexpressing *AtMYC2* and *AtMYB2* showed enhanced osmotic stress tolerance, as measured by electrolyte leakage from cells (146), although their salt stress tolerance is not known. Overexpression of ABA- and abiotic stress-inducible *Craterostigma plantagineum MYB10* enhanced salinity and desiccation tolerance of transgenic *Arabidopsis* plants. These transgenics also showed ABA hypersensitivity and altered sugar sensing. *In vitro* promoter binding assay showed that CpMYB10 binds to the *LEA Cp11-24* promoter (148).

C-repeat Binding Proteins

CBFs (C-repeat binding proteins) or dehydration-responsive-element binding proteins (*DREB*s) belong to the EREBP/AP2 domain transcription factor family. CBFs activate the expression of *LEA*/*COR* genes through *DRE*/*CRT* *cis*-elements in response to abiotic stress. *Arabidopsis DREB*s are classified into two classes: *DREB1* (*DREB1A=CBF3*, *DREB1B=CBF1*, *DREB1C=CBF2* and *CBF4*) and *DREB2* (*DREB2A* and *DREB2B*). The expression of *CBF1*, *CBF2* and *CBF3* is induced by cold stress, whereas that of *CBF4* is induced by drought stress. The expression of *DREB2A* and *DREB2B* is induced by dehydration and salt stress (132, 149, 150). Similar to *Arabidopsis DREB2*, rice *OsDREB2A* is induced by dehydration and salt stress (151). Osmotic stress–induced expression of *CBF4* appears to be mainly mediated by ABA (150). ABA has been shown to induce *CBF1*, *CBF2,* and *CBF3*, although their ABA-induced expression level is significantly lower than with cold stress (134). Transgenic plants overexpressing *CBF* (*CBF1, 3,* and 4) genes showed constitutive activation of *DRE*/*CRT* *cis*-element-dependent *COR* gene expression (149-153). Transcriptional activation of

COR genes by CBF transcription factors is conserved across plant species such as *Arabidopsis*, wheat, *B. napus* (154), barley, and rice (151). Transcriptome analysis of *CBF*-overexpressing transgenic *Arabidopsis* showed that approximately 13 *LEA*/*dehydrin* genes are under the transcriptional control of CBFs (155). Recently, ICE1 (inducer of CBF expression 1), a MYC-type basic helix-loop-helix transcription factor, as an upstream regulator of CBFs under cold stress was identified in *Arabidopsis* (156) (Figure 4). Upstream transcription factors that regulate the expression of *DREB2*/*CBFs* during salt stress have yet to be identified.

In tobacco *Tsi1* (tobacco-stress-induced-gene 1, a member of the EREBP/AP2 transcription factor family), gene expression is rapidly induced by salt but not drought or ABA. The overexpression of *TSI1* in tobacco enhanced the retention of chlorophyll content when leaves were floated in 400 mM NaCl solution for 48 or 72 h (157). Further detailed studies are needed to identify the targets of TSI1.

Transgenic *Arabidopsis* overexpressing *CBF1* or *CBF3* driven by the CaMV35S promoter or *CBF3* expression under the transcriptional control of the stress-responsive *RD29A* promoter showed enhanced tolerance to salt, drought, and freezing stress (149, 152-154). Transgenic wheat plants expressing *RD29A::CBF3* showed enhanced osmotic stress tolerance (158). Overexpression of the rice *OsDREB1A* gene in *Arabidopsis* resulted in the activation of target *LEA* genes and conferred salt and other abiotic stress tolerance (151). Constitutive overexpression of *CBF1* or *CBF3* resulted in growth abnormalities of the transgenic plants (149, 152-154, 159, 160). This problem has been overcome by the use of a stress-responsive promoter to drive the expression of *CBF*s (153, 158). Salt and abiotic stress tolerance of *CBF*-overexpressing transgenic plants was attributed to the enhanced expression of *LEA* genes (153, 154), accumulation of compatible osmolytes (161), and enhanced oxidative stress tolerance (159, 160). Genome-wide expression analysis showed that *CBF* overexpression also induces transcription factors such as AP2 domain proteins (RAP2.1 and RAP2.6), putative zinc finger protein, and R2R3-MYB73 (155), which might regulate genes involved in osmolyte biosynthesis and antioxidant defense. This evidence shows that the expression of several genes can be manipulated in transgenic plants engineered with a single CBF transcription factor and that enhanced expression of *LEA* genes is critical for salt and other abiotic stress tolerance.

OXIDATIVE STRESS MANAGEMENT

Reactive oxygen species (ROS), namely, superoxide radicals ($O_2^{\cdot -}$), hydrogen peroxide (H_2O_2), and hydroxyl radicals ($OH^{\cdot}$) are produced in aerobic cellular processes such as mitochondrial and chloroplast electron transport and oxidation of glycolate (photorespiration), xanthine, and glucose. ROS cause oxidative damage to membrane lipids, proteins, and nucleic acids. Hence, organisms have evolved various antioxidants and detoxifying enzymes to scavenge ROS efficiently. Antioxidant enzymes employed by plants are ascorbate, glutathione, α-tocopherol, and carotenoids, whereas detoxifying enzymes include superoxide dismutase (SOD), catalase, peroxidase, and enzymes of the ascorbate-glutathione

cycle. SOD converts superoxide to H_2O_2, which is detoxified to water and oxygen by the catalase and/or ascorbate-glutathione cycle. Salt stress induces the accumulation of ROS and enhances the expression of ROS-detoxifying enzymes (5, 162-164). Alleviation of oxidative damage by scavenging ROS is an important strategy of plants to tolerate stress (3, 165-167). Hence, several efforts have been made to improve salt tolerance by engineering ROS-detoxifying enzymes.

Transgenic plants overexpressing ROS-scavenging enzymes such as SOD (168), ascorbate peroxidase (APX) (169), and glutathione S-transferase/glutathione peroxidase (GST/GPX) (170, 171) showed increased tolerance to osmotic, temperature, and oxidative stress. The overexpression of the tobacco *NtGST/GPX* gene in transgenic tobacco plants improved salt- and chilling-stress tolerance because of enhanced ROS scavenging and prevention of membrane damage (170, 171). Transgenic tobacco plants overexpressing *AtAPX* targeted to the chloroplasts showed enhanced tolerance to salinity and oxidative stress (172). The *Arabidopsis pst1* (*photoautotrophic salt tolerance 1*) mutant is more tolerant to salt stress than is the wild type. The salt tolerance of this mutant was attributed to higher activities of SOD and APX than in the wild-type *Arabidopsis* (173). These evidences show that ROS detoxification is an important trait of plant salt tolerance.

Salt stress (5, 162) and ABA (174, 175) induce enhanced production of H_2O_2. ABA-dependent ROS production is catalyzed by NADPH oxidase, as revealed by analysis of the *atrbohD/F* double mutant of *Arabidopsis*, which is impaired in ABA-induced ROS production (176). ABA-elicited H_2O_2 production is negatively regulated by the ABI2 protein (177). H_2O_2 acts as a systemic molecule in regulating the expression of *GST* and *GPX* genes (178). The accumulation of H_2O_2 in leaves of catalase-deficient tobacco plants was sufficient to induce the production of defense proteins (GPX, PR-1) locally as well as systemically (179). Promoter analysis of the salt stress-inducible *Citrus sinensis GPX1* (phospholipid hydroperoxide) gene suggests that *GPX1* upregulation under salinity is mediated by H_2O_2 but not superoxide (172). Promoters of genes that encode ROS-detoxifying enzymes contain antioxidant-responsive elements (ARE), ABA-responsive elements (ABRE), nuclear factor (NF)-κB redox-regulated transcription factor recognition sequences, heat shock elements (HSE), and redox-regulated transcription factor Y-box *cis*-elements (180). Hence, ABA, as well as H_2O_2, may act as a second messenger to regulate antioxidant defense genes during salinity. Oxidative stress signaling is probably mediated by the MAPK cascade in plants (180, 181).

Pyramiding of chloroplastic and mitochondrial Mn-SOD in alfalfa resulted in lower biomass production compared with that in transgenic plants expressing either of the Mn-SODs (182). Engineered alterations in antioxidant systems may alter the pool size of ROS, which are involved in developmental, biotic, and abiotic stress signaling (175, 183). In field environments, crop plants often experience more than one biotic and abiotic stress. Critical evaluation of the engineered alterations in the antioxidant system on crop productivity in the normal environment, as well as under multiple stress environments in field conditions, and understanding of the signaling components that regulate ROS detoxification during salinity are needed to use this trait for genetically engineering plant salt tolerance.

MAPK SIGNALING PATHWAY

ROS signaling in plants under various stresses is mediated by mitogen-activated protein kinase (MAPK) signaling pathways (184, 185). Salt stress triggers the activation and enhances gene expression of MAPK signaling cascades, some components of which are common for both salt and ROS (181, 186). The *Arabidopsis* genome encodes approximately 60 MAPKKKs but only approximately 10 MAPKKs and 20 MAPKs (187). Hence, signals perceived by the 60 MAPKKKs must be transduced through 10 MAPKKs to 20 MAPKs. Thus, MAPK cascades offer potential nodes for stress, hormonal, and developmental signal cross-talk. Salt stress activates *Arabidopsis* AtMEKK1 (=MAPKKK) (188), AtMKK2 (=MAPKK) (189) and MAPKs (ATMPK3, ATMPK4, and ATMPK6) (190, 191). The active form of AtMEKK1 has been shown to activate AtMPK4 *in vitro* (192). Yeast 2-hybrid analysis, *in vitro* and *in vivo* protein kinase assays, and analysis of *mkk2* null mutants have led to the identification of a MAPK signaling pathway consisting of AtMEKK1, AtMEK1/AtMKK2, and AtMPK4/AtMPK6 (188, 189, 191) that transduces salt and other abiotic stress signals in *Arabidopsis*. Transgenic *Arabidopsis* plants overexpressing *AtMKK2* showed constitutive AtMPK4 and AtMPK6 activity and enhanced salt (germination on 150 mM NaCl medium) and freezing tolerance, whereas *mkk2* mutant plants exhibited impaired activation of AtMPK4 and AtMPK6 and thus hypersensitivity to salt and cold stress (189). In addition to salinity, H_2O_2 activates AtMPK3 and AtMPK6 (193), probably through H_2O_2-activated ANP1 (=MAPKKK) (194). Transgenic tobacco plants overexpressing a constitutively active tobacco *ANP1* ortholog, *NPK1*, exhibited constitutive AtMPK3 and AtMPK6 activity and enhanced salt-(300 mM NaCl for 3 days), drought-, and cold-stress tolerance (194).

Gene expression analysis of *AtMKK2*- and *ANP1*-overexpressing transgenic *Arabidopsis* plants led to the identification of target genes of this MAPK pathway. Overexpression of the active form of ANP1 showed activation of the *GST6* and *HSP18.2* promoters but not the *RD29A* promoter. A single amino acid mutation in the ATP-binding site of ANP1 abolished the ANP1 effect on these promoters (194). Microarray analysis of the transcriptome profile of *MKK2*-overexpressing plants identified approximately 152 target genes. Up-regulated genes include CBF2, RAV1, RAV2, MYB, and WRKY transcription factors, which may further regulate the expression of sub-regulons (189).

The *Arabidopsis* MAPK phosphatase 1 (*mkp1*) mutant exhibits salinity tolerance but hypersensitivity to genotoxic stress induced by UV-C. In a yeast 2-hybrid screen, MKP1 interacted with AtMPK3, 4, and 6. Microarray analysis of *mkp1* revealed that AtMKP1 negatively regulates a putative Na^+/H^+ antiporter AT4G23700 (195). Hence, MKP1 may negatively regulate salt stress signaling through AtMPK4. *Arabidopsis* nucleoside diphosphate kinase 2 (AtNDPK2) has been shown to interact with and activate AtMPK3 and AtMPK6 in yeast 2-hybrid and transgenic *Arabidopsis* plants overexpressing *AtNDPK2*. Further, these transgenic plants accumulated lower levels of ROS and showed enhanced tolerance to salinity and other abiotic stress. A deletion mutation of *AtNDPK2* impaired AtMPK3 and AtMPK6 activities. This evi-

dence suggests that AtNDPK2 is a positive regulator of stress signaling through MAPK pathways (193). In rice, the gene expression as well as kinase activity of *OsMAPK5* is regulated by ABA and biotic and abiotic stresses such as salt, drought, wounding, and cold. Transgenic rice overexpressing *OsMAPK5* showed increased tolerance to several abiotic stresses, including salt stress (196). These evidences show that diverse abiotic stress signals converge at MAPK cascades to regulate stress tolerance. Thus, in *Arabidopsis*, MAPK cascades consisting of AtMEKK1/ANP1, AtMEK1/AtMKK2, and AtMPK3/AtMPK4/ AtMPK6 may transduce salt-stress signaling. These MAPK cascades are further fine-tuned by a negative regulator, AtMKP1, and a positive regulator, AtNDPK1 (Figure 5).

MOLECULAR BREEDING

Selection for yield under field stress conditions across environments is time and labor consuming. Hence, the identification of component physiological traits of salt tolerance, which are linked to stress tolerance in yield, will enhance the pace of breeding programs. These physiological traits often are controlled by multiple genes and show continuous variation in segregating populations. These types of traits are called quantitative traits, and the regions of chromosomes

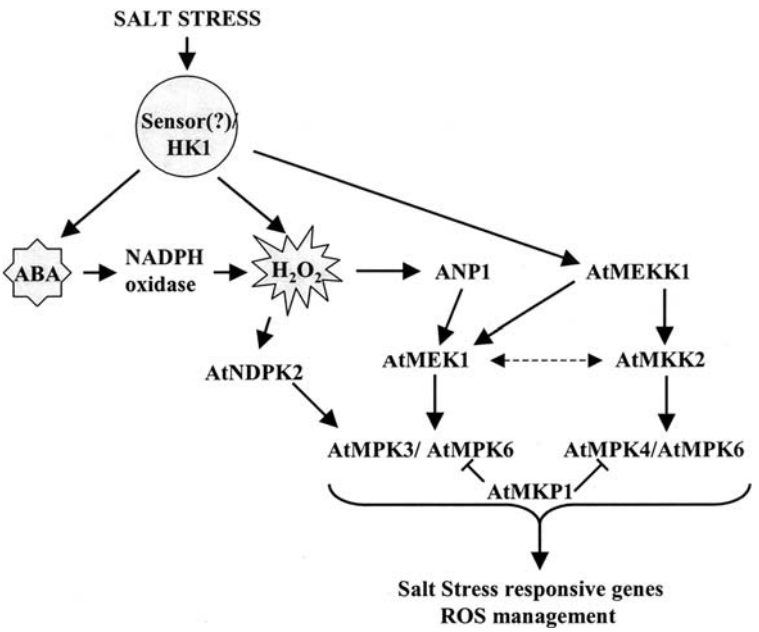


Figure 5. MAPK signaling pathways during salt stress in *Arabidopsis*. An unknown sensor perceives and transduces the salt-stress signals through MAPK pathways. Salt-stress sensors activate MAPK cascades either in an ABA and reactive oxygen species (ROS) dependent or independent pathway. Activated MAPK (ANP1 and AtMEKK1 = MAPKKK; AtMEK1=MAPKK; AtMPK3, 4 and 6 = MAPK) cascades regulate salt-stress responsive genes and salt tolerance. AtNDPK2 is a positive regulator of AtMPK3 and 6, whereas AtMKP1 is a negative regulator of AtMPK4 and 6.

controlling these traits are called quantitative trait loci (QTLs). Identifying QTLs with use of molecular markers is the primary step for marker-assisted breeding and candidate gene cloning. The application of molecular markers to identify QTLs for physiological traits has helped to identify QTLs linked to salt-stress tolerance in different plant species (Table 2).

Table 2. QTLs for salt-stress tolerance in different plant species. Some examples showing the number of QTLs, contribution of individual QTLs, and combined effect of QTLs on phenotypic variation.

Plant species	Mapping population	Component trait of salt tolerance	No. of QTLs and their contribution	Reference
Rice	RIL	Na^+, K^+ uptake and concentration	16	197
Rice	RIL	Dry mass; Na^+/K^+ ratio	11; Individual QTLs contributed to 6-19% variation	198
Rice	RIL	Na^+, K^+ absorption	2	199
Rice	F2-F3	Na^+, K^+ uptake and concentration	2 major (one each for 48.5% and 40.1% variation in Na^+ and K^+ concentration, respectively) + minors	200
*Lycopersico n*spp.	Two different F2 populations	Fruit weight Fruit No.	4 (Cross1), 6 (Cross2) 10 (Cross1), 6 (Cross2); contribution of individual QTLs vary from 6-25%	201
*Lycopersico n*spp.	Inbred backcross (BC1 selfed)	Salt tolerance during germination	7, All QTLs accounted for only 45% variation; individual QTLs contributed to 6.5-15.6% variation	202
*Lycopersico n*spp.	Inbred backcross (BC1 selfed)	Salt tolerance during vegetative stage	5QTLs, Individual QTLs contributed to 5.7-17.7% variation, with the combined effects being about 46% of the phenotypic variation	203
Arabidopsis	RIL	Salt tolerance during germination and seedling growth	11 (6 for germination explaining 32% variation +5 for vegetative growth explaining 38% variation); individual QTLs contributed to 5-14% variation	204

In tomato, a major QTL (*fwTG48-TG180*) that accounted for 58% variation in fruit weight under control conditions contributed to only 14% variation with salt stress. However, the same QTL contributed to 17% and 8% of the genotypic variation under control conditions and salt stress, respectively, in another F2 population. The detection of approximately 50% or more of QTLs for salt tolerance depends on the salinity stress (201). Thus, QTLs are stress sensitive, and proper regulation of gene expression is critical for salinity tolerance. Further differential sensitivity of different phenological phases of plant development to salinity stress is evident from the results of QTL analysis. QTLs associated with tolerance at germination differ from those of vegetative growth (202-204). QTL analyses clearly establish that 1) salt tolerance is governed by multiple genes; 2) the contribution of individual significant QTLs can vary from 5% to 50%, depending upon the complexity of the trait; 3) the stress responsiveness of QTLs indicates the crucial role of gene regulation during stress; and 4) QTLs for tolerance at different phenological phases specify the changes in salt tolerance mechanism during plant development.

If a QTL can be considered as a cluster of related genes that may be under the transcriptional control of one or more regulatory genes, one or more QTLs may also be under the transcriptional control of a single regulatory gene. The identification of gene(s) contributing to major QTLs and genetic transfer (breeding/genetic engineering) of a single regulatory gene that controls the expression of several target genes will significantly enhance the pace of development of salt-tolerant crops.

CONCLUSIONS AND PROSPECTS

During the past decade, the applications of molecular tools such as gene disruption and transgenic approaches have significantly enhanced our knowledge of salt-stress tolerance. Significant progress has been made toward understanding salt-stress signaling that controls ion homeostasis and salt tolerance. The SOS pathway regulates ion homeostasis during salt stress in *Arabidopsis*. Salt-stress sensor-induced cytosolic Ca^{2+} signals are perceived by SOS3, which in turn activates the SOS2 kinase. The activated SOS2 kinase regulates sodium efflux and sequesters sodium into the vacuole by activating Na^+/H^+ antiporters of plasma membrane and tonoplast, respectively. Osmotic homeostasis and stress damage control appear to be regulated by salt stress–induced ABA, ROS, a putative osmosensory histidine kinase (AtHK1), and MAPK cascades. However, components and targets of these signaling pathways are not yet understood. CBFs, bZIP, MYB, and MYC types of transcription factors induce *LEA* gene expression during osmotic stress. Molecular, genetic, and cell biological approaches to identify signaling components and biochemical characterization of signaling complexes will be required to further understand salt-stress signaling pathways and their use in crop improvement.

The transgenic approach demonstrates the possibilities of gene transfer across organisms and engineering salt tolerance by manipulating a single gene or a few genes. Genetic engineering of ion transporters has been shown to significantly enhance salt tolerance (65, 77, 78). Transgenic manipulation of signaling

molecules and transcription factors will be advantageous, because engineering a single gene can change the expression of several target genes involved in stress response and provide multiple abiotic stress tolerance (68, 149, 152, 153, 189, 196). Often, constitutive overexpression of signaling components, osmoprotectants and stress-responsive genes, results in reduced plant size and other growth abnormalities, even under normal growth conditions. Kasuga et al. (153) demonstrated that the use of a stress-responsive promoter could overcome this problem. Hence, the selection of stress-responsive and tissue-specific promoters for engineering the stress-tolerance trait is critical. The overexpression of osmoprotectant and antioxidant systems has been shown to protect transgenic plants from salt stress. Some of the osmoprotectants, such as polyols and trehalose, overproduced in transgenics, are often associated with growth defects and sterility. Engineering for antioxidant systems may alter the pool size of H_2O_2, a signaling molecule involved in developmental and stress signaling. Hence, careful examination is needed in employing these traits to engineer salt-tolerant crops.

Most of the transgenics discussed here are model plants, and stress tolerance was assessed at the vegetative phase of growth in controlled conditions for very short durations. Often, transgenic plants are not evaluated under realistic stress conditions (2). In most cases, very high salt-stress levels are applied to show clearly the survival of transgenic plants and death of control plants, rather than their productivity under long-term realistic salinity levels. Hence, the effect of stress in relation to plant ontogeny should be assessed at realistic stress levels and under combinations that occur in nature, by using transgenic crop plants in the field. The identification of QTLs for salt tolerance in different crops will be needed for precise molecular breeding for salt-tolerant crops. The application of marker-assisted selection to QTLs of major effects should help in improving salt tolerance of crop plants. In the near future, pyramiding regulatory genes controlling the various aspects of salt tolerance (i.e., ionic and osmotic homeostasis, and damage control) in a single transgenic plant is expected to yield salt-tolerant crop plants with a very high level of tolerance to salt and osmotic stress.

ACKNOWLEDGMENTS

Research in our laboratory is supported by grants from the U.S. National Institutes of Health, the U.S. National Science Foundation, and the U.S. Department of Agriculture.

REFERENCES

1 Munns, R. (2002) Plant Cell Environ. 25, 239-250.
2 Flowers, T.J. (2004) J. Exp. Bot. 55, 307-319.
3 Zhu, J.-K. (2002) Annu. Rev. Plant Biol. 53, 247-273.
4 Tester, M. and Davenport, R.A. (2003) Ann. Bot. 91, 503-527.
5 Hernandez, J.A., Ferrer, M.A., Jimenez, A., Barcelo, A.R. and Sevilla, F. (2001) Plant Physiol. 127, 817-831.

6 Maas, E.V. (1990) in Agricultural Salinity Assessment and Management, (K.K. Tanji, ed.), Chapter 13, pp. 262-304, ASCE Manuals and Reports on Engineering No. 71, American Society of Civil Engineers, New York.
7 Netondo, G.W., Onyango, J.C. and Beck, E. (2004) Crop Sci. 44, 806-811.
8 Zeng, L. and Shannon, M.C. (2000) Crop Sci. 40, 996-1003.
9 Munns, R. and Rawson, H.M. (1999) Aust. J. Plant Physiol. 26, 459-464.
10 West, G., Inze, D. and Beemster, G.T. (2004) Plant Physiol. 135, 1050-1058.
11 Burssens, S., Himanen, K., van de Cotte, B., Beeckman, T., Van Montagu, M., Inze, D. and Verbruggen, N. (2000). Planta 211, 632-40.
12 Wang, H., Qi, Q., Schorr, P., Cutler, A.J., Crosby, W. and Fowke, L.C. (1998) Plant J. 15, 501-510.
13 Sun, K., Hunt, K. and Hauser, B.A. (2004) Plant Physiol. 135, 2358-2367.
14 Kreps, J.A., Wu, Y., Chang, H.S., Zhu, T., Wang, X. and Harper J.F. (2002) Plant Physiol. 130, 2129-2141.
15 Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. (2002). Plant J. 31, 279-292.
16 Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) Plant Physiol. 133, 1755-1767.
17 Zhu, J.K. (2003) Curr. Opin. Plant Biol. 6, 441-445.
18 Liu, W., Fairbairn, D.J., Reid, R.J. and Schachtman, D.P. (2001) Plant Physiol. 127, 283-294.
19 Urao, T., Yakubov, B., Satoh, R., Yamaguchi-Shinozaki, K., Seki, M., Hirayama, T. and Shinozaki, K. (1999) Plant Cell 11, 1743-1754.
20 Kurth, E., Cramer, G.R., Lauchli, A. and Epstein, E. (1986) Plant Physiol. 99, 1461-1468.
21 Schroeder, J.I. and Hagiwara, S. (1989) Nature 338, 427-430.
22 Knight, H., Trewavas, A.J. and Knight, M.R. (1997) Plant J. 12, 1067-1078.
23 Sanders, D., Brownlee, C. and Harper, J.F. (1999) Plant Cell 11, 691-706.
24 DeWald, D.B., Torabinejad, J., Jones, C.A., Shope, J.C., Cangelosi, A.R., Thompson, J.E., Prestwich, G.D. and Hama, H. (2001) Plant Physiol. 126, 759-769.
25 Takahashi, S., Katagiri, T., Hirayama, T., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2001) Plant Cell Physiol. 42, 214-222.
26 Xiong, L., Lee, B.-H., Ishitani, M., Lee, H., Zhang, C. and Zhu, J.-K. (2001) Genes Dev. 15, 1971-1984.
27 Jia, W., Wang, Y., Zhang, S. and Zhang, J. (2002) J. Exp. Bot. 53:2201-2206.

28 Xiong, L., Schumaker, K.S. and Zhu, J.–K. (2002) Plant Cell 14, S165-S183.
29 Xiong, L. and Zhu, J.–K. (2003) Plant Physiol. 133, 29-36.
30 Leung, J. and Giraudat, J. (1998) Annu. Rev. Plant Physiol. Plant Mol. Biol. 49, 199-222.
31 Schroeder, J.I., Allen, G.J., Hugouvieux, V., Kwak, J.M. and Waner, D. (2001) Annu. Rev. Plant Physiol. Plant Mol. Biol. 52, 627-658.
32 Wu, Y., Kuzma, J., Marechal, E., Graeff, R., Lee, H.C., Foster, R. and Chua, N-H. (1997) Science 278, 2126-2130.
33 Sanchez, J.P., Duque, P. and Chua, N.-H. (2004) Plant J. 38, 381-395.
34 Wang, X-Q., Ullah, H., Jones, A. and Assmann, S. (2001) Science 292, 2070-2072.
35 Munns, R., Hare, R.A., James, R.A. and Rebetzke, G.J. (1999) Aust. J. Agric. Res. 51, 69-74.
36 Inan, G., Zhang, Q., Li, P., Wang, Z., Cao, Z., Zhang, H., Zhang, C., Quist, T.M., Goodwin, S.M., Zhu, J., Shi, H., Damsz, B., Charbaji, T., Gong, Q., Ma, S., Fredricksen, M., Galbraith, D.W., Jenks, M.A., Rhodes, D., Hasegawa, P.M., Bohnert, H.J., Joly, R.J., Bressan, R.A. and Zhu, J.-K. (2004) Plant Physiol. 135, 1718-1737.
37 Karahara, I., Ikeda, A., Kondo, T. and Uetake, Y. (2004) Planta 219, 41-47.
38 Garcia, A., Rizzo, C.A., Ud-Din, J., Bartos, S.L., Senadhira, D., Flowers, T.J. and Yeo, A.R. (1997) Plant Cell Environ. 20, 1167-1174.
39 Yeo, A.R., Flowers, S.A., Rao, G., Welfare, K., Senanayake, N. and Flowers, T.J. (1999) Plant Cell Environ. 22, 559-565.
40 Santa–Maria, G.E., Rubio, F., Dubcovsky, J. and Rodgriguez-Navarro, A. (1997) Plant Cell 9, 2281-2289.
41 Fu, H–H. and Luan, S. (1998) Plant Cell 10, 63-73.
42 Su, H., Golldack, D., Zhao, C. and Bohnert, H.J. (2002) Plant Physiol. 129, 1482-1493.
43 Rubio, F., Gassmann, W. and Schroeder, J.I. (1995) Science 270, 1660-1663.
44 Gorham, J., Bridges, J., Dubcovsky, J., Dvorak, J., Hollington, P. A., Luo, M.C. and Khan, J.A. (1997) New Phytol. 137, 109-116.
45 Uozumi, N., Kim, E.J., Rubio, F., Yamaguchi, T., Muto, S., Tsuboi, A., Bakker, E.P., Nakamura, T. and Schroeder, J.I. (2000) Plant Physiol. 122, 1249-1259.
46 Golldack, D., Su, H., Quigley, F., Kamasani, U.R., Munoz-Garay, C., Balderas, E., Popova, O.V., Bennett, J., Bohnert, H.J. and Pantoja, O. (2002) Plant J. 31, 529-542.
47 Laurie, S., Feeney, K.A., Maathuis, F.J.M., Heard, P.J., Brown, S.J. and Leigh, R.A. (2002) Plant J. 32, 139-149.
48 Bordás, M., Montesinos, C., Dabauza, M., Salvador, A., Roig, L.A., Serrano, R. and Moreno, V. (1997) Transgenic Res. 5, 1-10.
49 Gisbert, C., Rus, A.M., Bolarín, M.C., López-Coronado, J.M., Arrillaga, I., Montesinos, C., Caro, M., Serrano, R. and Moreno, V. (2000) Plant Physiol. 123, 393-402.

50 Rus, A.M., Estañ, M.T., Gisbert, C., Garcia-Sogo, B., Serrano, R., Caro, M., Moreno, V. and Bolarín, M.C. (2001) Plant Cell Environ. 24, 875-880.
51 Ellul, P., Ríos, G., Atarés, A., Roig, L.A., Serrano, R. and Moreno, V. (2003) Thoer. Appl. Genet. 107, 462-469.
52 Espinosa-Ruiz, A., Belles, J.M., Serrano, R. and Culianez–Macla, F.A. (1999) Plant J. 20, 529-539.
53 Maathuis, F.J.M. and Sanders, D. (2001) Plant Physiol. 127, 1617-1625.
54 Donaldson, L., Ludidi, N., Knight, M.R., Gehring, C. and Denby, K. (2004) FEBS Lett. 569, 317-320.
55 Shi, H., Xiong, L., Stevenson, B., Lu, T. and Zhu, J.–K. (2002) Plant Cell 14, 575-588.
56 Liu, J. and Zhu J.-K. (1998) Science 280:1943-1945.
57 Ishitani, M., Liu, J., Halfter, U., Kim, C-S., Shi, W. and Zhu, J.-K. (2000) Plant Cell 12, 1667-1677.
58 Liu, J., Ishitani, M., Halfter, U., Kim, C.S. and Zhu, J.-K. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 3730-3734.
59 Halfter, U., Ishitani, M. and Zhu, J.-K. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 3735-3740.
60 Guo, Y., Halfter, U., Ishitani, M. and Zhu, J.-K. (2001). Plant Cell 13, 1383-1400.
61 Shi, H., Ishitani, M., Kim, C.-S. and Zhu, J.-K. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 6896-6901.
62 Shi, H., Quintero, F.J., Pardo, J.M. and Zhu, J.-K. (2002b). Plant Cell 14, 465-477.
63 Qiu, Q.S., Barkla, B.J., Vera-Estrella, R., Zhu, J.-K. and Schumaker, K.S. (2003) Plant Physiol. 132, 1041-1052.
64 Qiu, Q.S., Guo, Y., Dietrich, M.A., Schumaker, K.S. and Zhu, J.-K. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 8436-8441.
65 Shi, H., Lee, B-H., Wu, S-J. and Zhu, J.-K. (2003) Nature Biotech. 21, 81-85.
66 Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J.-K. and Shinozaki, K. (2004) Plant Physiol. 135, 1697-1709.
67 Quintero, F.J., Ohta, M., Shi, H., Zhu, J.-K. and Pardo, J.M. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 9061-9066.
68 Guo, Y., Qiu, Q.S., Quintero, F.J., Pardo, J.M., Ohta, M., Zhang, C., Schumaker, K.S. and Zhu, J.-K. (2004) Plant Cell 16, 435-449.
69 Fukuda, A., Chiba, K., Maeda, M., Nakamura, A., Maeshima, M. and Tanaka, Y. (2004) J. Exp. Bot. 55, 585-594.
70 Gaxiola, R.A., Li, J., Undurraga, S., Dang, L.M., Allen, G.J., Alper, S.L. and Fink, G.R. (2001) Proc. Nat. Acad. Sci. U.S.A. 98, 11444-11449.
71 Zhao, L., Zhang, F., Guo, J., Yang, Y., Li, B. and Zhang, L. (2004) Plant Physiol. 134, 849-857.
72 Gaxiola, R.A., Rao, R., Sherman, A., Grisafi, P., Alper, S.L. and Fink, G.R. (1999) Proc. Nat. Acad. Sci. U.S.A. 96, 1480-1485.

73 Shi, H. and Zhu, J.-K. (2002). Plant Mol. Biol. 50, 543-550.
74 Fukuda, A., Nakamura, A. and Tanaka, Y. (1999) Biochim. Biophys. Acta 1446, 149-155.
75 Wu, C.A., Yang, G.D., Meng, Q.W. and Zheng, C.C. (2004) Plant Cell Physiol. 45, 600-607.
76 Apse, M.P., Aharon, G.S., Snedden, W.S. and Blumwald, E. (1999) Science 285, 1256-1258.
77 Zhang, H.-X. and Blumwald, E. (2001) Nature Biotech. 19, 765-768.
78 Zhang, H.-X., Hodson, J.N., Williams, J.P. and Blumwald, E. (2001) Proc. Nat. Acad. Sci. U.S.A. 98, 12832-12836
79 Ohta, M., Hayashi, Y., Nakashima, A., Hamada, A., Tanaka, A., Nakamura, T. and Hayakawa, T. (2002) FEBS Lett. 532, 279-282.
80 Qiu, Q.S., Guo, Y., Quintero, F.J., Pardo, J.M., Schumaker, K.S. and Zhu, J.-K. (2004) J. Biol. Chem. 279, 207-215.
81 Guo, Y., Xiong, L., Song, C.P., Gong, D., Halfter, U. and Zhu, J.-K. (2002) Dev. Cell 3, 233-244.
82 Cheng, N.H., Pittman, J.K., Zhu, J.-K. and Hirschi, K.D. (2004) J. Biol. Chem. 279, 2922-2926.
83 Berthomieu, P., Conéjéro, G., Nublat, A., Brackenbury, W.J., Lambert, C., Savio, C., Uozumi, N., Oiki, S., Yamada, K., Cellier, F., Gosti, F., Simonneau, T., Essah, P.A., Tester, M., Véry, A.-A., Sentenac, H. and Casse, F. (2003) EMBO J. 22, 2004-2014.
84 Maser, P., Eckelman, B., Vaidyanathan, R., Horie, T., Fairbairn, D.J., Kubo, M., Yamagami, M., Yamaguchi, K., Nishimura, M., Uozumi, N., Robertson, W., Sussman, M.R. and Schroeder, J.I. (2002) FEBS Lett. 531, 157-161.
85 Rus, A., Yokoi, S., Sharkhuu, A., Reddy, M., Lee, B-H., Matsumoto, T.K., Koiwa, H., Zhu, J.-K., Bressan, R.A. and Hasegawa, P.M. (2001) Proc. Nat. Acad. Sci. U.S.A. 98, 14150-14155.
86 Ohta, M., Guo, Y., Halfter, U. and Zhu, J.-K. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 11771-11776.
87 Bohnert, H.J. and Jensen, R.G. (1996) Trends Biotech. 14, 89-97.
88 Smirnoff, N. and Cumbes, Q.J. (1989) Phytochem. Anal. 28,1057-1060.
89 Chen, T.H.H. and Murata, N. (2002) Curr. Opin. Plant Biol. 5, 250-257.
90 Kasinathan, V. and Wingler, A. (2004) Physiol. Plant. 121, 101-107.
91 Urano, K., Yoshiba, Y., Nanjo, T., Ito, T., Yamaguchi-Shinozaki K. and Shinozaki, K. (2004) Biochem. Biophys. Res. Commun. 313, 369-375.
92 Mou, Z., Wang, X., Fu, Z., Dai, Y., Han, C., Ouyang, J., Bao, F., Hu, Y. and Li, J. (2002) Plant Cell 14, 2031-2043.
93 Hayashi, H., Alia, Mustardy, L., Deshnium, P., Ida, M. and Murata, N. (1997) Plant J. 12, 133-142.
94 Sulpice, R., Tsukaya, H., Nonaka, H., Mustardy, L., Chen, T.H. and Murata, N. (2003) Plant J. 36, 165-176.
95 Sakamoto, A, Alia, and Murata, N. (1998) Plant Mol. Biol. 38, 1011-1019.

96 Prasad, K.V.S.K., Sharmila, P., Kumar, P.A. and Saradhi, P.P. (2000) Mol. Breed. 6, 489-499.
97 Huang, J., Hirji, R., Adam, L., Rozwadowski, K.L., Hammerlindl, J.K., Keller, W.A. and Selvaraj, G. (2000) Plant Physiol. 122, 747-756.
98 Holmstrom, K.O., Somersalo, S., Mandal, A., Palva, T.E. and Welin, B. (2000) J. Exp. Bot. 51, 177-185.
99 Guo, B.H., Zhang, Y.M., Li, H.J., Du, L.Q., Li, Y.X., Zhang, J.S., Chen, S.Y. and Zhu, Z.Q. (2000) Acta Bot. Sinica 42, 279-283.
100 Kishitani, S., Takanami, T., Suzuki, M., Oikawa, M., Yokoi, S., Ishitani, M., Alvarez-Nakase, A.M., Takabe, T. and Takabe, T. (2000) Plant Cell Environ. 23, 107-114.
101 Kishor, P.B.K., Hong, Z., Miao, G.H., Hu, C.A.A. and Verma, D.P.S. (1995) Plant Physiol. 108, 1387-1394.
102 Hong, Z., Lakkineni, K., Zhang, Z. and Verma, D.P.S. (2000) Plant Physiol. 122, 1129-1136.
103 Zhu, B., Su, J., Chang, M.C., Verma, D.P.S., Fan, Y.L. and Wu, R. (1998) Plant Sci. 139, 41-48.
104 Nanjo, T., Kobayashi, T.M., Yoshida, Y., Kakubari, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1999) FEBS Lett. 461, 205-210.
105 De Ronde, J.A., Spreeth, M.H. and Cress, W.A. (2000) Plant Growth Regul. 32, 13-26.
106 Garg, A.K., Kim, J.K., Owens, T.G., Ranwala, A.P., Choi, Y.D., Kochian, L.V. and Wu, R.J. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 15898-15903.
107 Jang, I.C., Oh, S.J., Seo, J.S., Choi, W.B., Song, S.I., Kim, C.H., Kim, Y.S., Seo, H.S., Choi, Y.D., Nahm, B.H. and Kim, J.K. (2003) Plant Physiol. 131, 516-524.
108 Tarczynski, M.C., Jensen, R.G. and Bohnert, H.J. (1993) Science 259, 508-510.
109 Thomas, J.C., Sepahi, M., Arendall, B. and Bohnert, H.J. (1995) Plant Cell Environ. 18, 801-806.
110 Karakas, B., Ozias-Akins, P., Stushnoff, C., Suefferheld, M. and Rieger, M. (1997) Plant Cell Environ. 20, 609-616.
111 Abebe, T., Guenzi, A.C., Martin, B. and Cushman, J.C. (2003) Plant Physiol. 131, 1748-1755.
112 Zhifang, G. and Loescher, W.H. (2003) Plant Cell Environ. 26, 275-283.
113 Sheveleva, E., Chmara, W., Bohnert, H.J. and Jensen, R.G. (1997) Plant Physiol. 115, 1211-1219.
114 Gao, M., Tao, R., Miura, K., Dandekar, A.M. and Sugiura, A. (2001) Plant Sci. 160, 837-845.
115 Sheveleva, E.V., Marquez, S., Chmara, W., Zegeer, A., Jensen, R.G. and Bohnert, H.J. (1998) Plant Physiol. 117, 831-839.
116 Romero, C., Belles, J.M., Vaya, J.L., Serrano, R. and Culianez–Macia, F.A. (1997) Planta 201, 293-297.
117 Urao, T. and Yamaguchi-Shinozaki, K. (2002). JIRCAS 2002 Annual Report 47-48.

118 Xiong, L., Ishitani, M., Lee, H. and Zhu, J.–K. (2001) Plant Cell 13, 2063-2083.
119 Xiong, L., Lee, H., Ishitani, M. and Zhu, J.–K. (2002) J. Biol. Chem. 277, 8588-8596.
120 Abraham, E., Rigo, G., Szekely, G., Nagy, R., Koncz, C. and Szabados L. (2003) Plant Mol. Biol. 51, 363-372.
121 Armengauda, P., Thieryc, L., Buhota, N., Grenier-de Marchb, G. and Savoure, A. (2004) Physiol. Plant. 120, 442-450.
122 Thiery, L., Leprince, A.S., Lefebvre, D., Ghars, M.A., Debarbieux, E. and Savoure, A. (2004) J. Biol. Chem. 279, 14812-14818.
123 Khedr, A.H.A., Abbas, M.A., Wahid, A.A.A., Quick, W.P. and Abogadallah, G.M. (2003) J. Exp. Bot. 54, 2553-2562.
124 Satoh, R., Nakashima, K., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2002) Plant Physiol. 130, 709-719.
125 Oono, Y., Seki, M., Nanjo, T., Narusaka, M., Fujita, M., Satoh, R., Satou, M., Sakurai, T., Ishida, J., Akiyama, K., Iida, K., Maruyama, K., Satoh, S., Yamaguchi–Shinozaki, K. and Shinozaki, K. (2003) Plant J. 34, 868-887.
126 Satoh, R., Fujita, Y., Nakashima, K., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004) Plant Cell Physiol. 45, 309-317.
127 Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000) Curr. Opin. Plant Biol. 3, 217-223.
128 Moons, A., Bauw, G., Prinsen, E., Van Montagu, M. and Van Der Straeten, D. (1995) Plant Physiol. 107, 177-186.
129 Ingram, J. and Bartels, D. (1996) Ann. Rev. Plant Physiol. Plant Mol. Biol. 47, 377-403.
130 Wise, M.J. and Tunnacliffe, A. (2004) Trends Plant Sci. 9, 13-17.
131 Xu, D., Duan, X., Wang, B., Hong, B., Ho, T.D. and Wu, R. (1996) Plant Physiol. 110, 249-257.
132 Thomashow, M.F. (1999) Ann. Rev. Plant Physiol. Plant Mol. Biol. 50, 571-599.
133 Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) Plant J. 34, 137-148.
134 Knight, H., Zarka, D.G., Okamoto, H., Thomashow, M.F. and Knight, M.R. (2004) Plant Physiol. 135, 1710-1717.
135 Urao, T., Katagiri, T., Mizoguchi, T., Yamaguchi-Shinozaki, K., Hayashida, N. and Shinozaki, K. (1994) Mol. Gen. Genet. 244, 331-340.
136 Saijo, Y., Hata, S., Kyozuka, J., Shimamoto, K. and Izui, K. (2000) Plant J. 23, 319-327.
137 Chehab, E.W., Patharkar, O.R., Hegeman, A.D., Taybi, T. and Cushman, J.C. (2004) Plant Physiol. 135, 1430-1446.
138 Patharkar, O.R. and Cushman, J.C. (2000) Plant J. 24, 679-691.
139 Sheen, J. (1996) Science 274, 1900-1902.
140 Townley, H.E. and Knight, M.R. (2002). Plant Physiol. 128, 1169-1172.
141 Viswanathan, C. and Zhu, J.-K. (2002) Philos. Trans. R. Soc. Lond. B Biol. Sci. 357, 877-886.

142 Hwang, I., Sze, H. and Harper, J.F. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 6224-6229.
143 Perruc, E., Charpenteau, M., Ramirez, B.C., Jauneau, A., Galaud, J.P., Ranjeva, R. and Ranty, B. (2004) Plant J. 38, 410-420.
144 Yang, T., Chaudhuri, S., Yang, L., Chen, Y. and Poovaiah, B.W. (2004) J. Biol. Chem. 10.1074/jbc.M402830200.
145 Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 11632-11637.
146 Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) Plant Cell 15, 63-78.
147 Kang, J.Y., Choi, H.I., Im, M.Y. and Kim, S.Y. (2002) Plant Cell 14, 343-357.
148 Villalobos, M.A., Bartels, D. and Iturriaga, G. (2004) Plant Physiol. 135, 309-24.
149 Liu, Q., Sakuma, Y., Abe, H., Kasuga, M., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998) Plant Cell 10, 1391-1406.
150 Haake, V., Cook, D., Riechmann, J.L., Pineda, O., Thomashow, M.F. and Zhang, J.Z. (2002) Plant Physiol. 130, 639-648.
151 Dubouzet, J.G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E.G., Miura, S., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003) Plant J. 33, 751-763.
152 Jaglo–Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998) Science 280, 104-106.
153 Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1999) Nature Biotech. 17, 287-291.
154 Jaglo, K.R., Kleff, S., Amundsen, K.L., Zhang, X., Haake, V., Zhang, J.Z., Deits, T., and Thomashow, M.F. (2001) Plant Physiol. 127, 910-917.
155 Fowler, S. and Thomashow, M.F. (2002) Plant Cell 14, 1675-1690.
156 Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.H., Hong, X., Agarwal, M. and Zhu, J.-K. (2003) Genes Dev. 17, 1043-1054.
157 Park, J.M., Park, C.J., Lee, S.B., Ham, B.K., Shin, R. and Paek, K.H. (2001) Plant Cell 13, 1035-1046.
158 Pellegrineschi, A., Ribaut, J.-M., Trethowan, R., Yamaguchi-Shinozaki, K. and Hoisington, D. (2002) JIRCAS Working Report 2002, 55-60.
159 Hsieh, T.H., Lee, J.T., Charng, Y.Y. and Chan, M.T. (2002) Plant Physiol. 130, 618 - 626.
160 Hsieh, T.H., Lee, L.T., Yang, P.T., Chiu, L.H., Charng, Y.Y., Wang, Y.C. and Chan, M.T. (2002) Plant Physiol. 129, 1086-1094.
161 Gilmour, S.J., Sebolt, A.M., Salazar, M.P., Everard, J.D. and Thomashow, M.F. (2000) Plant Physiol. 124, 1854-1865.
162 Gomez, J.M., Hernandez, J.A., Jimenez, A., del Rio, L.A. and Sevilla, F. (1999) Free Radic. Res. 31, S11-S18.
163 Lee, D.H., Kim, Y.S. and Lee, C.B. (2001) J. Plant Physiol. 158, 737-745.

164 Sreenivasulu, N., Miranda, M., Prakash, H.S., Wobus, U. and Weschke, W. (2004) J. Plant Physiol. 161, 467-77.
165 Noctor, G. and Foyer, C.H. (1998) Ann. Rev. Plant Physiol. Plant Mol. Biol. 49, 249-279.
166 Hasegawa, P.M., Bressan, R.A., Zhu, J.-K. and Bohnert, H.J. (2000) Ann. Rev. Plant Physiol. Plant Mol. Biol. 51, 463-499.
167 Møller, I.M. (2001) Annu. Rev. Plant Physiol. Plant Mol. Biol. 52, 561-591.
168 Alscher, R.G., Erturk, N. and Heath, L.S. (2002) J. Exp. Bot. 53, 1331-1341.
169 Wang J., H. Zhang, and R.D. Allen (1999) Plant Cell Physiol. 40, 725-732.
170 Roxas, V.P., Smith Jr, R.K., Allen, E.R. and Allen, R.D. (1997) Nat. Biotechnol. 15, 988-991.
171 Roxas, V.P., Lodhi, S.A., Garrett, D.K., Mahan, J.R. and Allen, R.D. (2000) Plant Cell Physiol. 41, 1229-1234.
172 Badawi, G.H, Kawano, N., Yamauchi, Y., Shimada, E., Sasaki, R., Kubo, A. and Tanaka, K. (2004) Physiol. Plant 121, 231-238.
173 Tsugane, K., Kobayashi, K., Niwa, Y., Ohba, Y., Wada, K. and Kobayashi, H. (1999) Plant Cell 11, 1195-1206.
174 Guan, L.Q.M., Zhao, J. and Scandalios, J.G. (2000). Plant J. 22, 87-95.
175 Pei, Z.M., Murata, Y., Benning, G., Thomine, S., Klusener, B., Allen, G.J., Grill, E. and Schroeder, J.I. (2000) Nature 406, 731-734.
176 Kwak, J.M., Mori, I.C., Pei, Z.M., Leonhardt, N., Torres, M.A., Dangl, J.L., Bloom, R.E., Bodde, S., Jones, J.D.G. and Schroeder, J.I. (2003) EMBO J. 22, 2623-2633.
177 Murata, Y., Pei, Z.M., Mori, I.C. and Schroeder, J.I. (2001) Plant Cell 12, 2513-2523.
178 Levine, A., Tenhaken, R., Dixon, R. and Lamb, C. (1994) Cell 79, 583-593.
179 Chamnongpol, S., Willekens, H., Moeder, W., Langebartels, C., Sandermann Jr, H., Van Montagu, M., Inzé, D. and Van Camp, W. (1998) Proc. Nat. Acad. Sci. U.S.A. 95, 5818-5823.
180 Vranova, E., Inze, D. and Van Breusegem, F. (2002) J. Exp. Bot. 53, 1227-1236.
181 Chinnusamy, V., Schumaker, K. and Zhu, J.-K. (2004) J. Exp. Bot. 55, 225-236.
182 Samis, K., Bowley, S. and McKersie, B. (2002) J. Exp. Bot. 53, 1343-1350.
183 Alvarez, M.E., Pennell, R.I., Meijer, P.-J., Ishikawa, A., Dixon, R.A. and Lamb, C. (1998) Cell 92, 773-784.
184 Apel, K. and Hirt, H. (2004) Annu. Rev. Plant Biol. 55, 373-399.
185 Laloi, C., Apel, K. and Danon, A. (2004) Curr. Opin. Plant Biol. 7, 323-328.
186 Chinnusamy, V. and Zhu, J.–K. (2003) Top. Curr. Genet. 4, 241-270.
187 Arabidopsis Genome Initiative. (2000) Nature 408, 796-815.

188 Ichimura, K., Mizoguchi, T., Irie, K., Morris, P., Giraudat, J., Matsumoto, K. and Shinozaki, K. (1998) Biochem. Biophys. Res. Comm. 253, 532-543.
189 Teige, M., Scheikl, E., Eulgem, T., Doczi, R., Ichimura, K., Shinozaki, K., Dangl, J.L. and Hirt, H. (2004) Mol. Cell 15, 141-152.
190 Mizoguchi, T., Irie, K., Hirayama, T., Hayashida, N., Yamaguchi-Shinozaki, K., Matsumoto, K. and Shinozaki, K. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 765-769.
191 Ichimura, K., Mizoguchi, T., Yoshida, R., Yuasa, T. and Shinozaki, K. (2000) Plant J. 24, 655-665.
192 Huang, Y., Li, H., Gupta, R., Morris, P.C., Luan, S. and Kieber, J.J. (2000b) Plant Physiol. 122, 1301-1310.
193 Moon, H., Lee, B., Choi, G., Shin, D., Prasad, D.T., Lee, O., Kwak, S-S., Kim, D.H., Nam, J., Bahk, J., Hong, J.C., Lee, S.Y., Cho, M.J., Lim, C.O. and Yun, D-J. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 358-363.
194 Kovtun, Y., Chiu, W-L., Tena, G. and Sheen, J. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 2940-2945.
195 Ulm, R., Ichimura, K., Mizoguchi, T., Peck, S.C., Zhu, T., Wang, X., Shinozaki, K. and Paszkowski, J. (2002) EMBO J. 21, 6483-6493.
196 Xiong, L. and Yang, Y. (2003) Plant Cell 15, 745-759.
197 Flowers, T.J., Koyama, M.L., Flowers, S.A., Sudhakar, C., Singh, K.P. and Yeo, A.R. (2000) J. Exp. Bot. 51, 99-106.
198 Koyama, M.L., Levesley, A., Koebner, R.M.D., Flowers, T.J. and Yeo, A.R. (2001) Plant Physiol. 125, 406-422.
199 Bonilla, P., Dvorak, J., Mackill, D.J., Deal, K. and Gregorio, G.B. (2002) Philipp. Agric. Sci. 85, 68-76.
200 Lin, H.X., Zhu, M.Z., Yano, M., Gao, J.P., Liang, Z.W., Su, W.A., Hu, X.H., Ren, Z.H., and Chao, D.Y. (2004). Theor. Appl. Genet. 108, 253-60.
201 Monforte, A.J., Asins, M.J. and Carbonell, E.A. (1997) Theor. Appl. Genet. 95, 706-713.
202 Foolad, M.R., Chen, F.Q. and Lin, G.Y. (1998) Theor. Appl. Genet. 97, 1133-1144.
203 Foolad, M.R. and Chen, F.Q. (1999) Theor. Appl. Genet.99, 235-243.
204 Quesada, V., Garcia-Martinez, S., Piqueras, P., Ponce, M.R. and Micol, J.L. (2002) Plant Physiol. 130, 951-63.

STRATEGIES FOR HIGH-THROUGHPUT GENE CLONING AND EXPRESSION

L. J. Dieckman, W. C. Hanly, and F. R. Collart

Biosciences Division
Argonne National Laboratory
Argonne, Illinois 60439

INTRODUCTION

The wealth of data from genomic sequencing projects has led to an increased interest in the development of strategies for high-throughput cloning and expression. The emphasis on the high-throughput component is in part attributable to the initiation of large scale programs such as structural genomics, which mandate the development of automated approaches to facilitate an increased rate of structure determination (1-3). The development of protein chips (4-6) and genomic-scale interaction screens (7, 8) has further stimulated the expansion of high-throughput cloning and expression strategies. However, the experimental approach for development of automated systems for gene cloning and expression is inherently different from classical methods for cloning and expression (9-11). The high-throughput capability of an automated system is achieved at the expense of system flexibility and, as a consequence of this constraint, these strategies usually incur a higher rate of target attrition than more traditional benchtop or low-throughput approaches. The establishment of an

Genetic Engineering, Volume 27, Edited by J. K. Setlow

automated process also requires a more global approach for the evaluation and implementation of cloning and expression protocols (12). Because modifications of established automation methods are expensive and demand significant amounts of time for rewriting and revalidation, protocols must be evaluated at the inception of the program with respect to their compatibility with other method protocols and for feasibility of implementation in an automated setting. We developed a high-throughput cloning and expression strategy from target to validated expression clone that provides a clone resource for the Midwest Center for Structural Genomics (MCSG) (13). This strategy evolved after evaluation of three critical elements common to many high-throughput processes: targets, methods, and screening requirements. Integration of these considerations into a series of methods results in an efficient process that has been scaled to generate thousands of *E. coli* clones. The pipeline incorporates molecular tools that facilitate implementation of parallel processes and allow scaling of the components to meet increasing throughput demands and adapt to changing target characteristics. This chapter will summarize key elements of this process and provide a perspective on high-throughput method development strategies.

AUTOMATION PLATFORM

Commercially available liquid handlers, hardware, and components with standard microplate formats enable integration of high-throughput automation into most basic research departments. The Molecular Biology Robot System at Argonne National Laboratory includes a plate transfer robot (ORCA, Beckman Coulter, Inc., Fullerton, CA) that traverses a 3-m rail system. Adjacent to the rail are a number of stations (such as pipetting workstations, a plate washer, shaker, heatblock, barcode reader, incubators, etc.) that perform the equivalent of standard laboratory molecular biology manipulations during an automation procedure. The liquid handling stations include Beckman Coulter Biomek 2000 and Multimek workstations. The Biomek 2000 workstation incorporates a filtration station for purification of plasmids and amplified fragment DNA, a gripper device to allow for movement of labware, and thermal reservoirs to allow for heating/cooling of microwell plates. The system is controlled *via* the SAMI NT software package that provides a graphical interface for the development, scheduling, and implementation of methods on the system. For most applications, the graphical interface enables benchtop-trained scientists to design and implement methods without the need for a dedicated automation specialist.

HIGH-THROUGHPUT STRATEGY

Targets

The cloning and expression pipeline producing clones for the MCSG is one of the Protein Structure Initiative pilot centers funded by the National Institute of General Medical Sciences at the National Institutes of Health (1). The structural genomic target set of the MCSG represents mostly microbial targets characterized as cytoplasmic proteins. The characteristics of the target set

suggest *Escherichia coli* as a logical first choice of expression system in view of its demonstrated utility for expression of microbial proteins (14, 15). This system represents an efficient approach to produce proteins quickly, in large amounts, and in a cost efficient manner (16, 17) and is the standard platform used by a number of structural genomics centers as the primary protein production platform (12, 18-20).

Although prokaryotic expression systems have many advantages for both small- and large-scale protein expression, they have some limitations due to the inability of a prokaryotic system to produce proteins in as complex a manner as a eukaryotic cell (21-23). However, a number of laboratories (9, 19) have implemented large-scale platforms based in whole or in part on the *E. coli* expression system and have developed high-throughput methods for successful expression of eukaryotic proteins. These platforms typically employ multiple strategies involving both genetic design and protein expression cassettes in *E. coli* to maximize the generation of soluble proteins or protein domains for downstream analysis (19, 22, 24). A common approach is to use a bacterial expression system as an initial platform and screen for the production of soluble proteins. Targets that fail in the initial round can then be routed through salvage pathways that utilize alternative expression strategies. This tiered strategy leverages the efficient and cost-effective high-throughput processes that are already available for the production of proteins in *E. coli* before proceeding to more expensive and time intensive alternative approaches.

The characteristics of the target set extend beyond the choice of expression system(s) and impact the core methods and design of the protein production pipeline. Targets for the structural genomics pilot centers are selected by a criterion of less than 30% sequence identity to sequences in the Protein Data Bank. This constraint results in a target group containing large numbers of uncharacterized and hypothetical proteins that represent a challenge for expression in a soluble form. For the structural genomics centers, the metric for success at the expression level is the production of a clone expressing soluble protein at a level that enables purification of a sufficient amount of protein for crystallization trials. Large-scale expression studies addressing cytoplasmic targets suggest a capability to express approximately 30-50% of targets in a soluble form (13, 25). However, it can be anticipated that the process for the generation of clones expressing soluble protein at the level required for crystallization screening will become more difficult as the "low-hanging fruit" component of the target set is depleted (26, 27). These considerations have led to the development and implementation of high-throughput screening strategies for identification of clones expressing soluble protein that allow for higher overall throughput, reduced costs, and significant improvement of the efficiency of the production process.

High-Throughput Methods

The cloning and expression strategy for protein production represents a critical decision element of the high-throughput production process. This decision impacts most elements of the process, including the front-end bioinformatics, vectors, and targets, and the selection of methods appropriate for implementation in

an automated environment. The cloning strategy also influences the outcome metrics of cost, efficiency. and throughput. For many of the structural genomic pilot centers, the need to insure a constant supply of validated expression clones for crystallization trials has led to the implementation of parallel cloning strategies to ameliorate some attrition due to the production of low solubility or insoluble proteins. This type of parallel strategy relies on a universal cloning site approach for cloning and expression of targets in multiple vectors and from many sources. Although several universal cloning site systems are presently available, we selected the Ligation Independent Cloning (LIC) method (28, 29) for implementation into our high-throughput protein production pipeline. The selection of this method was based on system characteristics that facilitated implementation of the process in an automated environment as well as global considerations such as cost and efficiency. Some of the major attributes of this system that impacted the decision to select the LIC method as a core cloning strategy are summarized below:

- Our analysis of the heterogeneous character (a large component of hypothetical and uncharacterized open reading frames, ORFs), of the targets selected for structural genomics indicated that multiple expression systems would be advantageous to achieve a representative array of clones that expressed a soluble protein product. The capability to utilize multiple vectors and hosts in an automated process to generate and screen expression clones is an essential component to increase the target success rate for many high-throughput protein expression strategies. In the LIC approach, universal cloning sites can be incorporated at the primer design stage, enabling a general cloning approach to most of the selected targets (30). This characteristic enables implementation of parallel methods that utilize multiple vectors. Although a variety of vectors are available for the *E. coli* hosts, the selection of an optimal vector is dependent on programmatic goals and automation requirements (9, 31, 32). Our experience and that of others, however, indicates it is often necessary to design an array of compatible vector systems to provide for flexibility of expression with different fusion tags and protease cleavage sites (33, 34).
- The LIC approach employs a directional cloning method that simplifies robotic implementation by reducing the number of processes necessary to generate a validated clone. The directional nature allows for direct expression screening and facilitates the development of efficient screening methods for soluble expression products.
- The LIC cloning method does not involve restriction enzymes. This consideration eliminates restriction site screening as a component of the target selection process and enables a single cloning protocol to be applied to all targets.
- A characteristic of the LIC cloning method and some other universal cloning systems is the relatively simple methodologies for clone preparation. In automated environments, reducing the method complexity facilitates implementation and eliminates the occurrence of processing errors. The LIC approach is easily adaptable to an automation system inasmuch as the method requires only one enzymatic step that occurs at room temperature. Furthermore, the time required for processing of plates is minimal and the reagents used in the process are stable at the cold block temperatures that are maintained during transport on the automation system.

- A critical step in the cloning process is the combination of the target DNA with the vector. In the LIC method, this occurs *via* an annealing reaction conducted at room temperature. Our studies of various fragment to vector ratios (Figure 1) indicate a wide tolerance for variation in the amount of target DNA fragment on the annealing reaction. This latitude eliminates the need for normalization of fragment concentrations prior to annealing, thus conserving time and simplifying the process for implementation of the method as an automated process.
- The LIC method is highly efficient and cost effective. In a large-scale study of more than 880 targets from *Bacillus subtilis*, our analysis of individual LIC expression clones produced in the microwell plates indicated an overall expression efficiency of approximately 70% for all targets and vectors (35). Manual analysis of multiple clones for each target (four clones from each target were analyzed by denaturing gel electrophoresis) demonstrated expression of the target protein in greater than 75% of screened clones (13). This high expression efficiency provides several options for implementing expression screening procedures. One option implemented at the MCSG includes screening of nonclonal plasmid stocks for expression and solubility prior to clone isolation and storage. This process is desirable when the expected rate of target attrition is likely to be high to avoid time intensive cloning procedures for nonproductive clones. In any case, the high expression efficiency associated with the LIC method minimizes the amount of downstream effort required for the selection and validation of individual expression clones.

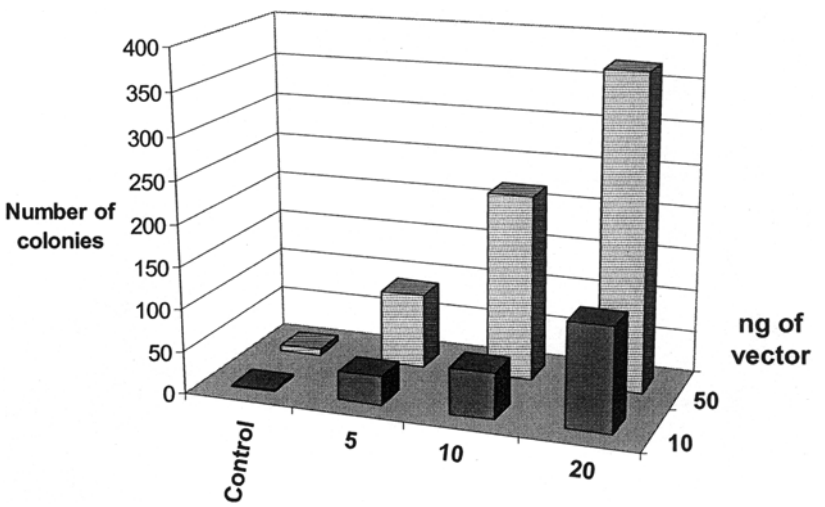


Figure 1. Dependence of colony formation at various fragment to vector ratios. The indicated amounts of LIC fragment and vector were annealed for 10 minutes prior to transformation. Control samples contained vector but no added LIC fragment.

The LIC cloning method was selected as a core strategy because these characteristics matched the requirements of the structural genomics program. Specific advantages include efficiency, cost, and ability to implement parallel approaches for different vectors. A disadvantage of the LIC approach is the reliance on the polymerase chain reaction (PCR) fragment as the cloning entry point. Due to the high attrition for structural genomics targets, individual targets are not cloned until a target is tested for soluble protein expression. Although this represents a cost-effective method for clone production, modifications of the LIC approach (36) as well as alternative approaches (20, 37) have been successfully implemented in the Protein Structure Initiative pilot centers.

Tag Detection Screening for Soluble Proteins

For the MCSG and many other high-throughput protein production centers, the measure of success for high-throughput cloning and expression component is the generation of a clone expressing a soluble protein product. The historical success rate for production of clones expressing soluble proteins (<50%) and the uncharacterized nature of the target group, suggested that implementation of a microwell plate–based screening method could reduce the amount of time spent on expression and solubility validation of nonproductive clones. This realization has led to the development of a number of high-throughput screening strategies for soluble proteins (38, 39) with approaches ranging from tag detection (40) to genetic endpoints (41, 42). For a high-throughput screening method to be effective it must be rapid and reproducible, and it must be able to predict which clones will be able to produce soluble proteins in culture. The day-to-day variations must be sufficiently low so that the test does not need to be repeated for reliable results. The purpose of the high-throughput screening is not to produce proteins of immediate use to the investigator but to indicate which clones will express soluble proteins for large-scale expression and further functional or structural studies. When working with large numbers of clones at a time, as is frequently done in high-throughput protein production, it is useful to evaluate carefully as many clones as possible before conducting manual and/or large-scale purification. The ultimate goal of the screening process is to identify a population of clones that are likely to give a high success rate in large scale protein purification (43-45).

We developed an expression screen for production of soluble protein that uses a tag detection strategy to screen for production of protein containing the 6x histidine component of the fusion tag (33). India HIS Probe-HRP (Pierce Chemical, Rockford, IL), a nickel activated horseradish peroxidase, is used for detection of the 6x histidine component of the fusion tag. The assay was developed using standard 96-well plates containing 88 target clones and 8 control wells. The four-day procedure enables processing of up to eight plates of clones with standard liquid handlers (13, 35). On the first day of the process, plasmid DNA is transformed into chemically competent BL21 cells, which are cultured overnight. The next day, the overnight cultures are diluted and incubated at 37°C (to log phase absorbance) before induction with isopropyl thiogalactoside. After two hours of further incubation, cells are lysed by a two-step

process with a sodium phosphate–buffered solution containing lysozyme, benzonase nuclease, and a 25% detergent solution. The tag detection assay is run on the third day but can be performed on the second day, provided sufficient time has elapsed for adsorption of the expressed proteins to the Immulon 4HBX plates.

The characteristics of the tag detection screen were assessed by intensive screening of 2 of the 10 plates of targets from a MCSG *Bacillus subtilis* genome screen (35). These plates, designated Bsub04 and Bsub08, were used for all of the following experiments. Reproducibility of the tag detection assay was assessed by conducting a screen on samples generated from the same bacterial growth culture and performing the assay on successive days. Differences were calculated by subtracting each sample's rank on day 1 from day 2 of the tag detection assay (the ranking procedure is described in the legend to Figure 2). An average deviation of approximately five positions was observed in the ranking of all 88 samples (Figure 2), suggesting this approach has value as a preliminary screen for detection of the fusion tag. The overall reproducibility of the experiment from bacterial transformation to the tag detection assay was assessed by averaging all ranking data for all of the assays from each plate (Bsub04 and Bsub08). We observed an overall standard deviation of approximately 12 rank positions averaged over all assays and plates (Table 1). These variations are attributable in part to the multiple pipetting and plate washing procedures on the robotic system, which occur over the three-day process. Variations in bacterial culture growth were also observed (not shown), and most likely contributed to the ranking differences found in Table 1.

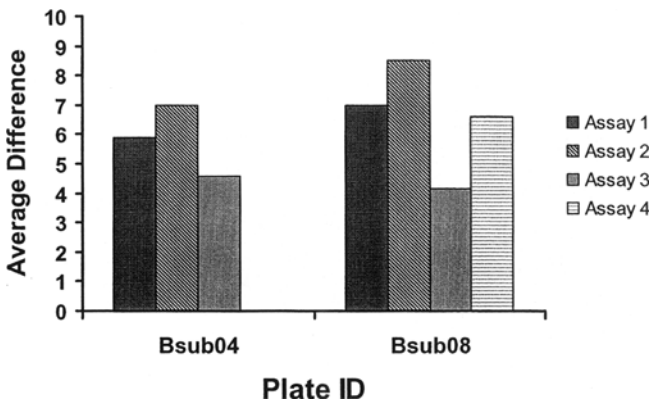


Figure 2. Day-to-day variation of target ranking in tag-detection screen with the same induced growth samples. The average differences and standard deviations were calculated from the absolute values of all 88 samples. The 88 plate samples in individual plates were ranked such that the well with the highest absorbance at 450 nm received a score of 1, whereas the well with the lowest absorbance at 450 nm received a score of 88. Ranking data were pooled for nine assay plates representing four different bacterial growth dates for Bsub04 clones and 11 assay plates representing five different bacterial growth dates for Bsub08 clones (Table 1). Average ranking score and standard deviations were calculated from the absolute values of all 88 samples for the nine assay plates.

Table 1. Overall reproducibility of tag detection assay.

	Assays	Range	Average Std. Deviation
Bsub04	9	1.00-25.56	11.01
Bsub08	11	2.73-27.38	13.40

The ability of the tag detection assay to predict expression of soluble proteins was evaluated by comparison of the results from the tag detection assay to those obtained from SDS-PAGE gel analysis of the soluble fractions. The data were pooled for all tag detection assays and each target in the plate assigned to one of four ranking groups (legend to Figure 2). Each of the groups contained 22 targets sorted so that the highest ranked group contained the targets with the highest scores 1-22 in the tag detection assay. (A score of 1 indicated the highest amount of color development on the assay plate.) For the Bsub08 plate, a total of 58 soluble positive expression clones were found. Twenty-one positive expression clones were found in the highest ranking group, with 20, 12, and 5 positive expression clones found in the remaining groups as ranked from highest to lowest scores (Figure 3). A similar trend was observed in the Bsub04 plate in that the targets from the tag detection assay with the highest scores were most likely to be associated with identification of a soluble protein band after SDS-PAGE analysis (Figure 3).

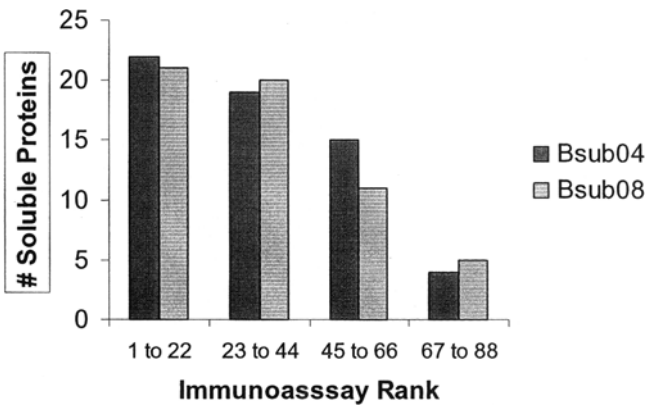


Figure 3. Summary of solubility *versus* relative rank as determined by the tag detection assay. For Bsub04, ranking data were pooled for nine assay plates representing four different bacterial growth dates. For Bsub08, ranking data were pooled for 11 assay plates representing five different bacterial growth dates. For solubility data, SDS-PAGE gels were evaluated by visually analyzing and scoring the gel according to the protein band found on the gel. If no protein band was apparent in the correct molecular weight region, the protein was given a score of 0. Scores of 1 and 2 were given to bands that were present indicating low and high solubility, respectively. All clones receiving a score over 0 were considered positive expression clones.

These data show that a positive expression clone can be predicted from assay rank. The highest-ranked 25%, or highest 22 from the Bsub04 assay ranking data, gave 22 soluble expression clones as analyzed by SDS-PAGE. Twenty-one out of 22 of these proteins received the highest solubility score of 2 (see legend to Fig. 3 for description of solubility level assignments). In the lowest-ranked 25% of the plate by assay, four soluble clones were found; three of these four received solubility scores of 1. The highest-ranked 25%, or highest 22 from the Bsub08 assay ranking data, gave 21 soluble expression clones as analyzed by SDS-PAGE. Sixteen out of 22 of these proteins received the highest solubility score. In the lowest-ranked 25% of the assay plate, five soluble clones were found; three of these received solubility scores of 1.

In order to test the ability of the assay to evaluate the soluble expression clones in the context of a single screen (the norm for an HTP production run), we analyzed tag detection results from a single plate and set of SDS-PAGE gels. For the Bsub04 plate, the highest ranked 22 clones on the assay plate gave 22 soluble expression clones as analyzed by SDS-PAGE (Figure 4). In the lowest-ranked 25% of the plate, one soluble expression clone was found by SDS-PAGE. The highest-ranked 25%, or highest 22 clones from the Bsub08 plate, gave 21 soluble expression clones as analyzed by SDS-PAGE. In the lowest-ranked 25% of the plate by tag detection assay, three soluble clones were found. These data show a positive expression clone can be predicted from the tag detection assay rank and thus eliminate downstream screening cost and time.

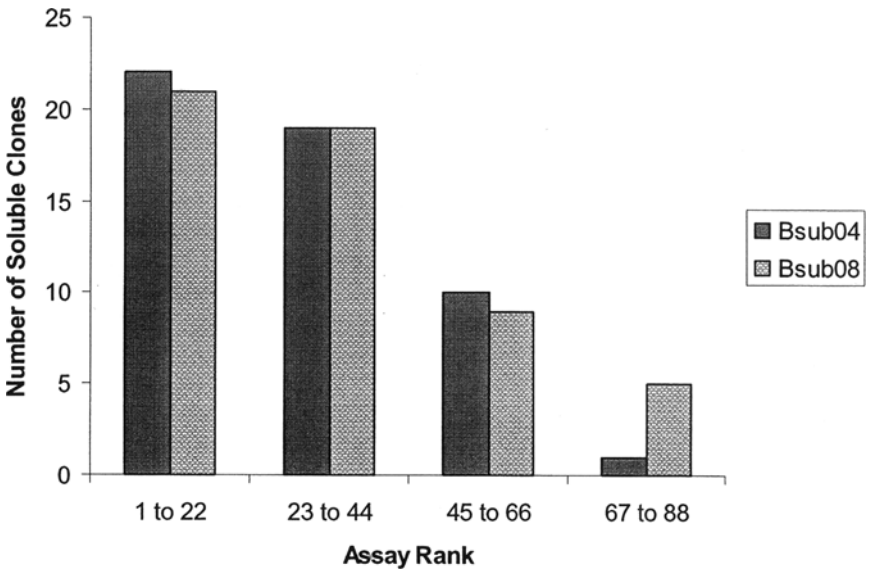


Figure 4. Prediction capability of the tag detection screen from a single growth plate. A single induced bacterial culture plate was used to generate two tag detection assay plates and one set of SDS-PAGE gels. Ranking data for the two days were averaged and plotted against the solubility data.

Although the data from the tag detection assay correlated well with the prediction of soluble clones, there were a few proteins in which the solubility rankings and SDS-PAGE results did not correlate. Several of the proteins that scored high in the tag detection assay ranking did not produce a soluble clone on SDS-PAGE. A protein that is cleaved or incompletely transcribed would produce a small peptide that may not be detectable on SDS-PAGE and yet give a strong signal on the tag detection assay. In addition, several proteins that received low rankings in the tag detection assay produced soluble proteins when screened by SDS-PAGE. These may represent proteins in which the his tag is buried or unavailable to the his-probe; they would not be detected on the tag detection screen even if they were expressed and soluble.

SUMMARY

High-throughput approaches for gene cloning and expression require the development of new, nonstandard tools for use by molecular biologists and biochemists. We have developed and implemented a series of methods that enable the production of expression constructs in 96-well plate format. A screening process is described that facilitates the identification of bacterial clones expressing soluble protein. Application of the solubility screen then provides a plate map that identifies the location of wells containing clones producing soluble proteins. A series of semi-automated methods can then be applied for validation of solubility and production of freezer stocks for the protein production group. This process provides an 80% success rate for the identification of clones producing soluble protein and results in a significant decrease in the level of effort required for the labor-intensive components of validation and preparation of freezer stocks. This process is customized for large-scale structural genomics programs that rely on the production of large amounts of soluble proteins for crystallization trials.

ACKNOWLEDGMENTS

The authors thank Andrzej Joachimiak and Fred Stevens for their support of the high-throughput cloning and expression efforts. This work was supported by National Institutes of Health Grant GM62414-01, A. Joachimiak, PI, and by the U.S. Department of Energy, Office of Health and Environmental Research, under contract W-31-109-Eng-38.

REFERENCES

1 Norvell, J.C. and Machalek, A.Z. (2000) Nat. Struct. Biol. 7 Suppl., 931.
2 Schmid, M.B. (2004) Nat. Rev. Microbiol. 2(9), 739-746.
3 Zhang, C. and Kim, S.H. (2003) Curr. Opin. Chem. Biol. 7(1), 28-32.
4 Jung, G.Y. and Stephanopoulos, G. (2004) Science 304, 428-431.
5 Lee, Y.S. and Mrksich, M. (2002) Trends Biotechnol. 20(12 Suppl.), S14-S18.
6 Phelan, M.L. and Nock, S. (2003) Proteomics 3(11), 2123-2134.

7 Lehner, B. and Fraser, A.G. (2004) Genome Biol. 5(9), R63.
8 Terradot, L. Durnell, N., Li, M., Li, J., Ory, J., Labigne, A., Legrain, P., Colland, F. and Waksman, G. (2004) Mol. Cell. Proteomics 3(8), 809-819.
9 Chambers, S.P. (2002) Drug Discov. Today 7(14), 759-765.
10 Endo, Y. and Sawasaki, T. (2003) Biotechnol. Adv. 21(8), 695-713.
11 Sarafanov, A. and Saenko, E. (2004) Anal. Biochem. 328(1), 98-100.
12 Lesley, S.A. (2001) Protein Expr. Purif. 22(2), 159-164.
13 Dieckman, L., Gu, M., Stols, L., Donnelly, M.I. and Collart, F.R. (2002) Protein Expr. Purif. 25(1), 1-7.
14 Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J., Beasley, S., Lan, N., Das, R., Gerstein, M., Arrowmith, C.H. and Edwards, A.M. (2003) Proteins 50(3), 392-399.
15 Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M. and Arrowsmith, C.H.. (2003) Acc. Chem. Res. 36(3), 183-189.
16 Baneyx, F. (1999) Curr. Opin. Biotechnol. 10(5), 411-421.
17 Baneyx, F. and Palumbo, J.L. (2003) Methods Mol. Biol. 205, 171-197.
18 Chance, M.R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G., Fajardo, J.E., Radhakannan, T. and Marinkovic, N. (2004) Genome Res. 14(10B), 2145-2154.
19 Folkers, G.E., van Buuren, B.N. and Kaptein, R. (2004) J. Struct. Funct. Genomics 5(1-2), 119-131.
20 Yokoyama, S. (2003) Curr. Opin. Chem. Biol. 7(1), 39-43.
21 Baneyx, F. (1999) Curr. Opin. Biotechnol. 10(5), 411-421.
22 Chambers, S.P., Austen, D.A., Fulghum, J.R. and Kim, W.M. (2004) Protein Expr. Purif. 36(1), 40-47.
23 Laage, R. and Langosch, D. (2001) Traffic 2(2), 99-104.
24 Uchimura, Y., Nakamura, M., Sugasawa, K., Nakao, M. and Saitoh, H. (2004) Anal. Biochem. 331(1), 204-206.
25 DiDonato, M., Deacon, A.M., Klock, H.E., McMullan, D. and Lesley, S.A. (2004) J. Struct. Funct. Genomics 5(1-2), 133-146.
26 Janda, I., Devedjiev, Y., Cooper, D., Chruszcz, M., Derewenda, U., Gabrys, A., Minor, W., Joachimiak, A. and Derewenda, Z.S. (2004) Acta Crystallogr. D Biol. Crystallogr. 60(Pt 6), 1101-1107.
27 Lundstrom, K. (2004) Comb. Chem. High Throughput. Screen. 7(5), 431-439.
28 Haun, R.S., Serventi, I.M. and Moss, J. (1992) Biotechniques 13(4), 515-518.
29 Aslanidis, C. and de Jong, P.J. (1990) Nucl. Acids Res. 18(20), 6069-6074.
30 Yoon, J.R., Laible, P.D., Gu, M., Scott, H.N. and Collart, F.R. (2002) Biotechniques 33(6), 1328-1333.
31 di Guan, C., Li, P., Riggs, P.D. and Inouye, H. (1988) Gene 67(1), 21-30.
32 Routzahn, K.M. and Waugh, D.S. (2002) J. Struct. Funct. Genomics 2(2), 83-92.
33 Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F.R. and Donnelly, M.I. (2002) Protein Expr. Purif. 25(1), 8-15.

34 Woestenenk, E.A., Hammarstrom, M., van den Berg, S., Hard, T. and Berglund, H. (2004) J. Struct. Funct. Genomics 5(3), 217-229.
35 Moy, S., Dieckman, L., Schiffer, M., Maltsev, N., Yu, G.X. and Collart, F.R. (2004) J. Struct. Funct. Genomics 5(1-2), 103-109.
36 Nguyen, H., Martinez, B., Oganesyan, N. and Kim, R. (2004) J. Struct. Funct. Genomics 5(1-2), 23-27.
37 Finley, J.B., Qiu, S.H., Luan, C.H. and Luo, M. (2004) Protein Expr. Purif. 34(1), 49-55.
38 Coleman, M.A., Lao, V.H., Segelke, B.W. and Beernink, P.T. (2004) J. Proteome Res. 3(5), 1024-1032.
39 Shih, Y.P., Kung, W.M., Chen, J.C., Yeh, C.H., Wang, A.H. and Wang, T.F. (2002) Protein Sci. 11(7), 1714-1719.
40 Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H. and Hard, T. (2002) Protein Sci. 11(2), 313-321.
41 Lesley, S.A., Graziano, J., Cho, C.Y., Knuth, M.W. and Klock, H.E. (2002) Protein Eng. 15(2), 153-160.
42 Waldo, G.S. (2003) Curr. Opin. Chem. Biol. 7(1), 33-38.
43 Millard, C.S., Stols, L., Quartey, P., Kim, Y., Dementieva, I. and Donnelly, M.I. (2003) Protein Expr. Purif. 29(2), 311-320.
44 Prinz, B., Schultchen, J., Rydzewski, R., Holz, C., Boettner, M., Stahl, U. and Lang, C. (2004) J. Struct. Funct. Genomics 5(1-2), 29-44.
45 Stols, L., Millard, C.S., Dementieva, I. and Donnelly, M.I. (2004) J. Struct. Funct. Genomics 5(1-2), 95-102.

MOLECULAR ROLES OF CHAPERONES IN ASSISTED FOLDING AND ASSEMBLY OF PROTEINS

Mark T. Fisher

Department of Biochemistry and Molecular Biology
University of Kansas Medical Center
Kansas City, Kansas 66160

INTRODUCTION

Within the last 15 years, there has been a major paradigm shift in our view of *in vivo* protein folding. This shift comes from the discovery of large families of special proteins called molecular chaperones that play essential roles in influencing or interacting with proteins as they proceed toward their folded or final assembled states. There are many different chaperone proteins that have been identified that interact with various protein folding populations, using a variety of direct or indirect mechanisms to influencing the protein folding reactions (Figure 1). In the strictest sense, a protein chaperone is a protein that aids in the folding of a substrate protein but does not become part of the final folded structure.

In vivo, proteins fold and assemble in very crowded molecular environments, containing protein concentrations that sometimes approach 200-300 mg/mL. In addition, the kinetics of protein folding and assembly reactions is complex and occurs over a large range of time (subμsec to hours). Conceptually, protein folding has been proposed to progress over a multitude of different energy states

Genetic Engineering, Volume 27, Edited by J. K. Setlow

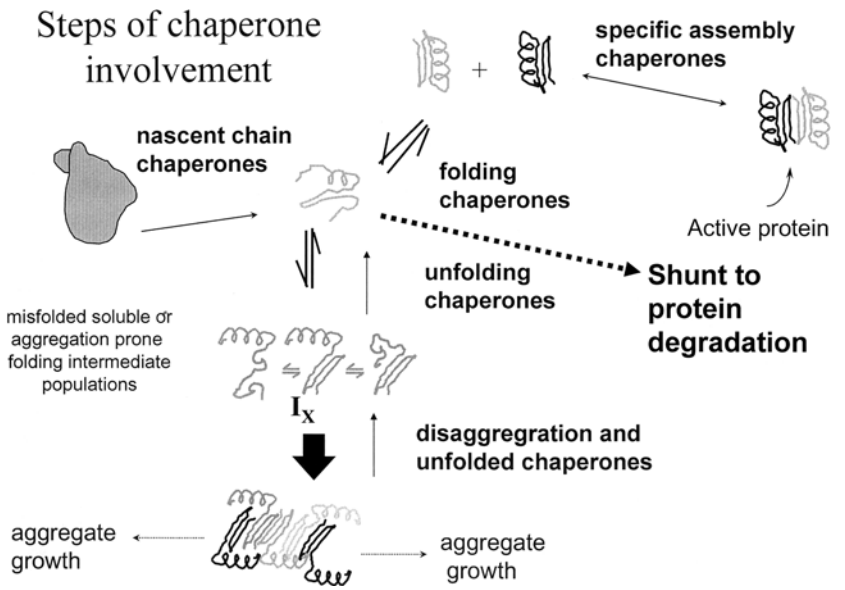


Figure 1. Chaperone proteins are found to play a role in almost every step of protein folding and assembly.

represented by a rugged folding landscape (1, 2). One prediction of this landscape feature is that folding protein rates should be heterogeneous, indicative of multiple rather than single folding trajectories (Figure 2). Indeed, folding heterogeneity is inferred from the presence of nonexponential kinetics as measured from fast kinetic experiments (3, 4).

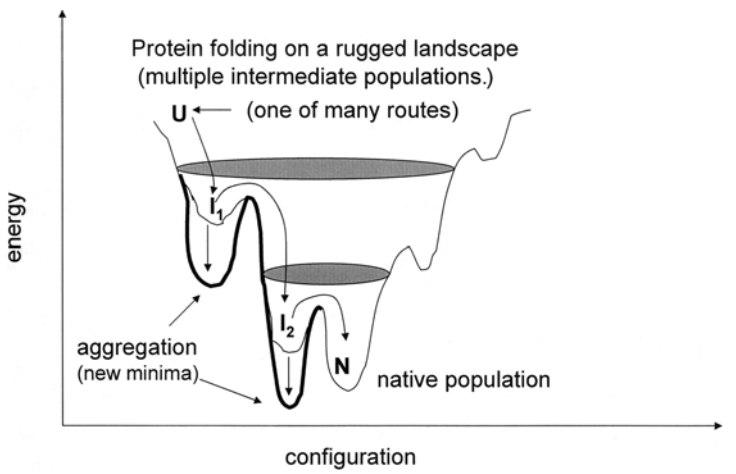


Figure 2. Folding funnel—Protein folding over a rugged energy surface. One route is highlighted (solid arrows) to include the possible changes in the energy surface of a series of folding intermediate populations that become trapped in lower energy minima resulting in further stabilizing aggregation reactions.

Unconstrained single-protein molecule measurements indicate that a multitude of small unfolding/folding steps are observed, suggestive of the presence of multiple small energy wells or trapping events, in agreement with the concept of a global folding landscape (5). Slower transitions were also observed due to coupled transitions or motions perhaps related to domain pairing events or other low-frequency vibrational modes. During protein folding, the presence of multiple fluctuating states is problematic, particularly in situations where these fluctuating conformers become momentarily stable (metastable states) and exist for extended periods of time. These semi-stable kinetic states can sometimes result in a buildup of folding intermediates, leading to deleterious protein misfolding aggregation reactions. Conceptually, simpler folding transitions will lead to a smoother protein folding energy landscape, decreasing the tendency to form misfolded or aggregated populations.

Chaperones' interactions with folding proteins depend on the folding kinetics of protein substrate and the physiochemical properties of the transient folding intermediates, whose formation is ultimately dictated by the primary sequence. However, these rates of folding/and or unfolding also depend on the environmental conditions in the cell (i.e., temperature, post-translational modification, etc.). These rates change when environmental parameters change or when specific missense mutations appear within the primary sequence. In the cell, various chaperones regulate and control the conformations that proteins acquire and are critical in maintaining protein homeostasis. Molecular chaperones interact with actively folding proteins; partially folded, kinetically trapped proteins; or even more globally unfolded states. Consequently, these protein chaperones prevent large-scale protein misfolding or aggregation by rescuing malfolded proteins or by shuttling them toward degradation pathways. A large subset of the chaperone proteins that we will discuss in this review were initially identified as heat shock proteins (Hsp). The levels of these Hsp molecular chaperones increase in response to cellular stress, where they serve to prevent or reverse protein misfolding and aggregation within the cell.

Understanding the general mechanisms of molecular chaperones is complicated because proteins have to achieve their final folded states by a variety of different mechanisms *in vivo*. For example, a significant portion of proteins that are synthesized in the cytoplasm have to be unfolded before they can be transported into other organelles within the cell. To participate in this transport process, premature folding has to be prevented and the protein substrate must be maintained in a partially folded state by molecular chaperones. In addition to protein transport, some proteins, particularly oligomeric proteins, temporarily exist as aggregation prone states before they acquire an assembly competent state. The accumulated presence of aggregation prone folding intermediates is often the major cause of general protein misfolding.

It is well established that the protein primary sequence dictates the rate and formation of the transient folding intermediates as well as the final fold within a particular environment. Understandably, the lifetime, fold, and subsequent folding rates of some folding intermediates are usually very sensitive to temperature variations. Any situation (mutation, environmental) that slows the normal folding rate may result in the formation of temperature-sensitive folding intermediates (6).

Lowered temperatures or even small molecule additives can help these proteins fold, sometimes to states that are just as stable as their native counterparts (7, 8). Unfortunately, successful *in vivo* folding of these mutant proteins cannot be achieved, leading to a number of protein-folding diseases. Interestingly, these mutant folding reactions can sometimes be reversed by chaperone proteins at physiological temperature and concentration conditions, leading to the prospect of designing chaperone-dependent therapeutic strategies to combat protein folding diseases.

Folding reactions often result in the formation of transient folding intermediates. Mechanistically, chaperone-substrate protein interactions depend on the transient formation of specific or general binding sites with the folding intermediate that will be recognized by the resident chaperone proteins. Thus, the binding interactions between chaperone proteins and folding intermediates are governed, in large part, by the thermodynamics and kinetic properties of the folding intermediates. The kinetics of the partitioning interaction also depend on the identities of the chaperones as well as on the concentrations of both of the interacting species.

Most of the current information about chaperone proteins has focused on the identification, classification, and documentation of the important chaperones or chaperone complexes. The process of categorizing and describing the multiple chaperones, co-chaperones, and their protein substrates are critical and absolutely necessary first steps in our understanding of chaperone mechanism and function. However, a better understanding of the molecular mechanism can be gained by examining how chaperones directly influence the properties of the folding protein. In this review, we shall focus on the chaperone mechanisms that influence protein folding and assembly, highlighting the increasingly important role that chaperones play in influencing the protein conformation of stable protein populations. Given the limited space of this review, we are not able to discuss the enzymic chaperones—protein disulfide isomerase (PDI), with its related reduction system, and the peptidyl prolyl *cis-trans* isomerase (PPI). Excellent descriptions of these two enzymic chaperones are found in reviews by Wilkinson and Gilbert (9) and Schiene and Fischer (10), respectively. In addition, we will not be able to discuss the molecular chaperones or proposed molecular mechanisms for protein folding and transport within the endoplasmic reticulum (ER). The reader is referred to an expert review on this system by Ellgaard and Helenius (11). However, a number of the chaperones discussed in this review do have homologous molecular counterparts located within the ER. In the first part of the review, the major chaperones involved in protein folding will be introduced. In the second major section, we shall examine and define the molecular mechanisms of chaperone-mediated protein folding for these chaperones.

MAJOR CHAPERONES IN PROTEIN FOLDING

The concept that proteins or biomolecules are able to direct the folding of other proteins was first suggested by Anfinsen and colleagues (12). Although this was a novel concept at the time, discovery and validation of such processes did not occur until the late 1980s. Pioneering genetic and structural work by C.

Georgopoulos and colleagues (13) provided important hints about the existence of molecular chaperones because they noticed that a number of the essential bacterial host proteins were essential to aid in the folding and assembly of a number of specific phage coat proteins. These initial host helper proteins that were initially identified turned out to be heat shock proteins. Homologous protein classes were subsequently found to assist in the folding and assembly of proteins in higher eukaryotes (14, 15). A series of landmark experiments launched the field of chaperone-mediated folding and transport and provided the first clues in the molecular mechanisms of these fascinating proteins (16-18).

A majority, but not all, of the molecular chaperones that have been studied over the past 20 years were initially classified as heat shock proteins (collectively called Hsp followed by the molecular mass). These chaperones show substantial increases in their levels following a heat stress to the organism (19). Fortunately, a number of these proteins are extremely easy to overproduce and purify, making them amenable to crucial biophysical studies. In terms of cellular mass, the major chaperones are very abundant and they are constitutively expressed. In this section we shall briefly describe the properties of the five major classes of heat shock proteins that function as chaperone proteins.

The Hsp70 Chaperone Class—Protein Folding, Unfolding

The Hsp70 chaperone class is one of the most prominent and diverse of all of the protein chaperones. Hsp70 proteins are involved in nascent and cellular protein folding, protein trafficking, protein unfolding, protein degradation, protein disassembly, and protein conformational regulation. Because of its abundance and ease of purification, the most commonly studied Hsp70 system is the DnaK/DnaJ/GrpE system isolated from *Escherichia coli*. Generally, the Hsp70 binds to an extended polypeptide chain that consists primarily of a stretch of aliphatic residues (typically around seven residues in length) that is usually buried in the interior of folded proteins (20, 21). These peptides bind tightly to the Hsp 70 (DnaK in *E. coli*) chaperones with tight binding affinities. The co-crystallization of the truncated Hsp70 peptide binding domain (DnaK) and a short hydrophobic peptide (NRLLLTG) (22) indicate that the Hsp70 class binds substrate peptides in an extended conformation.

The different Hsp70 isoforms are variable in sequence within their substrate binding domain and cavity, lending strong credence to the hypothesis that the differences in the specificity of action of the various Hsp70s are determined by the differences in the binding affinity of various polypeptides (see review 23). In addition to this apparent substrate binding variability, each nucleotide hydrolysis factor (NHF) and nucleotide exchange factor (NEF) interacts with a specific Hsp70 isoform. In general, the NHF and NEF are not interchangeable among other Hsp70 system variants. Like Hsp70, the hydrolysis factor (Hsp40 or DnaJ or J-like protein) also associates with the partially folded substrate (Figure 3). The simultaneous binding of Hsp70 and the nucleotide exchange factor Hsp40 to the same protein substrate forms a localized ternary complex where the Hsp40 can more effectively stimulate ATP hydrolysis rates of Hsp70 (24). Because in *E. coli* the Hsp40 (DnaJ) is present in such low amounts compared with Hsp70 (25),

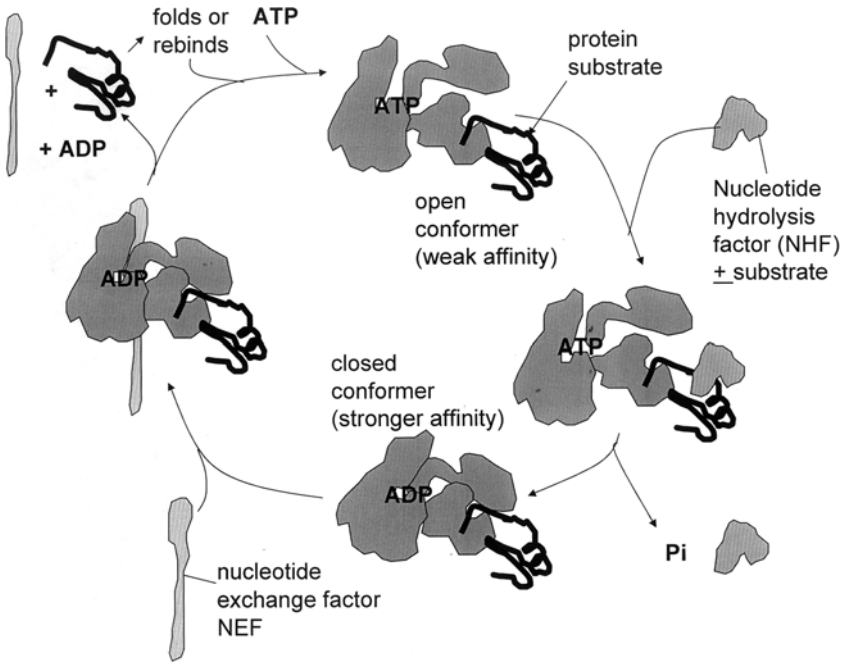


Figure 3. Cycle of Hsp70. The open conformer of Hsp70 binds a polypeptide in the presence of ATP. The association rate is rapid for this conformer. A nucleotide hydrolysis factor (commonly DnaJ in *E. coli* or other J-like proteins in eukaryotes) forms a ternary complex with DnaK and the partially folded polypeptide, leading to an acceleration of the hydrolysis reaction and resulting in a closed (tight-binding) conformer of Hsp70 (DnaK in *E. coli*). A nucleotide exchange factor (GrpE in *E. coli*; BAG in eukaryotes) accelerates the exchange rate of bound ADP, ATP rebinds, and the protein substrate dissociates, continues to fold, or rebinds. Once a new protein substrate binds, the cycle begins anew.

the formation of this ternary complex with the protein substrate increases the effective intramolecular collision frequency between Hsp70 and Hsp40. Following ATP hydrolysis, GrpE binds to Hsp70 (DnaK) and stimulates ADP dissociation and an ATP exchange, followed by the dissociation of GrpE and the protein substrate from DnaK in a concerted manner (26). After release, the protein can be transported, fold, assemble, proceed toward degradation, interact with other chaperones, or rebind to the Hsp70 chaperone again to restart the cycle.

The Chaperonin Class—Versatile Protein Folders

The chaperonin classes of molecular chaperones (group I or Hsp60/10 and group II CCT eukaryotic chaperonin) are large oligomeric allosteric ring systems. The group I chaperonins (GroE chaperonins in *E. coli* and mtcpn60/10 in mitochondria and chloroplasts) are, in most cases, absolutely essential chaperone proteins. The group I chaperonin is made up of a dimer of homologous heptamers, arranged in a double-stacked ring structure, with each of the seven

homologous subunits having a molecular mass of around 60 kDa. The class II chaperonins are also arranged as a double-stacked ring system but usually contain eight rather than the seven membered rings found in the group I chaperone class. The eukaryotic cytoplasmic form, CCT (chaperonin containing TCP-1 or sometimes called TriC) contains anywhere from two to eight related subunits per ring. The archaebacterial forms of the TCP-1 subfamily of chaperonins contain eight or nine subunits in the ring. Whereas the eukaryotic chaperonins are not as essential for cell viability as the group I chaperonins, there are numerous disease states that occur when these eukaryotic chaperonins are missing or are malfunctioning (27).

The arrangement of the subunits in the chaperonins forms a continuous binding surface around the inner rings. These chaperonins interact with partially folded, quasi-native folding intermediates or newly imported proteins. The eubacterial chaperonin (GroE in *E. coli*) binds a multitude of protein substrates and this chaperonin class is estimated to assist in the folding of 3-10% of the total proteins within bacteria (28). Hartl's group has designed efficient traps with an engineered GroE chaperonin set in an effort to identify nascent polypeptide chains that require the chaperonin. These generally interact through tight binding interactions, but recently it has been shown through *in vitro* experiments that weaker binding interactions between protein substrates and the chaperonins may be necessary to maintain active conformations (29-32).

Group I Chaperonins

The group I chaperonin system (particularly the GroE chaperonin from *E. coli*) is the most extensively studied and characterized of all the molecular chaperone proteins. The protein binding component of the group I chaperonins is comprised of a large double ring of heptamers (~800 kDa) arranged in such a manner that they contain an extensive hydrophobic binding surface within their inner ring cavities, enabling them to bind easily a multitude of hydrophobic folding intermediates. In addition to binding the hydrophobic protein substrate, the group I chaperonins can bind ATP and a small single ring heptameric cochaperonin factor (~70 kDa), allowing the protein binding component to switch its binding strength from a strong to a weak affinity, allowing the bound hydrophobic protein substrate to dissociate and fold. The group I chaperonins are also referred to as Hsp60/Hsp10 or chaperonin 60/10 (cpn60/10), classified according to their molecular masses. The group I chaperonin is able to bind a very large subset of different protein conformations and the interaction depends on the degree of surface hydrophobicity of the protein substrate. In general, partially folded protein states are perhaps the most prominent substrate for the chaperonin proteins because this form often contains exposed hydrophobic regions. However, the group I chaperonins can bind a variety of early and late folding intermediates, allowing it to bind a wide array of protein substrates. In addition to its substrate binding versatility, the GroE group I chaperonin employs a number of distinct molecular mechanisms to fold proteins.

The steps involved in the general mechanism of this chaperonin-assisted or -mediated folding are described as a series of vectorial processes driven by

ATP binding and hydrolysis (Figure 4). The GroEL oligomer is allosterically regulated by ATP binding, evident as nested cooperativity, which involves positive intra-ring cooperativity and negative inter-ring cooperativity (33). The GroEL oligomer contains 14 homologous subunits (57 kDa each), and each subunit contains three functional domains. The apical domain binds protein substrates and the cochaperonin GroES, the equatorial domain binds and hydrolyzes ATP, and the intermediate domain links and transmits allosteric signals between the apical and equatorial domains. Within the heptamer ring, GroEL allosterically binds and hydrolyzes seven ATP molecules in a positive cooperative manner (34, 35). In addition to ATP binding and hydrolysis, the protein substrate binding influences the intra-ring positive cooperativity by opposing the intra-ring ATP binding (36). Within the heptamer, the antagonistic protein substrate and ATP binding controls concerted shifts between two opposing global conformations. One global conformation, referred to as the taut or T state, is observed to bind protein substrate favorably and has a weaker affinity for ATP. ATP is proposed to preferably bind to the other global conformation, the relaxed or R state, resulting in the observed decrease in the substrate protein binding. The protein substrate affinity is further weakened when the cochaperonin GroES preferentially binds to the nucleotide-bound

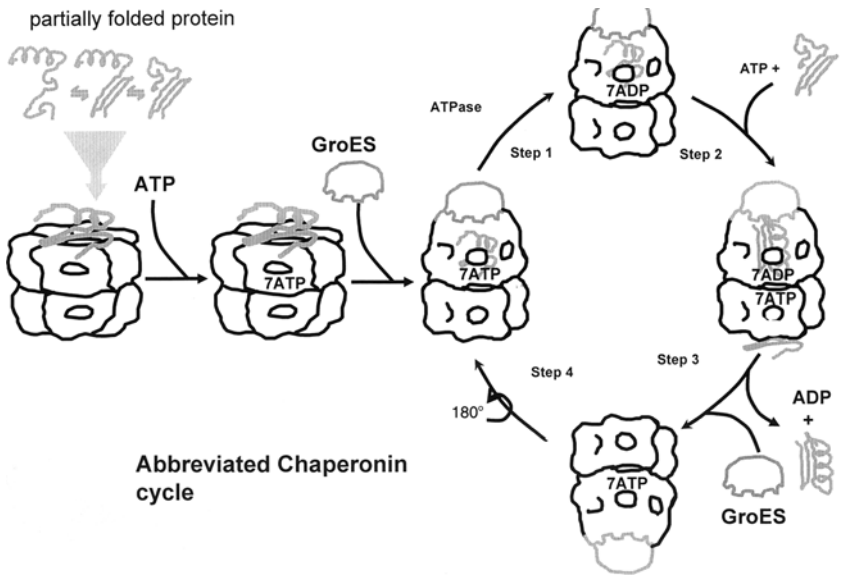


Figure 4. Group I chaperonin mechanism Step 1—In the presence of bound substrate and ATP, the cochaperonin GroES binds to GroEL, encapsulates, and releases the protein substrate inside a nanostructured interior cavity. Step 2—The protein substrate folds within the cavity as ATP is hydrolyzed to ADP. This particular step is presumed to be the slow step in the chaperonin cycle. Step 3—ATP and another protein substrate bind to the opposing ring, facilitating the release of trapped (and hopefully) folded substrate and ADP. Step 4—The opposing ring is now encapsulated with the next GroES cochaperonin molecule and the cycle continues.

GroEL, triggering massive conformational changes and creating a nano-structured chamber inside the GroEL-GroES complex, which momentarily encapsulates small- to medium-sized folding protein substrates (~20-50 kDa), allowing them to fold sequestered from bulk solvent (37).

The influence of the protein substrate on the GroEL inter-ring allostery is most important during the functional chaperonin cycle. The particular allostery manifests negative cooperative effects on inter-ring ATP, polypeptide, and GroES binding. When ATP, polypeptide, or GroES bind to one ring, the opposing ring affinities for ATP, polypeptide, or GroES binding decrease. Functionally, the long-range inter-ring transmission of binding energy of the polypeptide (over ~140Å) has dramatic effects on the chaperonin cycle. Horwich and co-workers observed that polypeptide binding on one ring facilitates the dissociation of bound GroES and entrapped protein on the opposing ring (38). Due to thermodynamic reciprocity, GroES binding to one heptamer facilitates the dissociation of the protein substrate from the opposite ring (39-42).

Group II Chaperonins

In contrast to the group I chaperonin class (GroE), a smaller amount of biochemical information is available to describe the mechanism of protein folding in the presence of the group II chaperonins (archaebacterial thermosome or eukaryotic CCT or TriC). However, much of what is known about the protein-folding mechanism of the class II chaperonins has come primarily from *in vitro* protein translation assays, substrate preferences, co-factor interactions, and structural details from cryo EM microscopy or X ray crystallography of particular intermediates in the chaperonin cycle. The eukaryotic chaperonins appear to be more specific than the group I chaperonins. The group II chaperonin has been found to interact with specific cytoskeletal proteins and various signal transduction proteins.

Although the CCT chaperonins are also double-ring structures like their group I counterparts (e.g., GroE chaperonins), they don't require a cochaperonin to encapsulate the nano-structured folding chamber. Another difference is seen when one compares the protein substrate interaction site of the group I and the group II CCT chaperonins. The group I chaperonins bind through exposed hydrophobic sites with partially unfolded hydrophobic substrates, whereas the eukaryotic group II CCT chaperonins prefer to interact with more specific quasi-native conformations of substrates like actin and tubulin through hydrophilic or charged residues. Since CCT is present in a much lower abundance than the more common molecular chaperone/heat shock proteins, the collisional frequency between CCT and the protein substrate may be enhanced by the prior formation of an auxiliary cofactor (prefoldin or GimC)–protein substrate complex (43).

Once the substrates associate with the CCT chaperonin, ATP binding forces the closure of a series of apical helical extensions to capture and encloses the protein substrate inside the newly formed internal cavity. Although this step is controversial, the substrate may either be released inside the cavity to fold or it may be actively folded into a collapsed form that remains bound to the roof of

the CCT (Figure 5) (44). More details of the molecular mechanism of the CCT-induced folding are presented later in this review.

Hsp90—The Multipurpose Chaperone

The Hsp90 chaperone family plays an interesting and diverse role in protein homeostasis. Unlike most chaperones, the Hsp90 class is not exclusively required nor is it involved in nascent protein folding. Even so, these chaperones are essential for viability, particularly in eukaryotes. Also, the eukaryotic Hsp90 contains two paralogs that interact with many proteins that are involved in signal transduction. Hsp90s are necessary to maintain active conformational states for steroid hormone receptors and various kinases. The mechanism of Hsp90 action is complicated because numerous cochaperone factors also associate with Hsp90 to modulate its ATPase activity, thus leading to a control of its protein substrate binding and release of the activated substrates (Figure 6). Presently, the only detailed molecular study completed to date suggests that the Hsp90 class may participate in activating proteins by using a general unfolding mechanism (21). Hence, the global mechanism for Hsp90 action will only be achieved when the detailed molecular interactions and mechanisms of binding between Hsp90 and other potential protein substrates are measured.

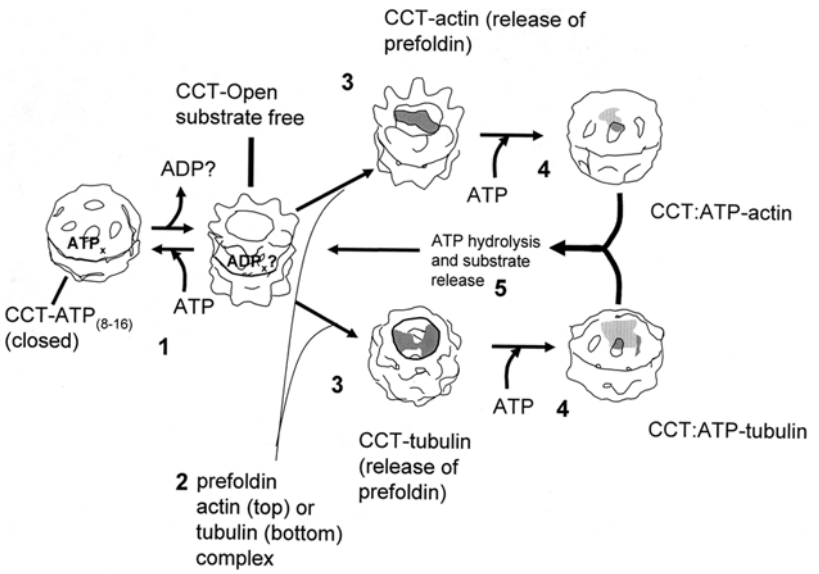


Figure 5. One possible mechanism of the CCT cycle with specific protein substrates, actin (upper branch) and tubulin (lower branch). This figure was modified from the discussion presented in (44). Step 1—The chaperonin cycles between a closed and open conformation, governed by ATP hydrolysis and/or ADP dissociation. Steps 2 and 3—The prefoldin protein substrate complex transports the partially folded substrate to CCT and dissociates resulting in a CCT protein substrate complex. Step 4—ATP binds to the chaperonin-protein substrate complex, momentarily encapsulating the substrate within the chaperonin cavity. Step 5—Subsequent hydrolysis of ATP induces the release of the bound substrate and cycles the CCT to the substrate accepting state (Step 2).

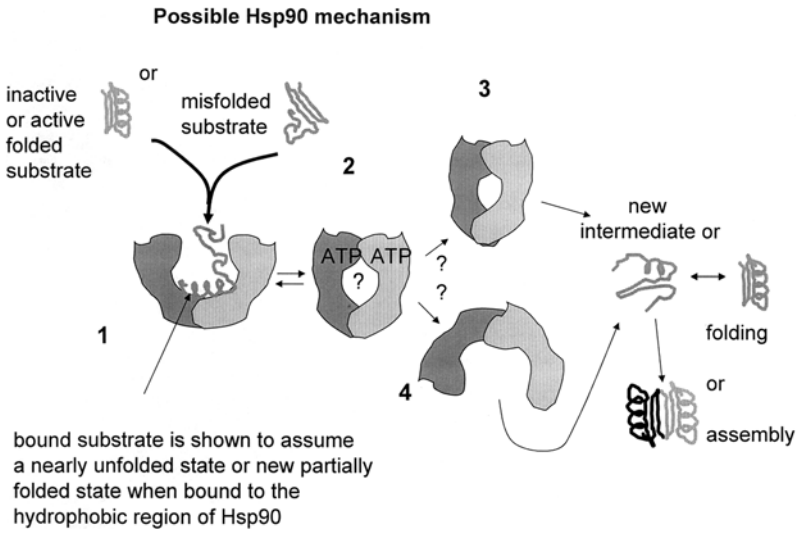


Figure 6. Possible Hsp90 mechanisms. The Hsp90 molecular chaperone is a dimer with an N-terminal ATP binding domain, a middle domain, and a C-terminal substrate binding domain. The Hsp90 is thought to alter the conformation of the interacting substrate protein (mechanism 1), unfold it, and release it to a different active state, either folded, active (partially unfolded), or assembly competent. The hypothetical mechanisms of ATP-dependent coupling and substrate release include a sequestration (mechanism 2), a coupled release (mechanism 3), or hinged release (through the ATP binding domains—mechanism 4) (based on the Hsp90 structure and discussion of data provided in references 21 and 45).

Hsp100 Class—Unfoldases

The Hsp100 chaperone proteins are specialized proteins that participate in disaggregating denatured proteins rather than in folding them directly. The disaggregated proteins can either proceed toward folding with the aid of other chaperone proteins or they can proceed toward proteolysis within large internal protease complexes. The Hsp100s, classified as members of the class I AAA ATPases, primarily exist as hexamers and contain two ATP binding domains per subunit. It appears that ATP is required for both function and assembly. As is usually the case, most of our knowledge about the Hsp100 class of proteins has been deduced from studying the bacterially or yeast-derived systems. As we shall see, the Hsp104 (yeast) or ClpB (bacterial) Hsp100 chaperone proteins are necessary in rescuing denatured proteins from their aggregated states so that they can be refolded by other chaperone proteins. These proteins were first characterized as proteins that are important for heat tolerance, where an increased intracellular concentration allows cells to endure heat stress–induced protein aggregation more efficiently. Another set of bacterial Hsp100 proteins (ClpA and ClpY) are involved in shuttling proteins toward degradation by the appropriate large molecular mass protease. ClpA(Y) proteins physically bind and facilitate the peptidase function of a specific protease called ClpP whereas the ClpB proteins do not interact with these peptidases.

To help reactivate and refold proteins, the unfolding function of some of the Hsp100 proteins require the assistance of the resident DnaK/Hsp70 systems. The prominent structural differences between the homologous ClpA and the ClpB systems are located within a middle domain where ClpB contains longer coiled-coil heptad repeats. Deletion of these regions in ClpB leads to low ATPase activity and inhibits proper ClpB assembly (46). Furthermore, most of these mutants were unable to protect null ClpB *E.coli* from surviving elevated heat stress. Initially the structural differences between ClpA and ClpB were thought to result in different interaction mechanisms with their respective protein substrates. As we shall see, these Hsp100 proteins may use very similar mechanisms of unfolding to either degrade or refold substrate proteins (Figure 7).

The Small Hsp Chaperone Class—Aggregation Prevention

The final class of molecular chaperones that we will consider in this review are the small heat shock proteins. These proteins are molecular chaperones that are mainly involved in preventing large-scale aggregation of heat-denatured proteins (47). These small heat shock proteins exist as oligomers that have the ability to dissociate, capture, and sequester partially folded aggregation-prone protein substrates. Once the protein substrates are captured, the sHsp-denatured protein substrate complexes reassemble into a larger oligomer (Figure 8) (48, 49). Upon return to normal temperatures or with the help of

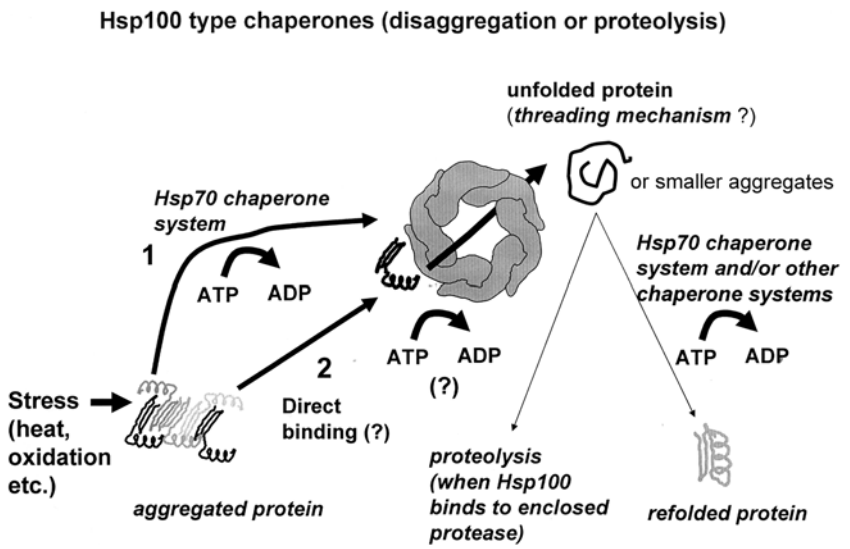


Figure 7. Hsp100s are primarily involved in protein disaggregation. The disaggregation can occur with the help of auxiliary Hsp70 proteins (pathway 1) or by direct binding of the Hsp100 to the aggregate (pathway 2). The individual proteins are either unfolded by a threading mechanism (pathway 1) or Hsp100 can act directly on the large aggregate to disrupt it into smaller, perhaps more manageable aggregates (pathway 2).

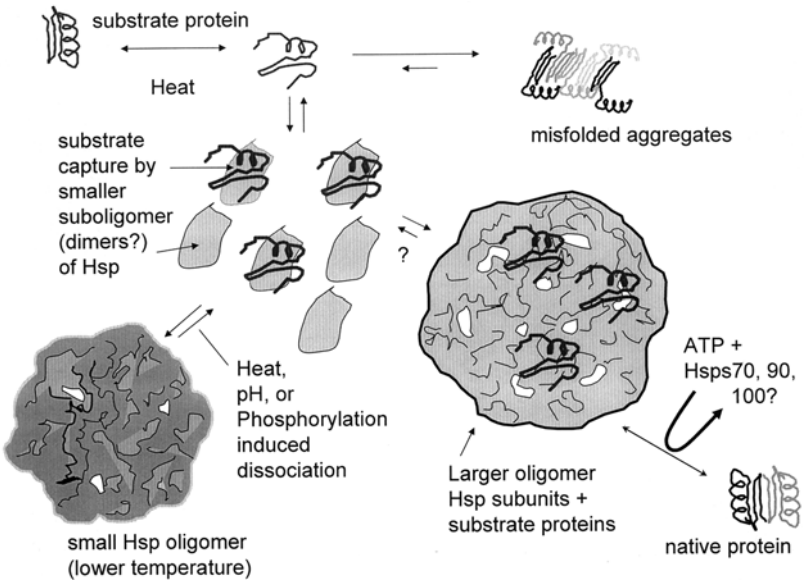


Figure 8. Current mechanism of sHsp. The small heat shock protein is initially present as an oligomer that undergoes a stress dependent dissociation to smaller oligomeric or subunits. These subunits associate with partially folded substrates, capturing these and preventing further accumulation of larger aggregates. The captured substrates can be stored or refolded with the aid of other prominent auxiliary chaperones.

other chaperones such as Hsp70, Hsp90, etc., the substrate protein can refold from its initial trapped state, presumably on the sHsp surface. Most of the small Hsps tested thus far exist as dynamic oligomers whose assembly stoichiometry and size depend on the temperature and the amount of denatured proteins that eventually become bound to the chaperone (Figure 8). In eukaryotes, the oligomers undergo a specific phosphorylation modification that may be an additional control feature for controlling oligomer size, affecting both sHsp oligomer stability and protein capture efficiencies (50). In addition to phosphorylation, there have also been recent reports that ATP can bind to the small Hsp α-crystallin protein (both ATP and α-crystallin present in the eye lens), leading to a stabilization of the oligomeric structure and facilitating capture of substrate proteins (51).

DEFINING THE MOLECULAR MECHANISMS OF CHAPERONE-MEDIATED PROTEIN FOLDING AND ASSEMBLY

Perhaps one of the most exciting areas of current molecule chaperone research is focused on elucidating the molecular mechanisms of chaperone-mediated protein folding. Early measurements of the molecular mechanisms of chaperone action were useful in our initial analysis of chaperone mechanism but they had the drawback of being imprecise and indirect. For example, most protein substrates that were used in chaperone research were enzymes, enabling

researchers to track endpoint folding rates and yields of activity regain as a function of time. Alternatively, because molecular chaperones in general inhibit protein aggregation, measuring the chaperone-dependent decrease in light scattering was also a popular method to examine chaperone effects on the protein folding reaction. Although these indirect measurements reflect the global effects of chaperone influences on protein folding, they cannot provide the sufficient molecular detail to ultimately define the molecular mechanisms involved in protein folding and assembly. Indeed, there are numerous instances where the above-mentioned indirect measurements have led to the false identification of proteins as chaperone proteins. For example, proteins such as casein, IgM, bovine serum albumin, or, in some cases, simple polypeptides (poly-L-Lysine) can also enhance folding yields, increase disaggregation, or decrease large-scale protein aggregation (52-54). Given the many different solution conditions that can enhance folding and decrease aggregation, it is not surprising that there are other examples where proteins that are not classified as true chaperones could appear to have properties of authentic chaperones. In the following sections, we will examine the effects that chaperones have on the protein substrate itself to get a clearer picture of the specific and unique folding mechanisms that are employed by authentic molecular chaperones. As we shall see, chaperones can fold particular proteins using a variety of different molecular mechanisms. Some chaperones provide very specific steric influences on protein folding, some induce localized folding, other chaperones unfold incorrectly folded proteins, and some simply serve as buffers to maintain active protein structure (protein homeostasis), whereas others appear to create nano-structured cavities leading to unique sequestered folding mechanisms.

Chaperones Inducing Folded Structure (Specific and Nonspecific)

It is well established that the formation of local protein fold motifs or short peptide secondary structures are dependent on the environment surrounding them (their context). This simple physical property is an important parameter to keep in mind when describing the potential molecular consequences of some protein-chaperone interactions and folding. For example, the interaction of loosely structured. transiently exposed hydrophobic regions from the protein substrates can interact with either general nonspecific hydrophobic regions or specific structured regions on chaperone proteins. Some chaperones can nonspecifically induce local secondary structure within unstructured regions, which may in turn cause the formation of more specific interior interactions, thus avoiding off-pathway misfolding or aggregation. In addition, it is also noted that chaperones may accelerate some protein domain pairing reactions in cases where domain-domain interactions of tertiary structures are kinetically slow in the absence of chaperone proteins (55). This section examines some specific and general chaperone-induced protein folding. In terms of the concept of the folding landscape, the formation of nonspecific or specific induced folded regions will change the shape of the folding funnel, narrowing or smoothing the folding funnel, allowing the polypeptide to avoid local kinetic traps or energy minima (Figure 9).

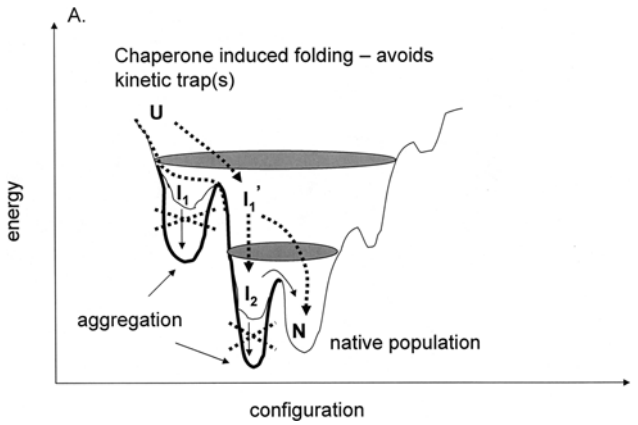


Figure 9. Local folding influences the folding landscape. Schematically, inducing a local fold in a folding protein may facilitate local context specific interactions within the protein, allowing protein substrate to bypass a particular folding and aggregation trap. This bypass can also describe limited but favorable intramolecular domain pairing interaction. The dotted lines schematically represent the influence that chaperone proteins have on reshaping the folding funnel, potentially changing the flux toward the native configuration.

Molecular Chaperones Involved in Specific Folding

There are a number of highly specific chaperones that only mediate the folding of very specific substrates. Presently, the cytoskeletal protein tubulin appears to have the most identified specific chaperone proteins associated with both its folding and assembly. Numerous chaperones are involved in stabilizing quasi-native intermediate species of the tubulin. Currently, the protein substrate tubulin appears to interact with the group II cytosolic chaperonin CCT. Single-particle cryo-electron microscopy reconstructions of the CCT-tubulin (or actin) complexes suggest that the protein substrates, tubulin and actin, form very specific complexes with CCT (Figure 5) (56). The reconstruction indicates that the volume of the partially folded tubulin almost matches the newly observed protein density that appears in the binding cavity of the eukaryotic chaperonin protein.

For the tubulin (α and β) subunits, more specific auxiliary chaperone interact with these proteins once they are released from the CCT chaperonin. The released tubulins interact with a set of four to five very specific chaperone proteins nebulously described as co-factors A, B, C, D, and E. Because these co-factor proteins are not part of the final dimer or fibril structures, they qualify as true chaperone proteins (57). These downstream chaperones directly interact with α tubulin, (cofactors A) and β tubulin (cofactor B), and the specific chaperone tubulin complexes form larger assemblies with the additional co-factors C, D, and E. In the presence of GTP, cochaperone C facilitates the chaperone complex-tubulin breakdown, finally resulting in the release and formation of the assembly competent $\alpha\beta$ tubulin dimer.

At a detailed molecular level, we still do not understand how tubulin (and actin) folding and assembly is specifically aided by all of these specific chaperones.

However, analysis of the structure of one of these specific co-factors indicates that it has a hydrophilic binding surface (58). This may make functional sense when one considers that the tubulins already are highly folded (called quasi-native states) prior to finally folding into their assembly competent forms (59). One hypothesis advanced by Valpuesta, Willison, and colleagues proposes that the CCT facilitates tubulin or actin folding by a direct mechanism (44). It is presumed that slow folding forms or stable kinetic intermediates may require the precise binding interactions provided of these specific auxiliary chaperones to surmount activation energy barriers that normally inhibit the formation of assembly competent tubulin subunits during unchaperoned folding. Both the CCT chaperonin and the auxiliary chaperones may directly facilitate the formation of the assembly competent forms of the tubulins by insuring that proper intra- and intermolecular interaction domains are brought into close proximity to each other within and between the heterologous subunits. In these cases, the favorable and specific chaperone binding interactions will literally inhibit the intramolecular movements within the folding intermediates, decreasing the entropic freedom of the partially folded states, providing the structure context to ultimately drive the acquisition of properly folded tertiary and quaternary assembly competent tubulin structures.

This mechanism predicts that there should be some specific interaction sites on the CCT chaperonin that will bind to specific regions of these structural proteins (56). The CCT chaperonin appears to facilitate the final fold acquisition for protein substrates that have already been significantly folded. Much like the Hsp90 class, the group II chaperonins may function as general buffer chaperones for proteins that are far removed from their largely unfolded nascent states. Although CCT certainly facilitates folding of some proteins within *in vitro* reticulocyte lysates, this mechanism need not involve nascent chain interactions. The CCT chaperone appears to preferentially interact with folded but inactive kinetically trapped conformers. Still, much more experimental work is needed, particularly with other recently determined substrates, before we can decipher the global molecular mechanism of these fascinating eukaryotic chaperonins.

One of the most detailed examples where very specific chaperones mediate specific assembly reactions involves a set of chaperone proteins called the pilin-specific PapD and FimC molecular chaperones. Unlike the CCT chaperonin, this chaperoned folding/assembly process does not require ATP to function nor do these specific chaperones help fold other nonspecific protein substrates. Important clues concerning the probable molecular mechanism of these chaperones have been gleaned from very specific and elegant structure studies of the chaperone-substrate complex. These highly specific pilin assembly chaperones are both encoded by specific pili operons, and these chaperones are absolutely required to provide specific steric information, thus insuring that the major pilin proteins (PapK and FimH) attain assembly competent conformations during pili growth (60, 61). The pili chaperones use a mechanism called strand complementation and strand exchange, akin to the specific domain swapping mechanisms found in a large number of oligomers (62). Through this mechanism, the pili protein presumably attains an assembly competent state by forming a complex with the pili chaperone, thus insuring that potential nonproductive interactive

hydrophobic surfaces remain unexposed. The growing pili are hypothesized to assemble when the specific chaperone strand is exchanged or replaced by a complementary strand component of the pili assembly complex (Figure 10). To test a prediction of this mechanism, Hultgen and co-workers demonstrated that they could replace the entire specific pili assembly chaperone FimG with the specific interacting but unstructured 13 amino acid strand. In the presence of the FimH pilin protein, the FimG derived unstructured polypeptide folds to form an inserted beta strand conformation that fills the exposed groove on the pili FimH protein, subsequently allowing this pilin protein to acquire its native assembly competent structure (63).

Molecular Chaperones Involved in Localized Folding

Some of the common nonspecific chaperone proteins also provide an amphipathic binding surface that can induce the formation of localized folded structure motifs. For instance, numerous investigators have shown that the group I GroEL chaperonin interact with predominantly unstructured, small polypeptide populations, shifting the solution folding equilibrium toward defined secondary structures. These induced motif structures were initially hypothesized to be formed through amphipathic interactions between hydrophobic faces of the chaperonin and the peptide. Direct evidence for this phenomenon has come from analyzing structures of small peptide-GroEL complexes. In one such experiment, Chen and Sigler specifically selected for a strong binding peptide using phage-panning

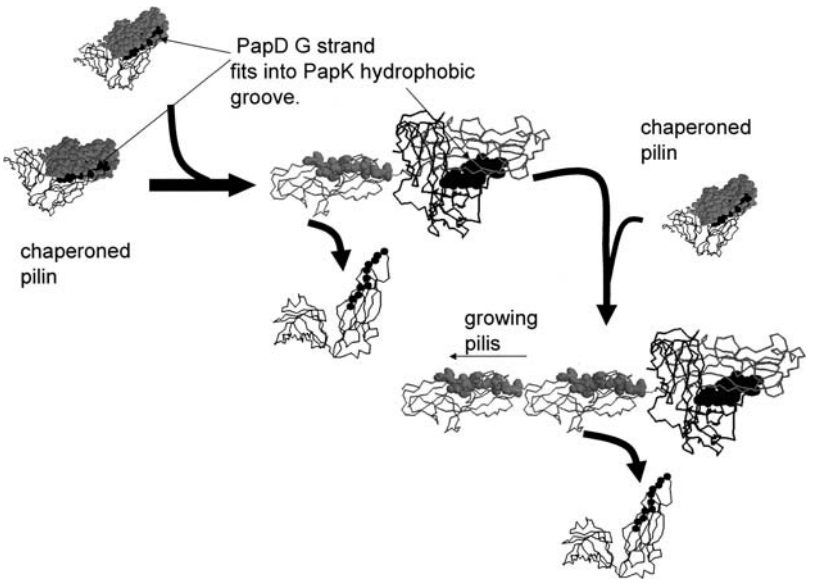


Figure 10. Example of a specific chaperone. The formation of specific domain swapped PapD-PapK complexes prevent PapK aggregation. Strand G of the PapD chaperone inserts into a hydrophobic groove of the PapK pilin protein. Chaperoned pilin protein is thought to grow the pili structure by replacing the G strand of PapD with the N-terminal portion of the growing pili. The pilin protein PapK can associate with the specific papD chaperone and the N-terminal structure of the next pilin protein. Structures were generated using RasMol version 2.6-UGB (Glaxo Wellcome Ltd.).

methods and analyzed the resolved X-ray crystal structure of the peptide-bound GroEL. The GroEL interaction with this unstructured 12 amino acid tight-binding peptide induces a structural shift to a highly populated anti-parallel hairpin. These studies indicate that the hydrophobic residues adopt a preferred orientation while binding to the chaperonin peptide binding sites, explaining the observed shifts toward higher populations of amphipathic secondary structures. Thus far, chaperonin-peptide model complexes have shown that GroEL can induce specific secondary structures such as α-helices, anti-parallel β turns and extended β-strands (64-68).

Even more surprising is the recent observation that the physiochemical nature of the internal chaperonin ring environment is hydrophobic enough apparently to stabilize and even maintain the native structure of some membrane proteins (69). For example, it has been shown that the chaperonin cavity binds hydrophobic membrane protein bacteriorhodopsin and can shuttle this protein directly between different lipid environments. More surprisingly, bacteriorhodopsin retains its chromophore while bound to the chaperonin, indicating that GroEL binds a fully functional, folded membrane protein. This GroEL captured membrane protein dissociates once ATP binds. GroEL also aids in the folding of other membrane proteins, including some containing as many as 14 transmembrane domains (69).

Other examples of induced specific interactions during folding may occur within the ribosome. It has been shown that the ribosome can provide a favorable internal environment for the nascent polypeptide chain of a membrane protein, facilitating the formation of preformed membrane helices within the ribosome channel even before the polypeptide exits the ribosome (70). The ribosome itself also exhibits some chaperone qualities (71). It is conceivable that the localized induction of protein secondary structure within the ribosome tunnel may be a common phenomenon. In this case, the ribosome-induced local folding in one environmental context may be stable enough to be transferred to another completely different solution environment (e.g., a lipid membrane).

Chaperones and Unfolding Mechanisms (Active and Passive)

There are a number of chaperone proteins that use unfolding mechanisms to help fold (or refold) substrate proteins. In this section, the mechanisms of protein unfolding by the Hsp60, Hsp70, Hsp90, and Hsp100 classes will be discussed. Two prominent mechanisms of unfolding, termed passive or active, are proposed to facilitate protein unfolding in the cell. In the passive mechanism, the chaperones will interact with a partially folded intermediate state that exists either as a prominent population or as an intermediate state that can be in rapid or slow equilibrium with its native folded state. Large-scale chaperone-dependent disassembly and unfolding can also occur during disaggregation reactions of large aggregated misfolded oligomers. The chaperone-dependent active mechanisms appear to require the binding or hydrolysis energy of ATP physically to unfold the protein substrate. As we shall see, protein unfolding reactions may precede protein folding, protein disaggregation, protein disassembly, or even protein degradation reactions.

The Hsp70 chaperones normally function by employing a passive unfolding mechanism. As mentioned previously, the Hsp70 class recognizes a peptide region that usually contains numerous stretches of hydrophobic residues that are normally sequestered on the interior of folded proteins. Thus, the transient exposure of these hydrophobic regions during folding allows the resident Hsp70 chaperones to bind and form tight complexes with the folding or unfolding protein substrate in question. In this way, the Hsp70 or Hsc70 (heat shock cognate) chaperones act as intermediate buffering chaperones wherein particular protein substrates are maintained in partially folded states. Eventually, these partially folded proteins interact more efficiently with downstream molecular chaperones such as the chaperonin 60s, Hsp90, or Hsp100 to refold, resume folding, maintain transport efficiency, or be degraded by the protease machinery. For example, *in vitro* protein folding experiments using GroEL show that the most efficient folding stoichiometries usually occur at concentration ratios of one GroEL oligomer per one folding protein. Interestingly, Hartl and co-workers found that including the complete *E. coli* Hsp70 system (DnaK/DnaJ/GrpE), enabled the substrate protein to fold from GroEL with high efficiency at substantially (10-fold) lower GroEL concentrations. These results indicate that the *E. coli* Hsp70 system prevents premature aggregation, maintaining the conformation of the folding substrate in a partially folded state that will evidentially interact with the GroE chaperonin and fold (72). Likewise, efficient protein transport from the cytoplasm to an internal organelle (such as the mitochondria) absolutely requires the interaction of the substrate with the Hsp70 class of proteins. If premature folding does occur, this Hsp70-dependent transport efficiency declines dramatically (73). In addition, as nascent polypeptide chains emerge from the ribosome, specific Hsp70 classes either in eukaryotes or prokaryotes (e.g., DnaK) are present to interact transiently with the newly formed polypeptide. Here again, preventing a rapid collapse will avoid the formation of kinetically trapped misfolded states.

In one detailed study, Hsp90 appears to interact with substrate proteins and induce a substantial amount of unfolding (21). Fersht and colleagues used cross-correlated relaxation-enhanced polarization transfer NMR spectroscopy to observe that the Hsp90-bound nuclear transcription factor p53 core domain shifts to an unfolded state when it is complexed to Hsp90. The unfolding was extensive enough to melt pre-existing helical and sheet secondary structural elements (21). The X-ray crystal structure of the dimeric carboxy-terminal protein binding region of Hsp90 has been determined to a resolution of 2.6 Å. This region has been implicated in substrate binding and it was found that this region contains a helical dimerization motif, with each monomer projecting a short flexible amphipathic helix into solution. This observation lead Agard and co-workers to propose that this region is responsible for forming complexes with protein substrates (45), perhaps leading to a general unfolding. In the near future, it will be extremely useful to examine more complexes between Hsp90 and other preferred substrates to determine whether the Hsp90 unfolding mechanism is common.

As mentioned previously, the group I chaperonins can participate in an unfolding reaction. At first glance, unfolding a protein may seem counterintuitive in a folding reaction, but in the context of the energy landscape, the unfolding of

an incorrectly folded protein may help the protein substrate avoid kinetic traps (2), allowing it to choose another energy surface to fold toward more thermodynamically stable native structures. This binding interaction between the chaperone and the misfolded protein results in a mass action shift toward more unfolded forms, thus conveying the outward appearance of a global unfolding mechanism. Indeed, the protein unfolding mechanism of GroEL is extensive enough to exchange most if not all of the initial internal peptide amide hydrogens of a number of protein substrates (74). Interestingly, the measured kinetic rate constants between folding intermediates and folded states are unchanged during partitioning, leading Frieden and Clark (75) to suggest that the unfolding mechanism is passive. In this instance, the function of the GroE chaperonins is to control the partitioning (binding) and release mechanisms through ATP binding and hydrolysis. From the energy landscape viewpoint, an alternating binding and unbinding mechanism of the chaperones, particularly for those that switch affinities through ATP binding and hydrolysis, offers a mechanism of kinetic proofreading (Figure 11) (1). In this instance, the preexisting trapped population binds to GroEL, shifting its conformation from one landscape surface to another through a passively unfolding reaction, allowing the original populated misfolded structures to refold (reset) toward a native configuration once the GroEL binding affinity is reversed by binding ATP and GroES and hydrolyzing ATP.

More direct evidence of this general unfolding mechanism comes from recent fluorescence energy transfer experiments designed by Lin and Rye (76). Here, the chaperonin GroEL initially forms a collision complex with a partially

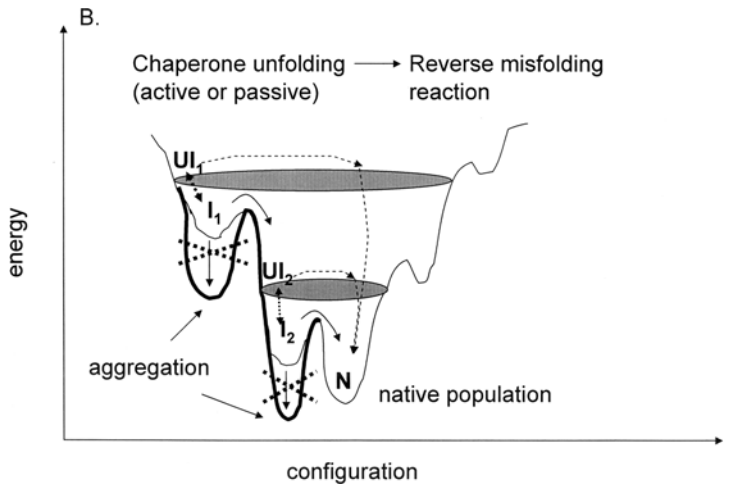


Figure 11. Unfolding influences on the folding landscape. In this instance, protein unfolding raises the apparent energy state of the trapped intermediate by unfolding it or in some cases disaggregating and unfolding the aggregate, reconfiguring the conformer to a higher energy unfolded state and allowing the protein to fold using a slightly different higher entropy pathway. The dotted lines and arrows represent alternative trajectories influenced by the various unfolding chaperones.

folded metastable state (misfolded) of ribulose bisphosphate decarboxylase (Rubisco) and the bound, Rubisco protein structure, is observed to undergo a global expansion. In this case, the Rubisco metastable state is apparently unfolded by simply partitioning onto GroEL. Once ATP and GroES are added to this complex, Rubisco refolds to a more collapsed form inside the enclosed GroES-GroEL complex, is released, and then assembles into an active Rubisco dimer. The initial step in this folding mechanism supports the passive thermodynamic partitioning mechanisms originally measured and proposed by Schmid and colleagues (77). As we shall see later in this review, GroEL may use this passive unfolding mechanism to protect natively folded proteins against heat denaturation, oxidation or aggregation (30, 32, 78).

Others suggest that GroEL may also employ an active unfolding mechanism (79). It was reported that an initial protected structural core of a metastable protein substrate (Rubisco) was exposed only after ATP was added. For active unfolding, GroEL actively or mechanically unfolds the protein substrate, presumably driven by the large conformational twists through ATP-driven movements in the apical domains that directly bind to the protein substrate (79). However, this active unfolding mechanism may not be universal because this particular unfolding mechanism has not been observed for other protein substrates. For instance, no predictable protection/deprotection shifts of internal residues were reported for the substrate malate dehydrogenase after ATP addition (80). Even so, it is still possible that the versatile GroEL chaperonin uses a couple of different unfolding mechanisms to facilitate folding.

The Hsp100 class of chaperones (ClpB or Clp A (Y) in *E. coli*; Hsp104 in yeast) utilize another interesting but somewhat mysterious mechanism of assisted folding through an unfolding mechanism. These ATP-dependent chaperone proteins directly unfold misfolded proteins or directly break up large insoluble aggregates (Figure 7). In *E. coli*, these Hsp100s exist as ring structures that contain a central pore surrounded by conserved residues, and they appear to unfold and thread proteins through this pore from either their amino- or carboxy-termini (81).

Clues into the molecular details of the unfolding mechanisms of the Hsp100s have relied on examining the structure and mechanisms surrounding the ClpA-ClpP interactions (82). ClpP is a self-compartmentalized protease that contains peptidase active sites sequestered inside a hollow oligomeric protein core (83). In order for proteolysis to occur within this sequestered active site, the substrate protein has to be presented to the interior of the protease in an unfolded or nearly unfolded state. In this regard, the roles of the Hsp100 like ClpA or ClpY hexamers in the ClpA(Y)-ClpP complexes are to unfold and thread the extended polypeptide chain into the interior of the enclosed ClpP protease. The *E. coli* disaggregation chaperone ClpB, like ClpA, is also thought to unfold proteins through a similar threading mechanism. If the interior channel is crucial to the unfoldase mechanism, then it is logical to assume that mutations at the interior conserved amino acid residues will affect function. Indeed, when targeted mutagenesis experiments were used to change conserved residues located within the ClpB hexameric central axial pore region, it was found that the homologous mutations tyr→phe preserved the unfoldase and thermotolerance functions

whereas decreased interaction tyr→ala mutations became defective in both functions (84). Further support for the threading mechanism was provided by Bukau and colleagues. In a series of elegant experiments, they constructed chimeras of the ClpB oligomer where the ClpP interacting region of Clp A was engineered onto Clp B. As predicted, this engineered ClpB-ClpA chimer was now able to interact with ClpP and function as an unfolding machine for this protease (Bukau, preliminary results, 2004). Interestingly, they also found that this particular construct was detrimental to cell growth in a ClpB depleted *E. coli* host background under heat stress conditions, suggesting that the general unfolding-resolublization-refolding mechanism of ClpB was crucial for cell survival.

In other studies, Lindquist and colleagues found that the yeast Hsp104 chaperone functions as an unfolding chaperone within certain concentration ranges. This group has found that yeast controls the aggregation of a translation termination protein called Sup35, apparently by varying the intracellular levels of Hsp104. At both high and low concentrations, Hsp104 facilitated Sup35 aggregation either by catalyzing the formation of an aggregation prone Sup35 form or by exposing more Sup35 hydrophobic surfaces on existing fibers to seed more aggregation. At intermediate cellular concentrations, Hsp104 apparently disaggregates preformed fibrils and, unlike the Clp proteins, it apparently performs this reaction without requiring the Hsp70/40 system. The exact molecular mechanism is not known but it is intriguing that, at least for Hsp104, one may be able to physically disrupt large amyloid-like fibrils. The key observation here is that the Hsp104 protein appears to directly interact with the preformed fibril (Figure 7), suggesting that Hsp104 may operate by a different mechanism (perhaps the crowbar mechanism of fibril disruption) (85). However, others present data suggesting that Hsp104 is unable to operate alone but only disaggregates fibers with the help of some as yet unidentified soluble yeast protein factors (perhaps the Hsp70/40 system) (86). In summary, the Hsp100 class of chaperones functions as a general ATP-driven unfoldase machine, enabling the organism to refold heat denatured proteins from large inactive protein aggregates.

Sequestered Folding—A New Paradigm for *In Vivo* Folding

One of the most interesting and unique mechanisms of chaperone assisted folding is used by the chaperonin class of molecular chaperones. *In vitro,* it has been demonstrated that the active chaperonin complex (both group I and perhaps II) forms a transient nano-structured protein chamber where a trapped protein substrate can fold in a sequestered environment, thus avoiding general protein aggregation. There are numerous instances where small proteins are captured, retained inside this unique protein environment, and proceed toward their active folded states (87, 88). Indeed, theoretical folding simulations show that protein folding within a confined space is favorable because such a physical constraint leads to the marked decrease in the entropic cost of collapsing a denatured state to its more compact folded structure. Thus, confining and limiting the conformational search space that a denatured protein substrate can explore may allow it to retain more native-like interactions, potentially enhancing folding rates and even leading to a predicted increase in the mean folding temperature (89, 90).

Conceptually, confining conformational search space can dramatically change the surface of the folding funnel, smoothing the energy landscape while avoiding deleterious low-energy kinetic traps (Figure 12).

Although this mechanism is very plausible, in reality these theoretical studies do not take into account the possibility that some of the physical properties of the nano-structured cavity walls may inhibit folding. In particular, these walls are probably not inert to all folding substrates because some strong hydrophilic interactions between the folding protein and the protein walls may occur. For some proteins, particularly those that are not intrinsic *E. coli* protein substrates, folding may be slow, progressing through a series of multiple rounds of rebinding, unfolding or folding, and sequestration until they acquire a folded state that no longer binds to the chaperonin (91, 92). Curiously, however, recent data collected by the Hartl group indicate that *E. coli* proteins that have been identified to absolutely require the complete *E. coli* GroE chaperonin system to fold do not have to undergo a multiple cycling mechanism (Hartl, preliminary results, 2004). Apparently, the authentic substrates may only require a brief interaction with the chaperonin. This observation makes evolutionary sense because the folding efficiencies of *E. coli* protein substrates, ultimately dictated by the protein primary sequence, have co-evolved with the *E. coli* chaperonin to avoid deleterious long-term interactions (i.e., multiple cycling) with the GroE chaperonin system. Indeed, it has been frequently noted that expressing slower folding extrinsic protein substrates inside *E. coli* sometimes can be detrimental to cell growth and/or cell division (93). In addition to the problems that may be encountered for

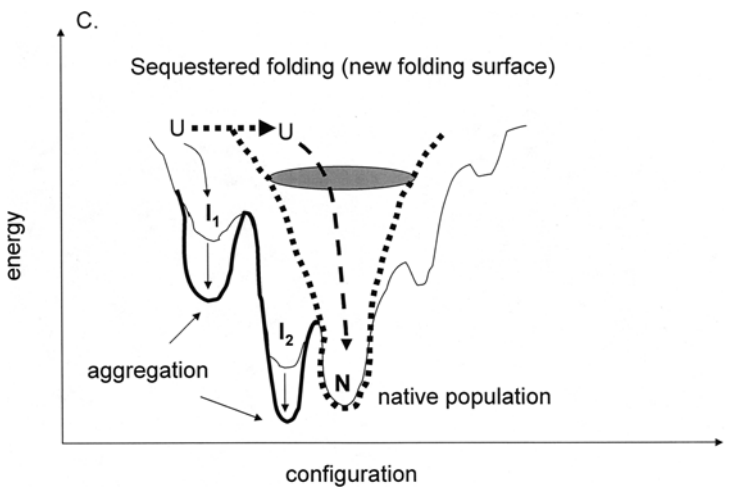


Figure 12. Sequestered folding dramatically alters the folding landscape (dotted surface). In this case, by constraining the entropic freedom of movement of the folding protein, a new energy surface may be formed where most of the kinetic folding traps are eliminated. This energy surface would also be applicable to very specific steric folding mechanisms. However, even though the energy minimum in this landscape is the properly folded protein, this does not preclude the possibility that sequestered folding could force the protein into a stable misfolded population.

overexpressed proteins, the sequestering mechanism cannot explain why the GroE chaperonins are unable to fold a number of small sized proteins such as actin or tubulin even though GroE interacts with these protein substrates. Here again, simply sequestering a folding protein inside the cavity does not automatically guarantee that the protein substrate will acquire its correct conformation. Even in situations where substrate protein interaction with the chamber walls may be weak, the non–*E. coli* substrate protein may simply fold into a trapped yet stable misfolded state.

Although most mechanistic diagrams and textbook descriptions of the chaperonin function still emphasize the sequestered folding mechanism of the group I chaperonin, it is now recognized that protein folding from the chaperonin does not always proceed through the complete encapsulation mechanism. For proteins that are too large to fit inside the defined GroEL-GroES nano-cavity, folding occurs *via* a negative cooperative mechanism (called *trans* folding) where the binding of GroES on the opposite ring from the large substrate bound ring is sufficient to facilitate the substrate protein release and folding. Large *E. coli* proteins (e.g., large subunits of RNA polymerase) have been identified to be authentic substrates for the GroEL chaperonin (32). Horwich and colleagues have now unequivocally demonstrated that a *trans* folding mechanism efficiently folds large proteins like aconitase (70 kDa) (40, 41). Thus, GroE-dependent folding does not always have to proceed *via* a strict encapsulation mechanism. In fact, Horwich's group engineered the covalent attachment of a single-ring GroES to double-ring GroEL and showed that even smaller protein substrates can fold from the opposite (*trans*) GroEL ring, albeit at a slower rate. Although the exact mechanism of folding from the unenclosed ring is unclear, it appears that in some cases, the previously described localized folding or unfolding mechanisms may be enough, in this case, to facilitate correct folding. In conclusion, it is safe to say that the group I chaperonin machine uses a number of diverse folding mechanisms to fold protein substrates.

Chaperones as "General Buffers" for Protein Folding

Most protein substrates that require chaperones to fold are identified by either immunoprecipitation (IPs) experiments (32) or chaperone knockout or depletion experiments (94). It is now apparent that some protein substrates transiently and weakly interact with the common chaperones, yet still require chaperones to fold and assemble properly. In this instance, some chaperone proteins (Hsp60, Hsp70s) may serve like a protein buffer to maintain active states of select metastable proteins or protein intermediates that are in equilibrium with their native conformations (29, 74). *In vitro*, it was found that an active cycling chaperonin system (GroE + ATP) was necessary to maintain rhodanese in its active conformation even though this enzyme was undergoing rapid inactivation/reactivation (oxidation/reduction) cycles (30). At physiological temperatures, numerous proteins lose their activity over time and it has been repeatedly demonstrated that a cycling chaperonin system will prevent this slow loss in activity (95). Likewise, the Hsp70/40 system also appears to be able to prevent heat denaturation and maintain proteins in their active states (96). In both of these examples,

the specific molecular mechanism may involve repeated passive partial unfolding/refolding transitions (30).

Transient interactions are also important in situations where metastable states of proteins form, particularly from proteins that exist in dynamic yet active oligomeric states. In many cases, these metastable states may dissociate and fold into kinetically trapped and improperly folded subunits that are unable to assemble or even aggregate. In a number of documented instances, brief interaction of these metastable forms with chaperone proteins transform these states back into their assembly competent forms (76, 97-99).

Because chaperone proteins often aid in shifting misfolded states toward active folded ones, chaperone proteins also work as internal homeostatic protein folding buffer systems that, for example, can fold mutant proteins to their active folded states, leading to an increased survivability of organisms containing potentially deleterious folding mutants. When present at high levels inside the cell, the chaperones can, in effect, hide the folding defects of these mutants by enabling these defective proteins to fold and maintain activity, resulting in an apparent native phenotype (100). Interestingly, since these metastable mutant phenotypes depend on the constant presence of the chaperone proteins, episodes of environmental stress resulting in transient increases in populations of unfolded or misfolded proteins flood the chaperone system and out-compete the mutant proteins for the normally protective resident chaperones, leading to their eventual misfolding or altered function. Rutherford and Lindquist (101) have proposed a potentially exciting chaperone-dependent molecular mechanism of evolution based on their ground-breaking experiments documenting the effects of heat stress and/or targeted Hsp90 inhibition on genetic development.

Chaperones are the interface between genotype and phenotype because they can broaden the genetic variation within a population. In this scenario, the chaperones buffer folding mutations or functional metastable proteins that are involved in regulating signal transduction and gene transcription, allowing these proteins to function. However, because they require interactions with chaperones (particularly Hsp90) to fold or function, they are particularly sensitive to environmental stress, which can result in marked changes in gene regulation and gene expression for the fringe populations. The most striking element about this stress perturbation is that it may play a crucial role in molecular evolution, particularly when changes in the gene regulation processes affect development. What is most interesting and evolutionarily relevant about these changes in the development program is that they can become inheritable stable defects. In this way, select populations of an organism can rapidly respond to the environment and even mimic rapid evolutionary selection processes in sort of a molecular version of punctuated evolution, initially triggered by modest environmental stress events (for reviews, see 102, 103).

Although the small heat shock proteins (sHsp) are not directly involved in folding, they also serve as protein buffers to specifically prevent or decrease protein aggregation. The sHsps crystallins are present in the eye lens to prevent large-scale aggregation, thereby preventing deleterious increases in light scattering within the lens. Although the oligomeric state of the sHsp protein is very dynamic, once the substrate becomes bound to the sHsp oligomer, it does not readily

exchange. For this reason, subsequent reactivation of the protein substrate from this bound state requires additional help from other resident ATP dependent chaperones (Figure 8) (48, 104).

In their dissociated states, the smaller suboligomeric sHsp form complexes with stress-denatured proteins by binding and sequestering or blocking potential aggregation prone sites on folding intermediates (Figure 8). Using a set of thermodynamically and structurally defined stable lysozyme mutants as their model substrate system, Koteiche and McHaourab (105) examined the binding interactions between various folding intermediates and the sHsp proteins and used these results to propose a plausible mechanism for determining how sHsp chaperones select their protein substrates. They found that the sHsp chaperone has two different binding modes (high and low capacities) for substrate proteins that are structurally regulated by temperature. The appearance of the high-capacity mode correlates with an increased unfolding state of the substrate protein and heat-dissociated sHsp subunits at heat shock temperatures whereas the lower-capacity mode physically correlates with an interaction between the smaller sHsp oligomer (Figure 8) and partially folded conformers that exist at normal temperatures. The advantage of controlling affinities and binding modes of sHsp with temperature allows the organism to avoid large scale unfolding and kinetic partitioning of normal substrates onto the sHsp at physiological temperatures.

Chaperone Networks

Although most of the mechanisms discussed thus far have focused on specific molecular mechanisms of individual classes of chaperone proteins, chaperones operate side by side as networks *in vivo*. As noted previously, early *in vitro* experiments with purified chaperones illustrated the advantages that chaperone networks bring to protein folding (72). Chaperone networks not only contribute to protein folding, they also play important roles in unfolding or assembly of heterologous protein complexes. For instance, it is now well established that the various steroid nuclear receptor proteins use very elaborate and stable multiple chaperone complexes involving Hsp90/Hsp70/Hsp40 and foldases such as peptidyl prolyl *cis-trans* isomerases to regulate function (for a review, see 106).

As will be discussed in a later section, general data from biological systems indicate that there are multiple isoforms and paralogs of the same chaperone protein families. In terms of chaperone networks, there are numerous examples where the loss of function of one particular chaperone species is compensated by another resident chaperone. For example, depletion of trigger factor (*cis-trans* peptidyl-prolyl isomerase) and Hsp70 (DnaK) can be tolerated in *E. coli* because the SecB chaperone, a promiscuous binding transport chaperone protein, can fill in as a general chaperone in addition to its normal role facilitating protein transport (107). As well, it has been demonstrated that some substrate proteins can fold with the Hsp70 class to the same extent as is observed with the chaperonin class. In these various examples, a considerable redundancy exists where chaperone functions are shared by different chaperone families. Other protein substrates, such as some of the WD-repeat proteins that require the CCT chaperonin to fold, also can interact with the isoform Hsp70 Ssb, when its nor-

mal chaperone co-factor prefoldin is depleted. In this instance, removing both the prefoldin and Ssb chaperones is lethal, suggesting that they may play indirect overlapping roles in protein folding (108).

From a functional standpoint, multiple chaperone networks are also intimately involved in most if not all protein trafficking processes. For example, transported proteins first interact with numerous cytosolic chaperones (i.e., Hsp70 system) to acquire and maintain their transport-competent, partially folded forms prior to their translocation into mitochondria. Once transport is complete, the newly transported substrate encounters the resident mitochondrial heat shock chaperones, Hsp60 and the Hsp70/40 system, once again resuming folding and assembly. It is important to note that during this process, the transported proteins are proteolytically modified (cleavage on N-terminal leader sequences), leading to a change in their folding kinetics that allows for the formation of a stable collapsed structure (109). As mentioned previously, chaperone networks also exist to facilitate protein disaggregation, refolding, or proteolysis. For example, the Hsp70/40 systems are necessary components of the Hsp100-dependent protein where they complement the disaggregation reaction and/or participate in the eventual rescue and refolding (Figure 7).

Role of Chaperones in Protein Assembly/Disassembly

There are numerous examples where highly specific chaperones directly aid in the assembly of proteins. Unlike the more common chaperone systems discussed in this review, these specific chaperones operate by sterically determining the assembly competent states. As discussed earlier, one of the best studied specific assembly systems involves the assembly of the *E. coli* PapK pili by the PapD chaperone. Other specific assembly systems include the specific phage-encoded chaperones called the scaffolding protein systems. In these latter systems, the scaffolding proteins are intimately involved in insuring that viral proheads assemble correctly. The scaffold proteins function as chaperones because they are not part of the final viral structure. These proteins are quite literally interior scaffolds that the prohead assembled viral coats interact with and, through this association, the viral coat assembly changes conformation, allowing it to progress toward the mature capsid conformation that is now able to accommodate and package viral DNA (110). These steric chaperone proteins are highly specific chaperones that can act at the monomer or oligomer level. These mechanisms are perhaps very similar in principle to examples where specific protein binding induces folding toward a new more active structure, insuring that extremely long-lived kinetic conformers (e.g., the folding of the alpha-lytic protease) can escape stable kinetic traps (111, 112).

In addition to these specific chaperone assembly systems, the common chaperones also participate indirectly in protein assembly reactions. For instance, the group I chaperonins facilitate assembly by interacting with partially folded monomers of oligomeric proteins, and insuring that they fold (or unfold) into assembly-competent states. GroE only induces the formation of the assembly-competent monomeric species through a transient mechanism but it doesn't participate directly in assembly (96, 113, 114). Even though chaperonins do not

participate in the actual assembly process, they can accelerate the association reaction by increasing the concentration of assembly-competent subunits. This has been shown *in vitro* where chaperonin-dependent oligomer formation can occur at concentrations that far exceed critical aggregation concentration limits (115-117).

Subunits of oligomeric proteins can attain assembly competence through either partial folding or unfolding mechanisms. As was mentioned before, Lin and Rye (76) observed the ribulose bisphosphate decarboxylase had to undergo a general unfolding reaction of a metastable state before it could once again collapse toward an assembly-competent species. In another assembly system, it was demonstrated that metastable GS monomers formed under low, albeit physiologically relevant, Mg^{2+} concentrations could properly assemble once they interacted with the complete GroE chaperonin system. In contrast to a general unfolding mechanism, however, it appears that the metastable GS monomers have to undergo a collapse because correct assembly is only observed following a tertiary structure collapse and the formation of a more compact structure (Fisher, preliminary results; 98). Because increased Mg^{2+}-binding (higher Mg^{2+}) or osmolytes can induce a folding collapse and facilitate assembly (98, 118), in this case, the chaperonin may collapse rather than unfold the metastable GS monomer. Here again, the versatility of the chaperonin system to unfold, induce local folding, or induce constrained collapse (sequestered folding) can lead to a variety of different mechanisms governing protein assembly.

Interestingly, chaperonins may possess a kinetic specificity for some assembly processes. For example, *E. coli* glutamine synthetase monomers commit to an assembly-competent form (no longer require rebinding to the *E. coli* chaperonin GroE) at a rate that is substantially faster than the actual renaturation rate (93). For folding of non-*E. coli* substrates such as mitochondrial ornithine decarboxylase or mitochondrial malate dehydrogenase with the *E. coli* chaperonin, the measured commitment rates are slow and mirror the activity regain rate, reflecting multiple binding and release cycles (88, 91, 95, 116). Since these substrates are not the authentic substrates for the *E. coli* chaperonin and because rapid *in vivo* assembly is probably subject to evolutionary pressure, transient interactions will be favored due to the limited number of intracellular chaperonins (28). For instance, successful microorganisms competing in natural environments certainly are sometimes best served by the rapid acquisition and utilization of limited food resources. It is logical to conclude that any advantage for producing rapidly assembling metabolic proteins will carry a selective advantage over those organisms that assemble their proteins at slower rates.

Although the chaperonins are not typically classified as specific steric chaperones, the group I cochaperonin GroES may sterically assist in the assembly of the GroEL chaperonin. Horowitz and colleagues noted that the *in vitro* assembly of the chaperonin GroEL appears to be accelerated when its cochaperonin GroES is present (119). The operon of the GroE chaperonins is arranged so that the smaller cochaperonin GroES monomers are synthesized first, followed by GroEL. This particular organization results in the formation of a functional heptamer of GroES before the tetradecameric GroEL is formed. Horowitz and colleagues suggest that the newly formed GroES potentially provides a structur-

al template upon which GroEL can form. However, this potentially interesting steric assembly reaction still has to be verified to occur *in vivo*.

Like the chaperonin, the Hsp70 classes also participate indirectly in assembly reactions. There have been numerous *in vitro* reports where the Hsp70 proteins also function to maintain assembly-competent states prior to assembly. Here again, the role of the chaperone protein is to extend the lifetime of the assembly-competent state (120).

Finally, one of the most critical assembly processes that involves chaperones is the assembly/disassembly control and regulation of transcription complexes. In particular, the Hsp70/40/90 systems are intimately involved in maintaining or controlling the assembly competence for a variety of transcription factors (121, 122). In addition, intracellular steroid receptors involved in transcriptional activation form large complexes with various chaperones [Hsp90, p23, and immunophilins (PPIs)], which evidently serve to hold and block the binding site of the steroid receptor. In this situation, the complex is locked in an incompletely folded inert state that becomes disassembled when the steroid hormone binds. Although much of our current understanding of the Hsp90-regulated complexes is still at the "oval biochemistry" stage, understanding the molecular details behind the formation and control of these complexes will prove to be extremely interesting and informative. Determining the exact molecular mechanisms and the dynamics of the chaperone-dependent assembly/disassembly control will certainly lead to exciting new developments in many fields.

FUTURE DIRECTIONS FOR CHAPERONE RESEARCH

The Specificity of Chaperone Proteins

Even though a majority of the chaperones interact with protein substrates in a nonspecific manner, biological data suggest that chaperones must exhibit some loosely defined substrate specificity that is governed by the properties of the protein substrate, the intracellular environmental differences, or the differences in the internal molecular dynamics of the chaperone itself (i.e., speed of conformation changes, binding and release cycle, etc.). With regard to the substrate protein properties, it has been firmly established that the physical properties of the substrate protein can influence its eventual partitioning and folding onto chaperone protein. There are numerous examples where changes in the folding landscape of a protein (e.g., through a single amino acid change) influences its folding rate or physical properties leading to enhanced interactions with the resident chaperone proteins. To understand this underlying specificity, it will be helpful to determine and compare the molecular details that lead to successful and unsuccessful chaperone-assisted folding for specific chaperone systems. Understanding this underlying specificity will help us determine how chaperones select specific substrate molecules out of the complex intracellular milieu.

Initial evidence of chaperone specificity comes from comparative studies between orthologs and paralogs of chaperone proteins within or between organisms. This loose substrate specificity for chaperone-assisted folding, assembly, or degradation is probably the reason why so many different isoforms or paralogs of

the same chaperone classes co-exist within the same organism (e.g., Hsp70, Hsp40 isoforms, Hsp60 isoforms, etc.). For example, it was observed that some protein substrates can fold from one particular chaperone yet are unable to fold from their homologous orthologs or paralogs even though the protein substrate in question can interact with all the chaperone isoforms. Some organisms contain and express multiple paralogs of the "promiscuous" chaperonin family within the same intracellular environment. Numerous microorganisms, including some that are pathological, simultaneously synthesize multiple chaperonin species (both GroEL and GroES-like species) (123, 124). Furthermore, some organisms appear to express both forms of the chaperonin family (group I and group II) simultaneously. Hartl and co-workers find that a thermosome-like chaperonin (group II) and a eubacterial-like chaperonin group I protein coexist within the same cell (125). Studies focused on identifying the protein substrates for these two specific chaperonin groups will be particularly illuminating for determining elements of chaperonin substrate specificity. Variable chaperonin function has been observed *in vivo* in instances where specific phage protein/host chaperonin relationships have co-evolved (126) and in cases where folding efficiencies and complementation differences are observed for different chaperonin paralogs (127).

This diversity in substrate recognition is also observed for the Hsp70 chaperones. Numerous Hsp70 isoforms coexist within their bacterial, yeast, or other general eukaryotes (23, 25, 128). These specific Hsp70 systems may facilitate folding of specific protein classes. For example, in *E. coli*, various iron sulfur proteins appear to require a specific set of constitutive heat shock cognate Hsp70-like chaperones, the Hsc66/Hsc20 class to fold (129). These proteins are very similar to the major *E. coli* DnaK/J-Hsp70/hsp40 chaperone system. In addition to the multiple classes of Hsp70-like proteins in the cytoplasm, the mitochondria also possess multiple Hsp70-like proteins that may play similar specific roles in facilitating mitochondrial protein folding and assembly.

This underlying chaperone specificity is also readily apparent when one compares the folding efficiency of the same protein using the chaperone systems between different organisms. For example, it has been shown that bacterial multidomain proteins such as beta-galactosidase use the *E.coli* Hsp70 trigger factor chaperone mechanism to fold effectively (more rapidly), while, conversely, these *E. coli* chaperones slow the folding of the eukaryotic multidomain protein like luciferase. Apparently, the efficient co-translational domain folding of luciferase observed in the eukaryotic system may not be as compatible with the bacterial chaperone system (130). In another study, a depletion of the prokaryotic Hsp70 (DnaK) in *E. coli* cannot be complemented by the expression of a very similar Archaea prokaryotic Hsp70 chaperone, even though *in vitro* experiments do show that some rescue of protein folding occurs (131). These findings suggest that differences in the eukaryotic and prokaryotic chaperones will differentially influence co- and post-translational protein folding. Differences in folding efficiencies (rates or yields or both) that depend on the identity of the chaperones used suggest that there exists a level of chaperone-substrate specificity.

Specificity may also depend on the dynamics of the chaperone machinery. For example, the kinetics of the Hsp60 chaperonin release of the folding protein from its nano-structured folding chamber has an impact on successful folding.

Buchner and colleagues (132) showed that extending the lifetime of a folding protein inside the internal nano-structured chamber is detrimental for folding, particularly for oligomeric proteins. In addition, the molecular dynamics that define the chaperonin cycle (binding-encapsulation-release) may also be of critical importance. Specific allosteric mutants developed by Horovitz and colleagues indicate that slowing the allosteric transitions and communications between the rings result in a marked decrease in the success of folding abilities of proteins that transiently interact with the chaperonin (133). Interestingly, slowing the conformational transitions of GroEL does not seem to affect slower folding proteins, but, as mentioned previously, the faster folding transitions in *E. coli* may be the dominating physiological interaction.

At the molecular level, there is both indirect and direct evidence supporting the notion that substrate protein binding also influences chaperone structure. The perturbing effect of substrates on chaperone dynamics and hence structure was first indirectly inferred by Horwich and co-workers when they compared the ATP-GroEL-GroES structure with the ADP-GroEL-GroES structure. Even though one structure (the ATP-bound form) is supposed to be the folding competent state of the chaperonin, no differences were evident between the structures (42). Horwich and colleagues have now shown that a tight binding substrate, such as MDH, imparts a significant load on the apical domain, particularly in the ADP-bound form of GroEL (134). Thus, the binding interaction between the substrate and GroEL do control the dynamics and allosteric structural changes of the chaperonin system. The biological diversity, coupled with the preliminary evidence that the substrate proteins directly affect the chaperonin (35, 36, 134), indicate that there is a critical gap in our understanding of the role of the most important but least studied ligand, namely the substrate protein, in the chaperonin cycle.

The protein substrates of the GroEL and DnaK chaperones as well as the chaperones themselves are able to structurally "mold" their binding site interactions, manifesting the promiscuous binding properties of these chaperones. To this end, more direct measurements of structures of protein substrate-GroEL complexes obtained from single-particle cryo-electron microscopy indicate that substrate protein binding interactions with GroEL result in a change in the entire GroEL structure. Molecular dynamics fit to an authentic protein substrate-bound GroEL structure provide a reasonable mechanism to explain the substrate protein–induced negative cooperative allostery for this system (135). The prediction that substrates may mold both themselves and the active site of a chaperone are born out from recent peptide-binding experiments. Most interestingly, protein substrate-induced conformational changes in GroEL appear to depend on the nature of the interacting peptide, strongly suggesting that these changes will be substrate dependent (136; A. Horwich, preliminary results).

Chaperones as *In Vivo* and *In Vitro* Folding Tools

Since chaperone proteins are required for proper folding and assembly *in vivo*, there has been significant interest in using these special proteins as potential folding tools *in vivo* and *in vitro*. *In vivo*, problem folding due to protein misfolding

and/or aggregation usually occurs during the overexpression of eukaryotic recombinant proteins within different host systems. It has become evident to some that the *in vivo* co-expression of chaperone proteins either individually or as mini-networks (Hsp70/40, Hsp60/10) can lead to substantially improved folding and assembly (for a review, see 137). Cellular conditions can be adjusted to augment the effects of expressed chaperone proteins. In a number of instances, proteins that are expressed with chaperones at lower temperatures can be produced in adequate soluble and properly folded quantities to enable researchers to crystallize their favorite overexpressed protein (138). In other instances, researchers have made genetic constructs that contain the genes encoding the particular chaperone system (either ER chaperones or cytoplasmic chaperones) and the target protein on the same construct, ensuring that, during their co-expression, there are adequate concentrations of chaperone proteins present to facilitate *in vivo* folding. This particular system has been moderately successful at folding difficult proteins such as membrane proteins.

In vitro methods that employ chaperone proteins have been limited in their use because their specificity in folding is not understood. However, this has not deterred investigators from constructing *in vitro* chaperone folding systems. Fersht and co-workers designed a bead-immobilized multiple chaperone system containing peptidyl prolyl *cis-trans* isomerase, protein disulfide isomerase, and the polypeptide binding fragment of GroEL (139). The immobilized chaperone beads could facilitate folding of a scorpion toxin isolated from inclusion bodies (cellular aggregates). Unfortunately, this novel approach does not work for a number of other proteins that have to fold with the complete chaperonin system (rather than the chaperonin fragment), thus limiting the scope of its application (140, 141). However, the notion of using chaperone proteins for general *in vitro* folding aids is still a viable pursuit.

Recently, another *in vitro* chaperone-assisted folding method has been developed that uses a combination of two physiological folding aids, namely the GroEL chaperonin, ATP, and a series of naturally occurring cellular osmolytes. The combination of these two folding aids results in a remarkable synergy for folding success *in vitro*. Specifically, this combinatorial approach has enabled researchers to fold proteins that could not fold with either of the two folding aids alone (99, 117, 118, 142, 143). This technique not only provides a potentially new screening method, it also simplifies the chaperonin mechanism. For example, proteins that fold only with the GroEL (large) and GroES (small) chaperonin components now can fold with GroEL and ATP or ADP (143). The mechanism of action depends on the osmolyte effects on the protein folding intermediate when it is bound to the chaperonin. The addition of cellular osmolytes to folding proteins favors the burial of peptide backbone (144) and initiates a rapid collapse. The chaperonin osmolyte system relies on a simple two-step procedure. The chaperonin/osmolyte folding process starts with the formation of a stable long-lived chaperonin-substrate protein complex. When osmolytes (e.g., glycerol, trimethyl amine N oxide (TMAO), sucrose, L-proline) are added to this preformed complex along with nucleotide, the osmolyte-induced collapse or forced folding of the bound target protein leads to a decrease in the intrinsic binding affinity between the folding intermediate and the GroEL chaperonin. This observation probably

explains why ADP binding to GroEL in the presence of osmolytes is sufficient to dissociate and initiate folding of numerous substrate proteins (143). Physiologically, the combination of high intracellular osmolyte concentrations and chaperone proteins is likely to have a dramatic impact on *in vivo* folding (145). In an applied sense, this combined chaperonin/osmolyte process has been demonstrated to fold proteins successfully from purified inclusion bodies at high concentrations (115, 117, 143) and, more importantly, aids in reversing misfolded protein states that have been found in some human disease states (99).

Chaperones and "Conformational Disease" Therapies

Chaperone proteins are intimately involved in inhibiting or influencing disease progression, particularly in diseases that result in an accumulation of misfolded proteins or aggregated states. Chaperones are also often associated with amyloid aggregates although their interactions are sometimes thought to occur during fibril growth (146). *In vivo*, chaperones have been implicated in regulating the progression of the aggregation resulting from extensions of repetitive polyglutamine tracts (Huntington's disease and spinocerebellar ataxias). In a series of elegant experiments, Plasterk, Morimoto, and colleagues used genome-wide RNA interference assays to identify proteins whose depletion resulted in an early onset of a modeled polyglutamine repeat aggregation disease in *Caenorhabditis elegans* (147). The identified targeted proteins fell into six internal protein homeostasis processes such as RNA metabolism, protein synthesis, protein folding (molecular chaperones), protein trafficking, protein degradation, and energy utilization. Interestingly, defects in the protein-folding chaperones were limited to a small group of specific chaperone proteins whose depletion was enough to induce an early progression of polyglutamine aggregation within this *in vivo* model. For example, six subunits of the eight total of the group II chaperonin (CCT), two Hsp70 proteins, and an N-terminal Hsp J domain were identified to be critical at protecting against early onset aggregation, once again emphasizing underlying chaperone specificity in the protein folding process. These results may be used to design therapies aimed at enhancing the levels of specific chaperones to delay or blunt specific disease progression.

One tantalizing therapeutic treatment aimed at controlling protein folding and trafficking flux through chaperone networks involves the use of a small molecule therapeutic called 4-phenylbutyrate, an approved general ammonia scavenger for urea cycle disorders and a known transcriptional regulator. This compound appears to down-regulate the Hsc70 protein levels and up-regulate Hsp70 by an undetermined mechanism. This control switch adjusts the flux between degradation and folding. Hsc70 is implicated in protein degradation of partially folded protein while Hsp70 is involved primarily in protein folding. Apparently, a target protein that is selected for degradation initially forms a stable complex with Hsc70, which then interacts with the chaperone-dependent E3 ubiquitin ligase called CHIP (*c*arboxyl terminus of *H*sc70-*i*nteracting *p*rotein), resulting in the ubiquitination of the target protein substrates and Hsc-70 (148, 149). Once ubiquitinated, the protein substrate is subsequently degraded by the proteasome. A down-regulation of Hsc70 and the simultaneous up-regulation of

Hsp70, either by 4PBA (4-phenylbutyric acid) or by genetic means (150), will increase the flux of the target protein toward its folded and functional state. Shunting protein flux from degradation and to folding is a useful procedure to correct trafficking defects encountered in protein conformational diseases found in cystic fibrosis transmembrane regulator (CFTR) folding and the folding of secretion competent Z variant of the elastase antitrypsin inhibitor (151, 152).

The information from molecular approaches that define the binding energetics, kinetics, and dynamics of chaperone-substrate protein functions will most certainly aid in our understanding and application of chaperone therapies for the growing number of protein folding diseases. Much work still needs to be done in the chaperone field, particularly in the area of identifying, characterizing, and understanding the molecular basis for the inherent chaperone specificity in folding and assembly. The most exciting developments in this field will come from understanding the molecular details of chaperone function in genetic transcription and molecular evolution.

REFERENCES

1 Gulukota, K. and Wolynes, P.G. (1994) Proc. Nat. Acad. Sci. U.S.A. 91, 9292-9296.
2 Dill, K.A. and Chan, H.S. (1997) Nat. Struct. Biol. 4, 10-19.
3 Sabelko, J., Ervin, J. and Gruebele, M. (1999) Proc. Nat. Acad. Sci. U.S.A. 96, 6031-6036.
4 Goldbeck, R.A., Thomas, Y.G., Chen, E., Esquerra, R.M. and Kliger, D.S. (1999). Proc. Nat. Acad. Sci. U.S.A. 96, 2782-2787.
5 Rhoades, E., Gussakovsky, E. and Haran, G. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 3197-3202.
6 Mitraki, A., and King, J. (1992) FEBS Lett. 307, 20-25.
7 Sturtevant, J.M., Yu, M.H., Haase-Pettingell, C. and King, J. (1989) J. Biol. Chem. 264,10693-10698.
8 Qu, B.H. and Thomas, P.J. (1996) J. Biol. Chem. 271, 7261-7264.
9 Wilkinson, B. and Gilbert, H.F. (2004) Biochim. Biophys. Acta 1699, 35-44.
10 Schiene, C. and Fischer, G. (2000) Curr. Opin. Struct. Biol. 10, 40-45.
11 Ellgaard, L. and Helenius, A. (2001) Curr. Opin. Cell Biol. 13, 431-437.
12 Epstein, C.J., Goldberger, R.F. and Anfinsen, C.B. (1963) Cold Spring Harbor Symp. Quant. Biol. 261, 12414-12417.
13 Georgopoulos, C.P., Hendrix, R.W., Casjens, S.R. and Kaiser, A.D. (1973) J. Mol. Biol. 76, 45-53.
14 Pelham, H.R. (1986) Cell 46, 959-961.
15 Ellis, R.J. and Hemmingsen, S.M. (1989) Trends Biochem. Sci. 14, 339-342.
16 Goloubinoff, P., Christeller, J.T., Gatenby, A.A. and Lorimer, G.H. (1989) Nature 342, 884-889.
17 Hartl, F.U., Pfanner, N, Nicholson, D.W. and Neupert, W. (1989) Biochim. Biophys. Acta 988, 1-45.
18 Cheng, M.Y., Hartl, F.U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L. and Horwich, A.L. (1989) Nature 337, 620-625.

19 Lindquist, S. and Craig, E.A. (1988) Ann. Rev. Genet. 22, 631-677.
20 Flynn, G.C., Pohl, J., Flocco, M.T. and Rothman, J.E. (1991) Nature, 353, 726-730.
21 Rüdiger, S, Freund, S.M.V., Veprintsev, D.B. and Fersht, A.R. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 11085-11090.
22 Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M.E. and Hendrickson, W.A. (1996) Science, 272, 1606-1614.
23 Erbse, A., Mayer, M.P. and Bukau, B. (2004) Biochem. Soc. Trans. 32, 617-621.
24 Han, W. and Christen, P. (2003) J. Biol. Chem. 278, 19038-19043.
25 Mayer, M.P., Schroder, H., Rudiger, S., Paal, K., Laufen, T. and Bukau, B. (2000) Nat. Struct. Biol. 7, 586-593.
26 Brehmer, D., Gassler, C., Rist, W., Mayer, M.P. and Bukau, B. (2004) J. Biol. Chem. 279, 27957-27964.
27 Slavotinek, A.M. and Biesecker, L.G. (2001) Trends Genet. 17, 528-535.
28 Lorimer, G.H. (1996) FASEB J. 10, 5-9.
29 Viitanen, P.V., Donaldson, G.K., Lorimer, G.H., Lubben, T.H. and Gatenby, A.A. (1991) Biochemistry 30, 9716-9723.
30 Smith, K.E., Voziyan, P.A. and Fisher, M.T. (1998) J. Biol. Chem. 273, 28677-28681.
31 Voziyan, P.A, Tieman, B., Low, C. M. and Fisher, M.T. (1998) J. Biol. Chem. 273, 25073-25078.
32 Houry, W.A., Frishman, D., Eckerskorn, D., Lottspeich, F. and Hartl, F.U. (1999) Nature 402, 147-154.
33 Horovitz, A., Fridmann, Y., Kafri, G. and Yifrach, O. (2001) J. Struct. Biol. 135, 104-114.
34 Todd, M.J., Viitanen, P.V. and Lorimer, G.H. (1994) Science 265, 659-666.
35 Yifrach, O. and Horovitz, A. (1995) Biochemistry 34, 5303-5308.
36 Yifrach, O. and Horovitz, A. (1996) J. Mol. Biol. 255, 356-361.
37 Saibil, H.R., Horwich, A.L. and Fenton, W.A. (2001) Adv. Prot. Chem. 19, 45-72.
38 Rye, H.S., Roseman, A.M., Chen, S., Furtak, K., Fenton, W.A., Saibil, H.R. and Horwich, A.L. (1999) Cell 97, 325-338.
39 Inbar, E. and Horovitz, A. (1997) Biochemistry 36, 12276-12281.
40 Chaudhuri, T.K., Farr, G.W., Fenton, W.A., Rospert, S. and Horwich, A.L. (2001) Cell 107, 235-246.
41 Farr, G.W., Fenton, W.A., Chaudhuri, T.K., Clare, D.K., Saibil, H.R. and Horwich, A.L. (2003) EMBO J. 22, 3220-3230.
42 Chaudhry, C., Farr, G.W., Todd, M.J., Rye, H.S., Brunger, A.T., Adams, P.D., Horwich, A.L. and Sigler, P.B, (2003) EMBO J. 22, 4877-4887.
43 Martin-Benito, J., Boskovic, J., Gomez-Puertas, P., Carrascosa, J.L., Simons, C.T., Lewis, S.A., Bartolini, F., Cowan, N.J. and Valpuesta, J.M. (2002) EMBO J. 21, 6377-6386.
44 Llorca, O., Martin-Benito, J., Grantham, J., Ritco-Vonsovici, M., Willison, K.R., Carrascosa, J.L. and Valpuesta, J.M. (2001) EMBO J. 20, 4065-4075.

45 Harris, S.F., Shiau, A.K. and Agard, D.A.. (2004) Structure 12, 1087-1097.
46 Kedzierska, S., Akoev, V., Barnett, M.E. and Zolkiewski, M. (2003) Biochemistry 42, 14242-14248.
47 Horwitz, J. (1992) Proc. Nat. Acad. Sci. U.S.A. 89, 10449-10453.
48 Lee, G.J., Roseman, A.M., Saibil, H.R. and Vierling, E. (1997) EMBO J. 16, 659-671
49 Haslbeck, M., Walke, S., Stromer, T., Ehrnsperger, M., White, H.E., Chen, S., Saibil, H.R. and Buchner, J. (1999) EMBO J. 18, 6744-6751.
50 Ito, H., Kamei, K., Iwamoto, I., Inaguma, Y. and Kato, K. (2001) Exp. Cell Res. 266, 213-221.
51 Biswas, A. and Das, K.P. (2004) J. Biol. Chem. 279, 42648-42657
52 Bhattacharyya, J. and Das, K.P. (1999) J. Biol. Chem. 274, 15505-15509.
53 Liegh–Edwards, K., Kueltzo, L.A., Fisher, M.T. and Middaugh, C.R. (2001) Arch. Biochem. Biophys.393, 14-21.
54 Strub, C., Schlieker, C., Bukau, B. and Mogk, A. (2003) FEBS Lett. 553, 125-130.
55 Coyle, J.E., Texter, F.L., Ashcroft, A.E., Masselos, D., Robinson, C.V. and Radford, S.E. (1999) Nat. Struct. Biol. 6, 683-690.
56 Llorca, O., Martin-Benito, J., Ritco-Vonsovici, M., Grantham, J., Hynes, G.M., Willison, K.R., Carrascosa, J.L. and Valpuesta, J.M. (2000) EMBO J. 19, 5971-5979.
57 Cowan, N.J. and Lewis, S.A. (2001) Adv. Protein Chem. 59, 73-104.
58 Steinbacher, S. (1999) Nat. Struct. Biol. 6, 1029-1032.
59 Cowan, N.J. and Lewis, S.A. (1999) Nat. Struct. Biol. 6, 990-991.
60 Choudhury, D., Thompson, A., Sojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J. and Knight, S. (1999). Science 285, 1061-1065.
61 Sauer, F.G., Futterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J. and Waksman, G. (1999) Science 285, 1058-1061.
62 Eisenberg, D. (1999) Science 285, 1021-1022.
63 Barnhart, M.M., Pinkner, J.S., Soto, G.E., Sauer, F.G., Langermann, S., Waksman, G., Frieden, C. and Hultgren, S.J. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 7709-7714.
64 Landry, S.J., Zeilstra-Ryalls, J., Fayet, O.L., Georgopoulos, C. and Gierasch L.M. (1993) Nature 364, 255-258.
65 Wang, Z., Feng, H.-P., Landry, S.J., Maxwell, J. and Gierasch, L.M. (1999) Biochemistry 38, 12537-12546.
66 Preuss, M., Hutchinson, J.P. and Miller, A.D. (1999) Biochemistry 38, 10272-10286.
67 Chen, L. and Sigler, P. (1999) Cell 99, 757-768.
68 Buckle, A.M., Zahn, R. and Fersht, A.R. (1997) Proc. Nat. Acad. Sci. U.S.A. 94, 3571-3575.
69 Deaton, J., Savva, C.G., Sun, J., Holzenburg, A., Berry, J. and Young, R. (2004) Protein Sci. 13, 1778-1786.
70 Woolhead, C.A., McCormick, P.J. and Johnson, A.E. (2004) Cell 116, 725-736.
71 Kudlicki, W., Coffman, A., Kramer, G. and Hardesty, B. (1997) Fold Des. 2, 101-108.

72 Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K. and Hartl, F.U. (1992) Nature 356, 683-689.
73 Matouschek, A., Pfanner, N. and Voos, W. (2000) EMBO Rep. 1,404-410.
74 Zahn, R., Perrett, S., Stenberg, G. and. Fersht, A.R. (1996) Science 271, 642-645.
75 Frieden, C. and Clark, A.C. (1997) Proc. Nat. Acad. Sci. U.S.A. 94, 5535-5538
76 Lin, Z. and Rye, H. (2004) Mol. Cell 16, 23-34.
77 Walter, S., Lorimer, G.H. and Schmid, F.X. (1996) Proc. Nat. Acad. Sci. U.S.A. 93, 9425-9430.
78 Martin, J., Horwich, A.L. and Hartl, F.U. (1992) Science 258, 995-998.
79 Shtilerman, M., Lorimer, G.H. and Englander, S.W. (1999) Science 284, 822-825.
80 Chen, J., Walter, S., Horwich, A.L. and Smith, D.L. (2001) Nat. Struct. Biol. 8, 721-728.
81 Singh, S.K., Grimaud, R., Hoskins, J.R., Wickner, S. and Maurizi, M. R. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 8898-8903.
82 Zeth, K., Raimond, B., Ravelli, R.B., Paal, K., Cusack, S., Bukau, B. and Dougan, D.A. (2002) Nat. Struct. Biol. 9, 906-911.
83 Wang, J., Hartling, J.A. and Flanagan, J.M. (1997) Cell 91, 447-456.
84 Lum, R., Tkach, J.M., Vierling, E. and Glover, J.R. (2004) J. Biol. Chem. 279, 29139-29146.
85 Shorter, J. and Lindquist S., (2004) Science 304, 1793-1797.
86 Inoue, Y., Taguchi, H., Kishimoto, A. and Yoshida M. (2004) J. Biol. Chem. (2004) 279, 52319-52323.
87 Hayer-Hartl, M.K., Weber, F. and Hartl, F.U. (1996) EMBO J. 15, 6111-6121.
88 Weissman, J.S., Hohl, C.M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K., Saibil, H.R., Fenton, W.A. and Horwich, A.L. (1995) Cell 83, 577-587.
89 Takagi, F., Koga, N. and Takada, S. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 11367-11372.
90 Thirumalai, D., Klimov, D.K. and Lorimer, G.H. (2003) Proc. Nat. Acad. Sci. U.S.A. 100, 11195-11197.
91 Fisher, M.T. and Yuan, X. (1994) J. Biol. Chem. 269, 29598-29601.
92 Weissman, J.S., Kashi, Y., Fenton, W.A. and Horwich, A.L. (1994) Cell 78, 693-702.
93 Fisher, M.T. (1998) Biochemistry (Moscow) 63, 453-472.
94 Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. and Bukau, B. (2001) Mol. Microbiol. 40, 397-413.
95 Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L. and Hartl, F.U. (1991) Nature 352, 36-42.
96 Michels, A.A., Kanon, B., Konings, A.W.T, Ohtsuka, K., Bensaude, O. and Kampinga, H.H. (1997) J. Biol. Chem. 272, 33282-33289.
97 Fedorov, A.N. and Baldwin, T.O. (1997) J. Mol. Biol. 268, 712-723.
98 Voziyan, P.A. and Fisher, M.T. (2000) Protein Sci. 9, 2405-2415.
99 Song, J.L. and Chuang, D.T. (2001) J. Biol. Chem. 276, 40241-40246.

100 Van Dyk, T.K., Gatenby, A.A. and LaRossa, R.A. (1989) Nature 342, 451-453.
101 Rutherford, S.L. and Lindquist, S. (1998) Nature 396, 336-342.
102 Rutherford, S.L. (2003) Nat. Rev. Gen. 4, 263-274.
103 Sangster, T.A., Lindquist, S. and Queitsch, C. (2004) BioEssays 26, 348-362.
104 Friedrich, K.L., Giese, K.C., Buan, N.R. and Vierling, E. (2004) J. Biol. Chem. 279, 1080-1089.
105 Koteiche, H.A. and McHaourab, H.S. (2003) J. Biol. Chem. 278, 10361-10367.
106 Pratt, W.B. and Toft, D.O. (2003) Exp. Biol. Med. 228, 111-133.
107 Ullers, R.S., Luirink, J., Harms, N., Schwager, F., Georgopoulos, C. and Genevaux, P. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 7583-7588.
108 Craig. E.A. (2003) Curr. Biol. 13, 904-905.
109 Hardy. S.J. and Randall. L.L. (1991) Science 251,.439-443.
110 Morais. M.C., Kanamaru. S., Badasso. M.O., Koti. J.S., Owen. B.A., McMurray. C.T., Anderson. D.L. and Rossmann. M.G. (2003) Nat. Struct. Biol. 10, 572-576.
111 Cunningham, E.L., Mau, T., Truhlar, S.M. and Agard, D.A. (2002) Biochemistry 41, 8860-8867.
112 Cunningham, E.L. and Agard, D.A. (2004) Protein Sci. 13, 325-331.
113 Zheng, X, Rosenberg, L.E, Kalousek, F. and Fenton, W.A. (1993) J. Biol. Chem. 268, 7489-7493.
114 Fisher, M.T. (1994) J. Biol. Chem. 269, 13629-13636.
115 Fisher, M.T. (1993) J. Biol. Chem. 268, 13777-13779.
116 Smith, K.E. and Fisher, M.T. (1995) J. Biol. Chem. 270, 21517-21523.
117 Voziyan, P.A., Jadhav, L. and Fisher, M.T. (2000) J. Pharm. Sci. 89, 1036-1045.
118 Voziyan, P.A. and Fisher, M.T. (2002) Arch. Biochem. Biophys. 397, 293-297.
119 Seale, J.W., Gorovits, B.M., Ybarra, J. and Horowitz, P.M. (1996) Biochemistry 35, 4079-4083.
120 Tyedmers, J., Kruse, M., Lerner, M., Demand, J., Höhfeld, J., Solsbacher, J., Volkmer, J. and Zimmermann, R. (2000) Eur. J. Biochem. 267, 3575-3580.
121 Freeman, B.C. and Yamamoto, K.R. (2002) Science 296, 2232-2235.
122 Morimoto, R.I (2002) Cell 110, 281-284.
123 Brocchieri, L. and Karlin, S. (2000) Protein Sci. 9, 476-486.
124 Ivic, A., Olden, D., Wallington, E.J., and Lund, P.A. (1997) Gene 194, 1-8.
125 Klunker, D., Haas, B., Hirtreiter, A., Figueiredo, L., Naylor, D.J., Pfeifer, G., Müller, V., Deppenmeier, U., Gottschalk, G., Hartl, F.U. and Hayer-Hartl, M. (2003) J. Biol. Chem. 278, 33256-33267.
126 Ang, D., Keppel, F., Klein, G., Richardson, A. and Georgopoulos, C. (2000) Ann. Rev. Genet. 34, 439-456.
127 George, R., Kelly, S.M., Price, N.C., Erbse, A., Fisher, M.T. and Lund, P.A. (2004) Biochem. Biophys. Res. Comm. 324, 822-828.
128 Ryan, M.T. and Pfanner, N. (2001) Adv. Protein Chem. 59, 223-242.

129 Vickery, L.E., Silberg, J.J. and Ta, D.T. (1997) Protein Sci. 6, 1047-1056.
130 Agashe, V.R., Guha, S., Chang, H.C., Genevaux, P., Hayer-Hartl, M., Stemp, M., Georgopoulos, C., Hartl, F.-U. and Barral, J.M. (2004) Cell 117, 199-209.
131 Zmijewski, M.A., Macario, A.J. and Lipinska, B. (2004) J. Mol. Biol. 336, 539-549.
132 Gallert, H., Rutkat, K. and Buchner, J., (2000) J. Biol. Chem. 275, 20424-20430.
133 Yifrach, O. and Horovitz, A. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 1521-1524.
134 Motojima, F., Chaudhry, C., Fenton, W.A., Farr, G.W. and Horwich, A.L. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 15005-15012.
135 Falke, S.R., Tama, F., Brooks, C.A. III, Gogol, E. and Fisher, M.T. (2004) J. Mol. Biol. (2005) 348, 219-230.
136 Wang, J. and Chen, L. (2003) J. Mol. Biol. 334, 489-499.
137 Baneyx, F. and Mujacic, M. (2004) Nat. Biotechnol. 22, 1399-1408.
138 Yoder, M.D., Thomas, L.M., Tremblay, J.M., Oliver, R.L., Yarbrough, L.R. and Helmkamp, G.M., Jr. (2001) J. Biol. Chem. 276, 9246-9252.
139 Altamirano, M.M., Garcia, C., Possani, L.D. and Fersht, A.R. (1999) Nat. Biotechnol. 17, 187-191.
140 Wang, J.D., Michelitsch, M.D. and Weissman, J.S. (1998) Proc. Nat. Acad. Sci. U.S.A. 95, 12163-12168.
141 Weber, F., Keppel, F., Georgopoulos, C., Hayer-Hartl, M.K. and Hartl, F.U. (1998) Nat. Struct. Biol. 5, 977-985.
142 Tieman, B.C., Johnston, M. and Fisher, M.T. (2001) J. Biol. Chem. 276, 44541-44550.
143 Voziyan, P.A , Johnston, M., Chao, A., Bomhoff, G. and Fisher, M.T. (2004) J. of Struct. and Functional Genomics 31, 1-6 (2005).
144 Liu, Y. and Bolen, D.W. (1995) Biochemistry 34, 12884-12891.
145 Diamant, S., Eliahu, N., Rosenthal, D. and Goloubinoff, P. (2001) J. Biol. Chem. 276, 39586-39591.
146 Lee, H.J. and Lee, S.J. (2002) J. Biol. Chem. 277: 48976-48983.
147 Nollen, E.A.A., Garcia, S.M., Haaften, G.V., Kim, S., Chavez, A., Morimoto, R.I. and Plasterk, R.H.A. (2004) Proc. Nat. Acad. Sci. U.S.A. 101, 6403-6408.
148 Berkovich, B., Stancovski, I., Mayer, A., Blumenfeld, N., Laszlo, A., Schwartz, A.L. and Ciehanover, A. (1997) J. Biol. Chem. 272, 9002-9010.
149 Jiang, J,, Ballinger, C.A., Wu, Y., Dai, Q., Cyr, D.M., Hohfeld, J. and Patterson, C. (2001) J. Biol. Chem. 276, 42938-42944.
150 Choo-Kang, L.R. and Zeitlin, P.L. (2001) Amer. J. Physiol. Lung Cell Mol. Physiol. 281, L58-L68.
151 Rubenstein, R.C., Egan, M.E. and Zeitlin, P.L. (1997) J. Clin. Invest. 100, 2457-2465.
152 Burrows, J.A.J., Willis, L.K. and Perlmutter, D.H. (2000) Proc. Nat. Acad. Sci. U.S.A. 97, 1796-1801.

ENGINEERING PLANTS FOR INCREASED NUTRITION AND ANTIOXIDANT CONTENT THROUGH THE MANIPULATION OF THE VITAMIN E PATHWAY

David K. Shintani

Department of Biochemistry and Molecular Biology
University of Nevada
Reno, Nevada 89557

INTRODUCTION

Commercial interest in the therapeutic application of phytochemicals has lead to resurgence in industrial and academic research in plant biochemistry and metabolic engineering. The biosynthesis of vitamin E represents one such area of intense activity. In addition to being an essential component of human and animal diets, consumption of vitamin E has been suggested to decrease the occurrence of some cancers and degenerative diseases, including prostate cancer, atherosclerosis, Alzheimer's disease, cataracts, and diabetes mellitus (1). Unfortunately, because the efficacy of some of these studies have been questioned, researchers are reluctant to widely prescribe increased daily intakes of vitamin E to the general public (1). However, due to its low toxicity, many physicians still consider vitamin E as a noninvasive prophylactic therapy for patients suffering from degenerative diseases such as Alzheimer's (2) and Parkinson's diseases (3). For this reason, the demand for vitamin E dietary supplements has been steadily increasing, driving genetic engineering and breeding efforts to biofortify food crops for increased vitamin E content.

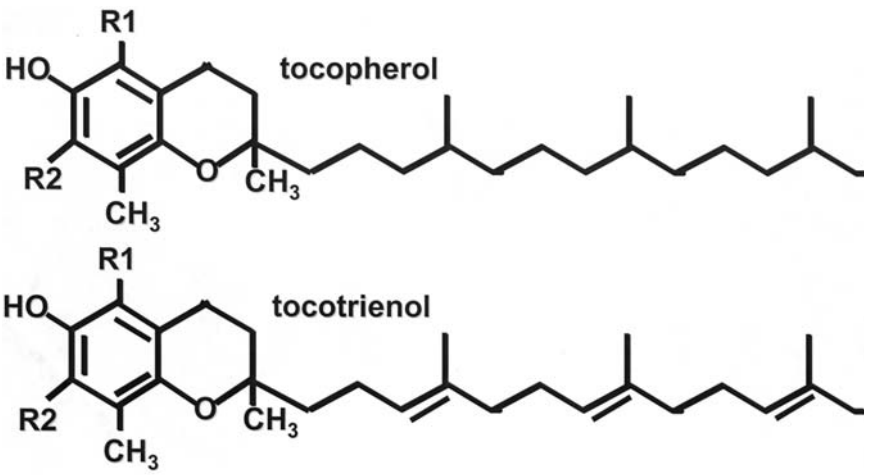
Genetic Engineering, Volume 27, Edited by J. K. Setlow

WHAT IS VITAMIN E?

Vitamin E comprises a group of compounds known collectively as tocols, which are lipophilic compounds possessing a polar chromanol head group attached to a nonpolar prenyl tail. Depending on the degree of saturation of the prenyl tail, tocols are designated as tocopherols or tocotrienols (Figure 1). Tocols with saturated prenyl tails derived from phytyl-pyrophosphate are designated as tocopherols, whereas those with polyunsaturated prenyl tails derived from geranylgeranyl pyrophosphate are designated as tocotrienols. Tocopherols and tocotrienols are further subclassified based on the number of methyl-groups and the position of these methyl groups on the chromanol head group. Tocols containing one methyl group are designated as δ-tocols. Tocols containing two methyl groups are designated as γ- or β-tocols. And fully methyl-substituted tocols containing three methyl groups are known as α-tocols (Figure 1).

NUTRITIONAL IMPORTANCE OF VITAMIN E

Vitamin E was first discovered by Evans and Bishop in 1922 (4). They found that a compound present in lettuce leaves was essential for fertility in rats such that deficiencies of this compound caused fetal resorption in pregnant animals. Other symptoms of vitamin E deficiencies observed in test animals included



R1	R2	Tocol Species
H	H	δ-tocopherol/tocotrienol
CH_3	H	β-tocopherol /tocotrienol
H	CH_3	γ-tocopherol /tocotrienol
CH_3	CH_3	α-tocopherol/tocotrienol

Figure 1. Chemical structures of tocopherols/tocotrienols.

testicular atrophy, necrotizing myopathy, central nervous system necrosis, erythrocyte hemolysis, and liver and kidney necrosis (5). The active substance isolated from lipid fractions of lettuce leaves was designated tocopherol from the Greek phrase meaning "to bear offspring."

Fortunately, vitamin E deficiencies are rare in humans. In fact, the recommended daily allowance (RDA) for vitamin E of 22 international units (IU) is easily achievable from even relatively poor diets. However, vitamin E deficiencies are observed in persons with severely impaired fat absorption, cystic fibrosis, and chronic cholestatic hepatobiliary disease and in prematurely born infants with low vitamin E reserves (5). In these cases, the deficiencies are manifested as severe anemia. Two rare congenital disorders, one resulting from an inability to synthesize very-low-density lipoproteins (VLDLs) (6) and the other from the lack of hepatic tocopherol transfer protein (7), prevent the normal absorption of vitamin E from the diet. Patients suffering from these genetic maladies suffer severe neurodegenerative disorders.

One International Unit (IU) of vitamin E equals 1 mg of dl-α-tocopheryl acetate or the specific quantity of a tocol species required to prevent resportion-gestation in 50% of female rats deprived of vitamin E. Of all tocols, α-tocopherol is by far the most potent form of vitamin E, possessing between 2 and 33 times higher vitamin E activity than all other tocopherol or tocotrienol species (Table 1). This is primarily due to the fact that animals and humans possess a tocopherol transfer protein (TTP) that is highly specific for the (R,R,R)-α-tocopherol steroisomer and ensures the preferential absorption and distribution of α-tocopherol throughout the body (8). As such, the biosynthesized α-tocopherol, which exists as the (R,R,R)-α-tocopherol steroisomer, is more potent than synthetic α-tocopherol, which is sold as a racemic mixture of more than 50 different steroisomers. Furthermore, α-tocopherol is better absorbed from a food source than from tablet/capsule supplements (9). These observations provide the primary motivation for increasing natural vitamin E production in food plants.

Table 1. Vitamin E activity of naturally occurring tocol species.

	IU*/mg (5)	**Relative Vitamin E Activity (5)**
D-α-tocopherol (RRR)	1.49	100%
D-β-tocopherol (RRR)	0.75	50%
D-γ-tocopherol (RRR)	0.15	10%
D-δ-tocopherol (RRR)	0.05	3%
D-α-tocotrienol	0.75	50%
D-β-tocotrienol	0.08	5%
D-γ-tocotrienol	not known	not known
D-d-tocotrienol	not known	not known

*IU = International Unit

ANTIOXIDANT FUNCTION

Tocopherols and tocotrienols function as antioxidants by scavenging and detoxifying free radicals and activated oxygen species. The tocol chromanol head group plays an important role in the inactivation of oxidizing agents (10). Singlet oxygen quenching, for example, is achieved through resonance energy transfer to the bicyclic aromatic chromanol ring. In regards to free radical quenching, the phenolic hydrogen on position 6 of the ring can be readily donated to reduce radical species, resulting in the formation of a new tocol radical, which can subsequently be recycled through its reduction with ascorbic acid or glutathione.

While the prenyl tails of tocols do not actively participate in redox reactions, it is thought that they may play a role in membrane stabilization (11). *In vitro* studies suggest that the interaction of the tocol prenyl tail with polyunsaturated fatty acid components of lipid membrane functions to broaden the temperature range of membrane gel and liquid-crystalline phase transitions, thereby extending the membrane stability over a greater temperature range (11). Surprisingly, even though α-tocopherol is the most potent form of vitamin E *in vivo*, the relative order of the *in vitro* antioxidant potential from highest to lowest ranges is δ-tocopherol > β- and γ-tocopherol > α-tocopherol (10).

OCCURRENCES AND ROLE OF TOCOPHEROLS AND TOCOTRIENOLS IN PLANT TISSUES

Vitamin E is only synthesized by photosynthetic organisms, including cynanobacteria, algae, and plants (12). In higher plants, tocol synthesis is localized within the plastid (13). Although all plants are capable of synthesizing tocols, the relative amounts and types produced vary not only between species, but also between different tissue types within a given species (14).

Oilseeds are by far the richest sources of vitamin E, having total tocol levels ranging from 200 to 1,200 μg per gram of oil (see Table 2) (14, 15). Sunflower and olive oils, which are commonly consumed in European/Mediterranean diets, are particularly good sources of vitamin E because they not only produce tocols in large abundance, but the primary tocol formed is α-tocopherol, the most potent form of vitamin E. Unfortunately, the most highly consumed vegetable oils in American diets (i.e., soybean, corn, and rapeseed oil) accumulate primarily γ-tocopherol, which has one-tenth of the vitamin E activity of α-tocopherol.

Oilseeds have been hypothesized to accumulate tocopherols to protect storage lipids from oxidative damage. While it has been known that tocols play an important role protecting extracted vegetable oils from oxidative damage, the assumed *in vivo* function of tocols as antioxidants in oilseeds was not proven. However, recently, it has been reported that the germination of seeds from *Arabidopsis* mutants incapable of producing tocopherols was significantly impaired due to the nonenzymatic oxidation of seed storage lipid and subsequent damage to the developing embryo (16). Products of lipid oxidation accumulated in the tocopherol mutant seedlings but were absent in wild-type plants, indicating a protective role for seed tocols.

Table 2. Tocopherol content and composition of important dietary sources of Vitamin E.

	Content (μg/gm food)	**% α-tocopherol**	**% γ-tocopherol**	**% β-tocopherol**	**% δ-tocopherol**
Oils (15)					
Soybean	1200	7.00	70.00	1.00	22.00
Corn(15)	1000	22.00	68.00	3.00	7.00
Sunflower	700	96.00	2.00	2.00	n.d.
Rapeseed	630	41.27	57.14	n.d.	1.59
Olive	240	100	trace	trace	trace
Green Vegetables (14)					
Spinach	30	62.67	4.67	trace	32.67
Parsley	25.3	68.77	7.11	trace	24.11
Cabbage	16.7	100	n.d.	n.d.	n.d.
Leek	9.2	100	n.d.	n.d.	n.d.
Lettuce	7.5	60	40	n.d.	n.d.

The evolutionary basis for some oils being rich in α-tocopherol and others being rich in γ-tocopherol may be related to the fatty acid composition of the oil. Oil seeds that predominately accumulate γ-tocopherol also produce relatively large amounts of polyunsaturated fatty acids (i.e., linoleic and linolinic acids), which are highly susceptible to oxidative damage. Such seeds may have evolved to accumulate γ-tocopherol instead of α-tocopherol because of γ-tocopherol's superior properties as a chemical antioxidant (10).

The tocol compositions of leaves and shoots are primarily made up of α-tocopherol (Table 2). However, compared with seeds that are rich in total tocol content, leaves and shoots possess relatively low tocol levels (i.e., 5-30 μg/gm fresh weight) (14). The role of α-tocopherol in photosynthetic tissues is not clear. The observation that α-tocopherol accumulates in leaves in response to abiotic stress (17) leads to the assumption that α-tocopherol protects the chloroplast from oxidative damage. However, recent studies cast doubt on this assumption. In *Arabidopsis*, it was shown that mutant plants devoid of tocopherol were no more susceptible to conditions known to promote oxidative damage (i.e., high and low temperature, paraquat application, and high light) than wild-type plants (18). Furthermore, transgenic *Arabidopsis* plants with elevated α-tocopherol levels were shown to be no more resistant to high light stress than wild-type plants (19). Further studies clearly need to be performed to deduce the role of tocopherols in green tissues.

Whereas tocopherols are ubiquitous among photosynthetic organisms, tocotrienols are primarily restricted to the seed tissues of monocot plants. Grains such as wheat, rye, and barley accumulate relatively high levels of both tocotrienols and tocopherols (15). Palm kernel oil is also a particularly rich source of tocotrienols (20).

TOCOL BIOSYNTHETIC PATHWAYS

Eloquent radiolabeling studies were conducted to deduce the biosynthetic route leading to the formation of α-tocopherol (Figure 2) (13, 21). These studies revealed that tocopherols are synthesized from two major precursors, homogentisate acid and phytyl-pyrophosphate (phytyl-PP). Homogentisate acid (HGA) is the aromatic precursor for the biosynthesis of tocols and plastiquinones (22). This compound is derived from p-hydroxyphenyl-pyruvate (HPP), a product of the shikimic acid pathway (23), and is formed from a reaction catalyzed by the enzyme HPP dioxygenase (HPPD). Due to its essential role in plastiquinone biosynthesis, HPPD has been an attractive herbicide target and has lead to the development of a relatively new class of triketone herbicides such as sulcotrione (24).

The other major pathway precursor, PDP, is derived from isopentenyl pyrophosphate (IPP) arising from the plastid localized methylerythritol phosphate (MEP) pathway (25). IPP is converted to dimethylallyl-pyrophosphate (DMAPP) by the plastid localized IPP isomerase. IPP and DMAPP then serve as substrates for a plastid localized geranylgeranyl pyrophosphate (GGPP) synthase in the synthesis of GGDP. GGDP can then serve as a precursor for tocotrienol biosynthesis or be reduced to phytyl-PP (PDP) in a reaction catalyzed by GGDP reductase (26).

In the first committed step in tocopherol biosynthesis, HGA and PDP are condensed to form 2-methyl-6-phytylbenzoquinol (MPBQ) (21) in a reaction catalyzed by the enzyme HGA phytyltransferase (HPT). MPBQ, the first prenylquinone intermediate, can then either be cyclized by tocopherol cyclase to form δ-tocopherol (18, 27, 28) or be methylated by MPBQ methyltransferase to form 2,3-dimethyl-6-phytylbenzoquinol (DMPBQ) (21). The tocopherol cyclase can then react with DMPBQ to form γ-tocopherol (18, 27, 28). δ-Tocopherol and γ-tocopherol then both serve as substrates for γ-tocopherol methyltransferase and form β-tocopherol and α-tocopherol respectively (29, 30). Whereas α-tocopherol is a terminal product, β-tocopherol could be methylated once more to form α-tocopherol in a yet to be defined methyltransferase reaction. The biosynthesis of tocotrienols occurs through an analogous set of reactions that differ only in the first step with homogentisate acid being condensed with GGDP, instead of PDP, to form 2-methyl-6-geranylgeranylplastoquinol in a reaction catalyzed by HGA geranylgeranyl transferase (HGGT) (Figure 2) (31).

ENGINEERING PLANTS FOR INCREASED TOCOL CONTENT

Flux through the tocopherol biosynthetic pathway is regulated by enzymes controlling the synthesis and condensation of the two pathway precursors, HGA and PDP. The enzymes primarily responsible for the synthesis of HGA and PDP are HPP dioxygenase (HPPD) and GGDP reductase (GGDPR), respectively (Figure 1). HPPD and GGDPR were selected as candidate regulatory enzymes based on *in vivo* observations showing that HGA and PDP are limiting for tocopherol biosynthesis. Tissue culture studies have shown that exogenously supplied HGA and PDP both caused significant increases in tocopherol biosynthesis (32, 33). Furthermore, when PDP accumulates upon dark senescence due

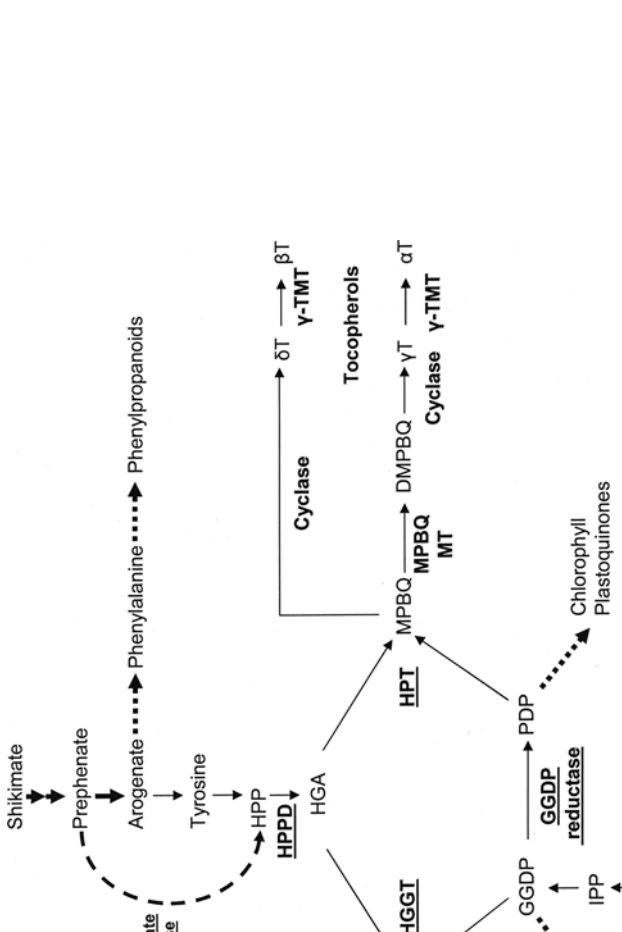


Figure 2. Schematic representation of tocopherol/tocotrienol biosynthesis in plants. Enzymes that are responsible for flux regulation of the pathway are underlined whereas enzymes that are responsible for an overall compositional change in tocol content are highlighted in gray. Dotted lines represent alternative pathways that compete for tocopherol precursors. Dashed line represents biosynthetic route present in yeast but normally absent in plants. Abbreviations: DMGGBQ, dimethylgeranylgeranylbenzoquinol; DMPBQ, dimethylphytylbenzoquinol; GGDP, geranylgeranyl diphosphate; HPP, p-hydroxyphenyl-pyruvate; HPPD, p-hydroxyphenyl-pyruvate dioxygenase; HGGT, homogentisate geranylgeranyl transferase; HPT, homogentisate phytyltransferase; IPP, isopentenyl pryophosphate; MEP, methylerythritol phosphate; MGGBQ, methylgeranylgeranylbenzoquinol; MPBQ, methylphytylbenzoquinol; MPBQMT, methylphytylbenzoquinol methyltransferase; g-TMT, g-tocopherol methyltransferase; T, tocopherol; T3, tocotrienol.

to chlorophyll breakdown, large increases in tocopherol biosynthesis are observed (34).

Transgenic approaches were used to test the hypothesis that rates of tocopherol biosynthesis were regulated by the availability of HGA. In these studies, the HPPD gene was overexpressed in the leaves and seeds of both *Arabidopsis* and tobacco (35, 36). Surprisingly, only small increases in tocopherol content were observed in both leaves and seeds of transgenic lines possessing large increases in HPPD activity (35, 36). These results indicate that HPPD is not the only factor controlling tocopherol biosynthetic flux. Interestingly, overexpression of the HPPD gene in plants also resulted in plants with increased resistance to sulcotrione (36, 37).

In an attempt to further increase HGA levels, Rippert et al. (37) overexpressed the yeast prephenate dehydrogenase, an enzyme not found in plants, in tobacco leaves to redirect the shikimic acid pathway flux away from the production of phenylpropanoid compounds and towards the synthesis of tocopherols (Figure 2). Prephenate dehydrogenase catalyzes the direct conversion of prephenate to HPP, bypassing three enzymatic steps that normally occur in production of HPP in plants. When the yeast prephenate dehydrogenase gene was expressed alone, only a slight increase in tocol biosynthesis was observed. However, when the prephenate dehydrogenase was co-overexpressed with the HPPD gene in transgenic tobacco leaves, a 10-fold increase in tocol production was observed. Surprisingly, the predominate tocol species produced were tocotrienols with α-tocotrienol present at 70% of the total. Because tocotrienols are not normally produced in dicot leaves, these results suggest that PDP is limiting and that the excess HGA produced is being condensed with GGDP to drive the production of tocotrienols. Therefore, it is likely that GGDPR activity is limiting in leaves.

Although similar experiments have not yet been performed to test directly the effect of increasing PDP levels on tocopherol yields through overexpression of the GGDP reductase, it has been shown that increasing the total flux through the MEP pathway by overexpressing the deoxyxylulose phosphate synthase gene resulted in a 40% increase in leaf tocopherol content (25).

In addition to precursor availability, tocopherol biosynthesis is regulated by the rate at which homogentisate phytyltransferase (HPT) condenses HGA and PDP to MPBQ, the first true tocopherol biosynthetic intermediate. When the HPT gene was overexpressed in transgenic Arabidopsis, a 10-fold increase in HPT activity translated into a 4.4-fold increase in leaf tocopherol over wild-type levels (19). Similar studies performed in seeds also resulted in increased tocopherol levels, but the magnitude of the increase was only 0.4- to 2-fold over wild-type levels (19, 38)

Recently, Cahoon et al. (31) cloned a gene encoding a seed-specific barley HPT whose sequence diverged significantly from previously characterized dicot HPTs. Functional analyses revealed that this gene encoded an enzyme that utilized GGDP as its prenyl substrate instead of PDP and led to the production of tocotrienols. Analogous to HPT in tocopherol biosynthesis, this novel enzyme, homogentisate geranylgeranyltransferase (HGGT), plays an important role in regulating pathway flux through tocotrienol biosynthesis. Evidence of this can be seen in studies where the overexpression of the HGGT gene in maize caused a

20-fold increase in tocotrienol levels and an 8-fold increase in total tocols (tocopherols + tocotrienols) (31). This result represents the largest increase in tocol production ever observed in plants and significantly increases the antioxidant potential of corn. Unfortunately, since dietary tocotrienols are not absorbed as well as α-tocopherol, the large increase in tocotrienol levels observed in the HGGT overexpressing maize seed did not add much to the vitamin E nutritional value of these plants. However, it has been reported that tocotrienols may have a novel therapeutic role in decreasing cholesterol levels in humans (39). Furthermore, because tocotrienols have superior *in vitro* antioxidant activities (10), transgenic plants with elevated tocotrienol levels could be used as sources of chemical antioxidants for industrial applications such as oxidative stabilizers for paints, coatings, and other lipophilic products.

Interestingly, when HGGT was overexpressed in *Arabidopsis* leaves the tocopherol content was not affected, but tocotrienols, which are not normally synthesized by dicot species, accumulated to very high levels (31). This result shows that HGGT and HPT are highly specific for their prenyl substrates, GGDP and PDP, respectively, and must compete with one another for HGA. Although overproduction of tocotrienols clearly shows that GGDP and HGA are abundantly available, the fact that HGGT overexpression had no effect on tocopherol production indicates that PDP must be limiting. Therefore, overexpression of GGDP reductase could significantly increase PDP availability and thus increase pathway flux.

ENGINEERING PLANTS WITH ALTERED TOCOL COMPOSITION

The enzymes important in determining tocol composition, methylphytylbenzoquinone methyltransferase (MPBQMT), tocopherol cyclase (TC), and γ-tocopherol methyltransferase (γ-TMT) appear in the terminal half of the tocopherol biosynthetic pathway. The latter enzyme, γ-TMT, is particularly important in determining the Vitamin E content of oilseeds, such as soybean, rapeseed and maize kernels. Instead of accumulating α-tocopherol, these crops accumulate γ-tocopherol, which has only one-tenth the vitamin E activity of α-tocopherol. It has been hypothesized that the enzyme responsible for the conversion of γ-tocopherol to α-tocopherol, γ-TMT, must be limiting in these oilseed crops. To test this hypothesis, the γ-TMT gene was overexpressed in *Arabidopsis* seed, which like many important oilseed crops accumulates γ-tocopherol as its predominant tocol species (29). Analysis of plants overexpressing the γ-TMT behind a seed-specific promoter resulted in 95% conversion of the γ-tocopherol to α-tocopherol (29).

In a subsequent study, van Eenennaam et al. (40) simultaneously overexpressed the γ-TMT and MPBQMT genes in soybean seeds. Like *Arabidopsis*, soybean seeds contain only low levels of α-tocopherol; however, instead of accumulating primarily γ-tocopherol, soybean seeds also accumulate significant amounts of δ-tocopherols (i.e., 22% of total tocopherols). Consistent with the results of Shintani and DellaPenna (29), the overexpression of the γ-TMT and MPBQMT genes resulted in an almost complete conversion of seed tocopherol pools to α-tocopherol, thus increasing the vitamin E activity of soybean oil seven-fold.

Interestingly, γ-TMT is not normally limiting in green tissues; it has been reported that pathway flux increases caused by overexpression of the HPT gene result in γ-tocopherol accumulation, indicating that under these conditions γ-TMT levels are limiting (19). Therefore, overexpression of the HPT and the γ-TMT genes is necessary to optimize α-tocopherol levels in green tissues.

So far, no studies have been performed to determine what role if any TC plays in shaping tissue tocol compositions. However, it is clear that manipulation of the γ-TMT and MPBQMT genes allows us the ability to engineer plants with «tailor-made» tocopherol compositions. We can now design plants that accumulate only α-, β-, γ-, or δ-tocopherol, which would not only impact the nutritional content of crop foods, but also allow for the production of natural antioxidants to replace synthetic antioxidants for food processing and industrial applications.

CONCLUSIONS

Recent advances in our understanding of tocopherol biosynthesis have helped us identify key regulatory enzymes controlling pathyway flux and tocopherol composition. This information is invaluable to plant breeders and biotechnologists who wish to create commercial crops with both elevated and tailored tocopherol compositions that can be utilized for food and industrial chemicals. The biofortification of crop foods with elevated vitamin E levels should be one of the major priorities for crop improvement. Our ability to develop novel crop foods that are biofortified for vitamin E will have a significant impact on the health of the general public. Although fiscal gains have been one of the primary motivating forces behind these efforts, altruistic aims have also driven this research. Specifically, groups are interested in developing foods that have been biofortified for vitamin E. The rationale being that substantial increases in the vitamin E content of food crops are needed to provide the public with dietary sources that can approach the therapeutic levels needed to achieve the desired health benefits of vitamin E. As yet, no such food crop exists and, with current sources, one would need to consume approximately 730 gm of soybean oil or more than 3.5 kg of spinach to obtain the minimum recommended therapeutic levels of vitamin E (100 IU/day) to prevent coronary heart disease. Neither option is practical and clearly illustrates the need for improved food crops with elevated levels of α-tocopherol. Biofortified plants would provide a sustainable alternative to a prescribed regimen of vitamin E supplementation that would be available to everyone regardless of income or class.

REFERENCES

1 Food and Nutrition Board (2000) in Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids (Institute of Medicine, eds.) pp. 186-283, National Academy Press, Washington D.C.
2 Berman, K. and Brodaty, H. (2004) CNS Drugs 18, 807-825.
3 Fariss, M.W. and Zhang, J.G. (2003) Toxicology 189, 129-146.
4 Evans, H.M. and Bishop, K.S. (1922) Science 56, 650-651.

5 Bender, B.A. (2003) in Nutritional Biochemistry of the Vitamins, 2nd ed. pp 109-130, Cambridge University Press, Cambridge, UK.
6 Muller, D.P., Lloyd, J.K. and Wolff, O.H. (1983) Lancet 1, 225-228.
7 Aparicio, J.M., Belanger-Quintana, A., Suarez, L., Mayo, D., Benitez, J., Diaz, M. and Escobar, H. (2001) J. Pediatr. Gastroenterol. Nutr. 33, 206-210.
8 Kaempf-Rotzoll, D.E., Traber, M.G. and Arai, H. (2003) Curr. Opin. Lipidol. 14, 249-254.
9 Leonard, S.W., Good, C.K., Gugger, E.T. and Traber, M.G. (2004) Am. J. Clin. Nutr. 79, 86-92.
10 Kamal-Eldin, A. and Appelqvist, L.A. (1996) Lipids 31, 671-701.
11 Fryer, M.J. (1992) Plant Cell Environ. 15, 381-392.
12 Hess, J.L. (1993) in Antioxidants in Higher Plants (Alscher, R.G. and Hess, J.L., eds.) pp. 111-134, CRC Press, Boca Raton, FL .
13 Soll, J., Kemmerling, M. and Schultz, G. (1980) Arch. Biochem. Biophys. 204, 544-550.
14 MacLaughlin, P.J. and Weihrauch, J.L. (1979) J. Am. Diet. Assoc. 75, 647-665.
15 Taylor, P. and Barnes, P. (17 October 1981) Chem. Ind., 722-726.
16 Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M. and DellaPenna, D. (2004) Plant Cell 16, 1419-1432.
17 Munne-Bosch, S. and Alegre, L. (2003) Plant Physiol. 131, 1816-1825.
18 Porfirova, S., Bergmuller, E., Tropf, S., Lemke, R. and Dormann, P. (2002) Proc. Nat. Acad. Sci. U.S.A. 99, 12495-12500.
19 Collakova, E. and DellaPenna, D. (2003) Plant Physiol. 131, 632-642.
20 Choo, Y.M., Ma, A.N., Chuah, C.H., Khor, H.T. and Bong, S.C. (2004) Lipids 39, 561-564.
21 Soll, J. and Schultz, G. (1980) Phytochemistry 19, 215-218.
22 Whistance, G.R. and Threlfall, D.R. (1970) Biochem. J. 117, 593-600.
23 Fiedler, E., Soll, J. and Schultz, G. (1982) Planta 155, 511-515.
24 Secor, J. (1994) Plant Physiol. 106, 1429-1433.
25 Estevez, J.M., Cantero, A., Reindl, A., Reichler, S. and Leon, P. (2001) J. Biol. Chem. 276, 22901-22909.
26 Soll, J. and Schultz, G. (1981) Biochem. Biophys. Res. Commun. 99, 907-912.
27 Stocker, A., Fretz, H., Frick, H., Ruttimann, A. and Woggon, W.D. (1996) Bioorg. Med. Chem. 4, 1129-1134.
28 Sattler S.E., Cahoon, E.B., Coughlan, S.J. and DellaPenna, D. (2003) Plant Physiol. 132, 2184-2195.
29 Shintani, D. and DellaPenna, D. (1998) Science 282, 2098-100.
30 d'Harlingue, A. and Camara, B. (1985) J. Biol. Chem. 260, 15200-15203.
31 Cahoon, E.B., Hall, S.E., Ripp, K.G., Ganzke, T.S., Hitz, W.D. and Coughlan, S.J. (2003) Nature Biotech. 21, 1082-1087.
32 Furuya, T., Yoshikawa, T., Kimura, T. and Kaneko, H. (1987) Phytochemistry 26, 2741-2747.
33 Caretto, S., Bray Speth, E., Fachechi, C., Gala, R., Zacheo, G. and Giovinazzo, G. (2004) Plant Cell Rep. 23, 174-179.

34 Moshe, R., Cojocaru, M., Gottlieb, H.E. and Goldschmidt, E.E. (1989) Plant Physiol. 89, 1028-1030.
35 Falk, J., Andersen, G. and Kernebeck, B.K.K. (2003) FEBS Lett. 540, 35-40.
36 Tsegaye, Y., Shintani, D.K. and DellaPenna, D. (2002) Plant Physiol. Biochem. 40, 913-920.
37 Rippert, P., Scimemi, C., Dubald, M. and Matringe, M. (2004) Plant Physiol. 134, 92-100.
38 Savige, B., Weiss, J.D., Wong, Y.H., Lassner, M.W., Mitsky, T.A., Shewmaker, C.K., Post–Beittenmiller, D. and Valentin, H.E. (2002) Plant Physiol. 129, 321-332.
39 Theriault, A., Chao, J.T., Wang, Q., Gapor, A. and Adeli, K. (1999) Clin. Biochem. 32, 309-319.
40 Van Eenennaam, A.L., Lincoln, K., Durrett, T.P., Valentin, H.E., Shewmaker, C.K., Thorne, G.M., Jiang, J., Baszis, S.R., Levering, C.K., Aasen, E.D., Hao, M., Stein, J.C., Norris, S.R. and Last, R.L. (2003) Plant Cell 15, 3007-3019.