Computer Analysis of Phytochrome Sequences from Five Species: Implications for the Mechanism of Action

Michael D. Partis

A.F.R.C. Institute of Horticultural Research, Worthing Road, Littlehampton, W. Sussex, BN 17, 6 LP, U.K.

and

Rudolf Grimm

Institut für Biologie II/Botanik, Albert-Ludwigs-Universität, Schanzlestraße 1, D-7800 Freiburg, Bundesrepublik Deutschland

Z. Naturforsch. **45c**, 987–998 (1990); received April 12, 1990

Phytochrome, Sequence Analysis, Amphiphilic Peptides, Membrane Interaction, Targeting Sequences

The amino acid sequences of phytochrome from *Avena sativa, Oryza sativa, Curcurbita pepo, Pisum sativum* and *Arabidopsis thaliana* have been analyzed with a variety of computer programs, with a view to identifying areas of the protein which contribute to the properties of this photoreceptor. A region at the C-terminus has been shown to be amphiphilic, and by analogy with surface-seeking peptides, may be responsible for interaction of phytochrome with lipid bilayers. Possible targeting sequences in phytochromes have been identified, including a series of four basic residues which correspond to those responsible for transport of nuclear-located proteins. Sites capable of post-translational modification have been found in monocot sequences, but not in dicot sequences. Areas of the phytochrome molecule which are exposed on the surface of the portein, and which are therefore capable of interaction with other cellular macromolecules, have been identified. Analogies with other biliproteins have been used to define minimum chromophore-protein interactions. Possible enzymic activities associated with phytochromes have been discussed with respect to local amino acid sequence similarity with enzymes.

Introduction

Although the control of gene transcription in plants by light is well documented, the mechanism and components of the signal transduction chain are largely unknown. The major photoreceptor responsible for detection of the primary stimulus is the chromoprotein phytochrome, which mediates a wide range of developmental processes. The initial event in phototransformation of the inactive form of phytochrome (Pr) to the active form (Pfr) is a rapid photoisomerization of the bilin chromophore [1]. A series of intermediates unique to the forward and backward reactions probably reflect conformational changes in the apoprotein. In the absence of fine structural information on the isoforms of phytochrome, physicochemical methods such as CD spectroscopy and the use of monoclonal antibodies as molecular probes have highlighted areas of the protein which may be important in phototransformation and in molecular recognition of the active form [2–6]. Following the isolation of c-DNA from *Avena sativa* and *Cucurbita pepo* [7, 8], these studies have been complemented by computer analysis of the derived amino acid sequence. The recent availability of new sequences for *Oryza sativa* [9], *Pisum sativum* [10] and *Arabidopsis thaliana* [11], and also of antibodies to known areas of the apoprotein [12, 14] provides an opportunity for verification and extension of these conclusions. We have examined these sequences with a variety of computer programs with a view to identifying areas of the phytochrome molecule which may be involved in the mode of action, and may serve as a focus for future research.

Methods

Secondary structure predictions by the Chou–Fasman [15] and Garnier–Osguthorpe–Robson [16] methods and surface probability calculations [17] were performed with the program PEPTIDESTRUCTURE (University of Wisconsin software). Pattern matching was carried out with the

Reprint requests to M. D. Partis.

Verlag der Zeitschrift für Naturforschung, D-7400 Tübingen
0341–0382/90/0900–0987 $ 01.30/0

program MATSCAN (University of Leeds software). Other analyses were performed with the following programs written in VAX-11 Pascal at IHR. Hydrophobic moments were calculated with a program MOMENT employing the algorithm of Eisenberg *et al.* [18]. Hydrophilicity and flexibility profiles were calculated with a program PREDICT using the methods of Hopp & Woods [19] and Karplus & Schulz [20]: this program was also used to identify possible surface structure. Detection of possible membrane-spanning sequences was carried out with a program MEMBRANE using the method of Eisenberg *et al.* [18]. Normalized consensus hydrophobicity values used in these programs were taken from [21]. Further information on the programs MOMENT, PREDICT and MEMBRANE is available from M.D.P. at IHR.

Table I. Comparison of secondary structure predictions with CD measurements.

	α-Helix	β-Sheet	β-Turn	Aperiodic
Avena sativa[1]	52.2	0	21.1	26.7
Avena sativa[2]	55.3	0	18.8	26.0
Avena sativa[3]	39.4	27.1	19.2	14.3
Avena sativa[4]	46.1	16.5	14.7	22.8
Cucurbita pepo[3]	40.8	29.2	17.4	12.6
Cucurbita pepo[4]	42.9	21.8	15.2	20.1
Oryza sativa[3]	39.3	27.3	18.6	14.8
Oryza sativa[4]	46.9	18.4	14.2	20.5
Pisum sativum[3]	39.8	28.5	18.7	13.1
Pisum sativum[4]	45.9	17.8	15.4	20.9

[1] Values calculated from CD measurements on Pr [5].
[2] Values calculated from CD measurements on Pfr [5].
[3] Values calculated by Chou-Fasman.
[4] Values calculated by Garnier-Osguthorpe-Robson.

Reliability of secondary structure predictions

In those cases where X-ray crystallographic data is available for proteins from a variety of species (globins, cytochrome c) or homologous proteins (trypsin-like serine proteases), conservation of conformation is the general rule, even when less than 40% of the amino acid residues are identical [22], and the overall shape of phytochromes should be conserved likewise. While the lengths of secondary structure elements such as α-helix or β-sheet may vary, the position of these within the tertiary structure should be retained. Surface structures are the most variable feature of homologous amino acid sequences, but at least some of these are conserved in phytochrome as judged by reactivity of monoclonal antibodies to phytochromes from a wide range of plant species [23, 24].

In the light of these expectations, how reliable are the secondary structure predictions from the known sequences? A guide to the accuracy of such predictions in the absence of X-ray crystallographic data can be given by calculation of secondary structure from CD spectra to a high degree of accuracy [25]. The theoretical proportions of different secondary structure elements for the four known sequences are compared with the calculated values from CD studies on *Avena* phytochrome in Table I. It is apparent that only β-turn predictions correspond to experimentally determined data, while α-helix is underestimated, and β-sheet is overestimated. Moreover, predictions for the four phytochrome sequences differ in many regions, with sheet replacing helix and *vice-versa;* a situation unknown in those homologous proteins whose structures have been so far studied [22]. Although the success of Chou–Fasman and Garnier–Osguthorpe–Robson predictions is only moderate (aproximately 50% overall, [69]), an additional contributory factor to the failure of the programs must be interactions between the chromophore and apoprotein. However, such interactions, while modifying the apoprotein structure, should not effect major changes in the proportions of structural elements (as distinct from the small conformational changes encountered in phototransformation, arising from chromophore-protein interactions).

Putative membrane-binding sequences

An appreciation of the limitations of the predictive methods is essential in the analysis of sequence data. For example, the assertion has been made [7] that there are no α-helices of sufficient length (21 residues) to span lipid bilayers. Since α-helical content is underestimated, only minimum lengths of helix can be identified, rendering the argument invalid without reference to other data. However, Eisenberg *et al.* [18] have successfully predicted membrane-spanning regions by identifying areas of net hydrophobicity; failures of analysis were only found for proteins containing mostly β-structure [21]. Application of this method to the

phytochrome sequence reveals no membrane penetration initiators of sufficient hydrophobicity to justify assignment of phytochrome as an integral membrane protein, either singly (<H> greater than 0.65) or in combination (<H> greater than 0.42, <H> combined greater than 1.1). This is not surprising, in that the bulk of phytochrome in the cell exhibits all of the properties of a soluble protein; however, there is a considerable body of evidence that at least some phytochrome-mediated processes occur at membranes, and that a proportion of phytochrome is associated with organelles [26].

It would seem likely, therefore, that any membrane-bound phytochrome would be a peripheral, rather than intrinsic, protein. The occurrence of amphiphilic regions is a common feature in proteins such as apolipoproteins, peptide toxins and hormones which interact with membrane surfaces [27] and in targeting presequences [28]. In this respect, a region near the C-terminus of phytochrome with a highly amphiphilic nature is of interest. Comparison of the hydrophobic moment and hydrophobicity of some membrane-associated peptides with this section of phytochrome is shown in Table II. The phytochrome peptide scores for both parameters show a striking resemblance to those for surface-associated amphiphilic peptides and mitochondrial presequences, demonstrating their suitability for interaction with organelles. However, it is not suggested that this portion of the protein functions as a classical mitochondrial targeting sequence; both the C-terminal location, and the lack of proteolytic processing [31] indicate that this is not the case.

Structurally important areas of the phytochrome protein should be conserved. In contrast to the rest of the C-terminal domain, conservation of sequence in the amphiphilic portion of the phytochrome molecule is good (Fig. 1), with the few amino acid substitutions being conservative. A consequent retention of the amphiphilic nature of the sequence is evident from the helical wheel and vector plots shown in Fig. 2, with hydrophobic residues such as leucine, isoleucine and valine on one face of the helix, and charged residues such as argi-

Table II. Amphiphilic and transmembrane peptides. Net hydrophobicity (<H>) per residue, and hydrophobic moment (<μH>) were calculated as described by Eisenberg [21].

Protein	Residues	<H>	<μH>
Surface-active			
Cecropin A	3–20	−0.22	0.80
Cecropin B	3–20	−0.23	0.77
δ-Hemolysin	9–26	0.08	0.70
Melittin	12–23	0.25	0.57
Mitochondrial presequences			
Cyt c peroxidase	1–18	−0.17	0.57
Cyt c oxidase IV	3–20	−0.20	0.69
Porin	2–19	−0.12	0.57
OTC	14–31	−0.31	0.56
Phytochromes			
Oryza	1093–1110	−0.36	0.67
Pisum	1087–1104	−0.21	0.64
Avena	1091–1108	−0.34	0.70
Cucurbita	1086–1103	−0.29	0.65
Arabidopsis A	1088–1105	−0.20	0.60
Arabidopsis B	1120–1137	−0.12	0.66
Arabidopsis C	1078–1095	−0.22	0.65
Transmembrane helices			
Bacteriorhodopsin	10–27	0.78	0.13
	42–59	0.67	0.14
	107–124	0.74	0.15
	135–152	0.82	0.04

```
                1081                                        1120
    Avena3      EEDNKEQSEE GLSLLVSRNL LRLMNGDVRH LREAGVSTFI
    Avena4      EEDNKEQSDE GLGLLVSRKL LRLMNGDVRH LREAGVSTFI
  Cucurbita     G.SEEDASEE GFSLLISRKL VKLMNGDVRY MREAGKSSFI
    Pisum       G.NNVLESEE GISLHISRKL LKLMNGDVRY LKEAGKSSFI
    Oryza       EDDNKEQSDE GMSLAVSRNL LRLMNGDVRH MREAGMSTFI
Arabidopsis A   G.TEEDVSEE GLSLMVSRKL VKLMNGDVQY LRQAGKSSFI
Arabidopsis B   H.SSRWTSPE GLGLSVCRKI LKLMNGEVQY IRESERSYFL
  Consensus     ..dnkeqSeE GlsLlvSRkL lrLMNGDVRy lreag.StFi
```

Fig. 1. Amphiphilic phytochrome sequences. Sequences were manually aligned for best fit. Conserved residues are indicated by upper case letters in the consensus sequence. The region of maximum hydrophobic moment is underlined on the consensus sequence.

nine and lysine on the other. Interestingly, the sequence for *Arabidopsis* phyC contains an additional glycine residue; amphiphilicity is retained, but the direction of the hydrophobic moment is rotated. An antibody raised against *Zea* phytochrome which binds to this region (Z−4A5, residues 1086−1106, [14]), also binds to phytochrome from the alga *Chaetomorpha* (P. Lindemann, personal

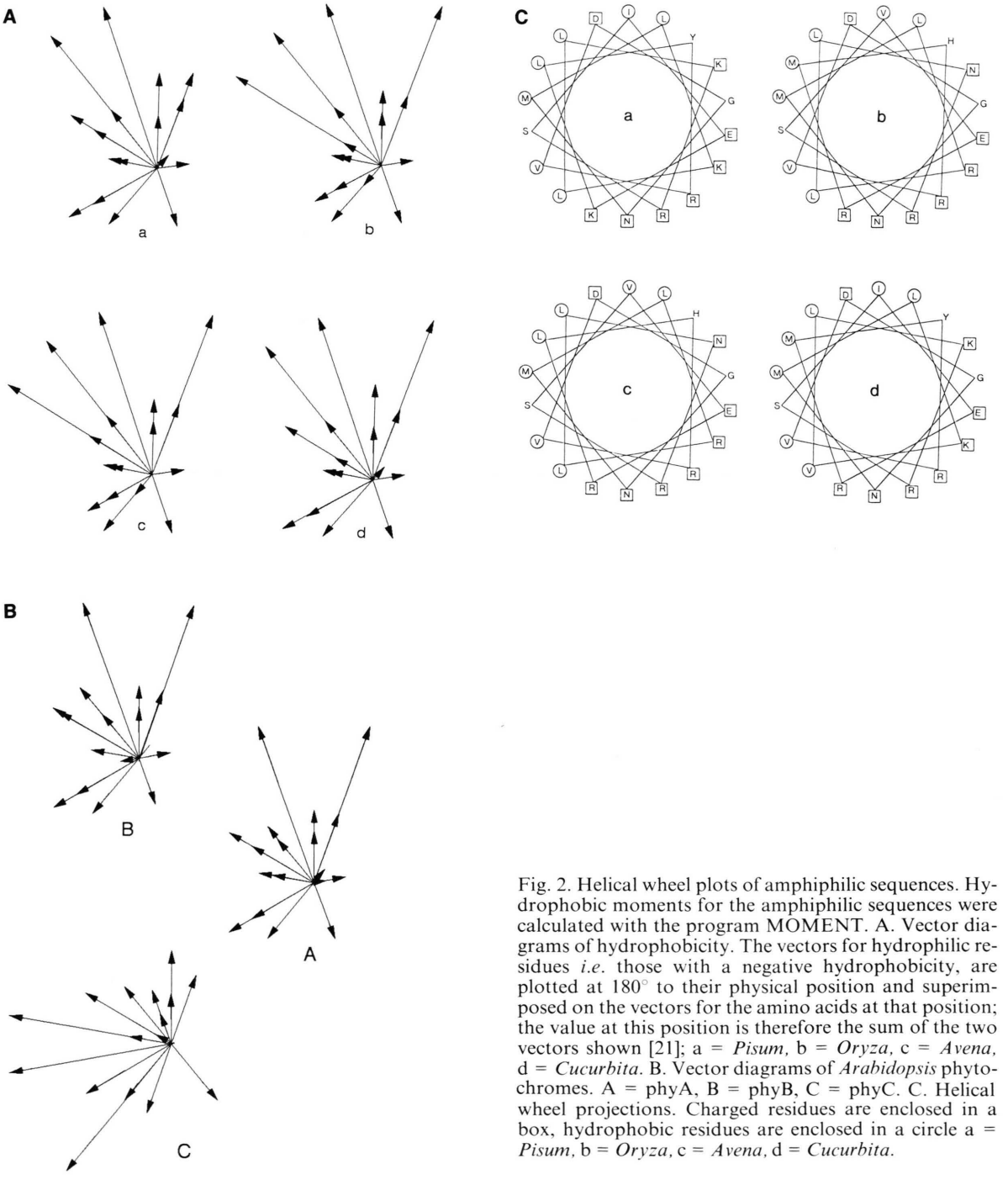

Fig. 2. Helical wheel plots of amphiphilic sequences. Hydrophobic moments for the amphiphilic sequences were calculated with the program MOMENT. A. Vector diagrams of hydrophobicity. The vectors for hydrophilic residues *i.e.* those with a negative hydrophobicity, are plotted at 180° to their physical position and superimposed on the vectors for the amino acids at that position; the value at this position is therefore the sum of the two vectors shown [21]; a = *Pisum*, b = *Oryza*, c = *Avena*, d = *Cucurbita*. B. Vector diagrams of *Arabidopsis* phytochromes. A = phyA, B = phyB, C = phyC. C. Helical wheel projections. Charged residues are enclosed in a box, hydrophobic residues are enclosed in a circle a = *Pisum*, b = *Oryza*, c = *Avena*, d = *Cucurbita*.

communication), indicating that conservation of this structure exists throughout the plant kingdom. This antibody shows preferential binding to the Pfr form [14], indicating that the region is more accessible in this conformer, and this may relate to the increased affinity of Pfr over Pr for membranes and liposomes [26].

Targeting sequences

The bulk of phytochrome is localized in the cytoplasm, where the Pr form appears to be distributed uniformly throughout the cytosol, whereas the Pfr form becomes sequestered within seconds after its formation from Pr. Nevertheless, one can speculate that a sub-population of phytochrome, or one of the different isoforms (*e.g.* those resulting from hexaploidy in *Avena* [7, 29], or those of a multi-gene family in *Arabidopsis* [11]) can reside in or on organelles, contributing to the diversity of the regulatory effects of the chromoprotein. Other factors which might influence the distribution of phytochrome between cytoplasm and membranes are the quantity and availability of interacting membrane proteins, and post-translational modification of phytochrome. The amino acid sequences which direct proteins to distinct cellular compartments have been extensively studied, although the mechanisms by which they operate are a matter of some debate [13].

Targeting to chloroplasts and mitochondria

A similarity between the N-terminal region of phytochrome, which is rich in serine, threonine and positively charged amino acids, and transit peptides of chloroplast proteins has been noted [5], and it has been suggested that this region may be involved in binding to membranes. The block structure which is a common feature of these peptides [13, 30] is not found in the phytochrome sequence, however, and the resemblance to such peptides is probably fortuitous. The number of acidic residues in this region would argue against a transit peptide role, as would the absence of a typical cleavage site [65]. Indeed, the phytochrome associated with membranes is identical in size to that in the cytosol [31], demonstrating that processing by a transit peptidase has not occurred. Additionally, partially degraded "large" phytochrome which has lost the N-terminus has been shown to retain the capacity to bind to detergents and liposomes [32, 66]. The *Arabidopsis* phyB protein probably has a 35-residue glycine-rich extension [11].

Essentially the same arguments can be made for and against targeting of phytochrome to mitochondria. Light-induced import of the chromoprotein into mitochondria has been described [67]. However, the lack of a strongly amphiphilic region at the N-terminus would suggest that phytochrome could not be transported by the MOM-19 route used for transfer of proteins destined for various mitochondrial compartments [13, 68], in which proteolytic processing of the signal peptide occurs. Cytochrome c, which is not transported in this way, relies on a high affinity to negatively charged phospholipid [13]. Phytochrome, however, shows a relatively low affinity for this kind of phospholipid [66].

Signal for directing proteins into microbodies

Microbody proteins lack cleavable topogenic sequences such as those found in mitochondrial and chloroplast precursor proteins, and are synthesized at their final sizes [33]. The information for targeting microbody proteins therefore resides in the amino acid sequence of the mature protein. It has been suggested that the tripeptide S – K/H – L is an essential element in the targeting sequence, which is located at the C-terminus [33, 34]. The tripeptide sequence is to be found at the carboxy terminus of two of the *Avena* phytochromes between residues 1038 and 1040. As the tripeptide is absent from *Oryza, Pisum* and *Cucurbita* phytochromes, and there is no evidence for location of phytochrome in microbodies, the significance of the occurrence of this feature in *Avena* phytochrome must be speculative.

Four basic residues in series

In most of the phytochrome sequences determined so far a conserved tetrapeptide composed of four basic residues is located C-terminal to the chromophore attachment site. The retention of this feature is all the more striking in that the surrounding amino acids show little homology. Any assignment of a function for this tetrapeptide must be speculative, but a common stretch of four basic residues is found within the internal topogenic sig-

nal sequence for nuclear location of proteins [33]. A proline adjacent to the first basic residue appears to be an important residue for this feature, and this is lacking in all of the phytochrome sequences. Not all nuclear-located proteins contain this proline residue. However, in those phytochrome sequences where the four basic residues are conserved, there is a preceding proline within six residues; where the basic residues are absent, this proline is also not present. Roberts [35] has reviewed signal-mediated protein transport into nuclei, and points out that the basic residues may also be flanked by amino acids with bulky side chains. This is true of all of the phytochrome sequences. Roberts [35] suggests that the function of these residues is to insulate the basic residue motif from the influence of surrounding sequence. In the Oat-4 *Avena* sequence the first basic residue is replaced by a glutamine residue, and in *Arabidopsis* phyB and phyC three of the basic residues are lost. If the series of basic residues is a targeting sequence, then this has important implications for the distribution of the phytochrome isoforms between organelles. Nagatani *et al.* [36] have discussed the evidence for phytochrome location in nuclei.

Potential modification of phytochrome

Although the amino acid sequences of phytochromes have been derived from c-DNAs, there is relatively little information from direct protein sequencing. This is particularly important with regard to identification of possible post-translational processing of the native proteins, which could affect either their tertiary structure and/or their potential activity in signal transduction. Hitherto, the only modification shown by direct sequencing is the bilin chromophore linked by a thioether bond to cysteine 321 [7, 37]. Additionally, phosphorylation of phytochrome is possible at three sites [38]. Consensus sequences which have been established as features for modification of amino acid residues can be found in monocot phytochromes. As these are absent in dicot phytochromes, the significance of these is uncertain. Differences between monocot and dicot phytochrome are not well documented, although there are some properties which are dissimilar, such as the dark reversion of dicot phytochrome *in vivo*.

N-Linkage of oligosaccharides

Proteins of both plant and animal cells contain N-linked as well as O-linked glycans that modify the physicochemical properties of the polypeptides. N-linkage occurs only on asparagine residues within the tripeptide signal sequence N – X – S/T located in peptide sequences which adopt the β-turn conformation [39]. The sequence N – A – S occurs in monocot phytochrome (*Avena* residues 806–808 and *Oryza* residues 808–810) at the beginning of a stretch of protein predicted to be β-turn. This feature is absent in the sequences of dicot phytochromes. However, a possible conserved N-linked glycosylation site exists at residues 372 in *Pisum*, 371 in *Cucurbita*, and 373 in *Arabidopsis* phyA (but not phyB or phyC). Evidence for the occurrence of carbohydrate in phytochrome is conflicting; Roux *et al.* [40] characterized *Avena* phytochrome as a glycoprotein with carbohydrate constituting about 4% of the protein by weight. In direct contrast is the report of Boeshore & Pratt [41] who were unable to detect carbohydrate in *Avena* phytochrome. Potential modification of asparagine-806 would increase the apparent molecular weight by about 1 kDa per monomer. Asparagine-linked glycans are classified as polymannoses, or as complex glycans, often containing 3 mannose, 2 N-acetylglucosamine, 1 xylose, and 1 fucose. A re-evaluation of the potential sugar content by gas-liquid chromatography has been reported (Keegstra & Quail in [42]); no sugar residue was present at greater than 0.3 residues/mole of monomer. Although this would argue that the bulk of phytochrome is not a glycoprotein, it is possible that a sub-population may be modified in this way.

Tyrosine sulfation

On the basis of analyses of the regions surrounding sulfation sites in proteins, Hortin *et al.* [43] have derived five rules for the prediction of such sites. All five criteria are met for the regions surrounding a tyrosine residue in all of the monocot phytochrome sequences, but this residue is replaced by a phenylalanine in *Pisum*, *Cucurbita* and *Arabidopsis* (Fig. 3). Interestingly, the tyrosine is located in a part of the native protein which is exposed in the Pr form [44, 45]. Little is known regarding tyrosine-sulfated proteins, although Hutt-

```
                1                                        ↓ 40
     Avena3     MSSSRPAS.. SSSSRNRQSS QARVLAQTTL DAELNAEYEE
     Avena4     MSSSRPAS.. SSSSRNRQSS RARVLAQTTL DAELNAEYEE
  Cucurbita     MSTSRPSQ.. SSSNSGRSRH STRIIAQTSV DANVQADFEE
     Pisum      MSTTRPSQ.. SSNNSGRSRN SARIIAQTTV DAKLHATFEE
     Oryza      MSSSRPTQCS SSSSRTRQSS RARILAQTTL DAELNAEYEE
Arabidopsis     MSGSRPTQ.. SSEGSRRSRH SARIIAQTTV DAKLHADFEE
  Consensus     MSssRP.q.. SSssr.Rqss .aRilAQTtl DAelnAe.EE

               41                                   70
     Avena3     SGDSFDYSKL VEAQRDGPPV QQGRSEKV.I
     Avena4     SGDSFDYSKL VEAQRDGPPV QQGRSEKV.I
  Cucurbita     SGNSFDYSSS VRVTSDVSGD QQPRSDKVTT
     Pisum      SGSSFDYSSS VRVSGSVDGD QQPRSNKVTT
     Oryza      YGDSFDYSKL VEAQRTTGPE QQARSEKV.I
Arabidopsis     SGSSFDYSTS VRVTGPVVEN QPPRSDKVTT
  Consensus     sGdSFDYSk. V..qrdvpp. QQpRSeKV.i
```

Fig. 3. N-terminal sequences of phytochromes. Sequences were aligned manually for best fit. Conserved residues are indicated by upper case letters in the consensus sequence. The arrow indicates a possible sulfation site in monocots.

ner [46] has speculated that this post-translational modification of protein may be widespread in plants.

Interaction with macromolecules

Identification of residues which are located at the surface of phytochrome is of importance not only because some of these residues may be involved in interactions with other cell components, but also because antibodies with a high probability of recognizing the native protein can be raised against peptides corresponding to these sites, and subsequently used as molecular probes. There are numerous methods for the prediction of surface location, based on charge [19], β-turn potential [47], flexibility [20], and combinations of these ([17]; PREDICT). Areas identified by these methods are shown in Table III. The list is not intended to be exhaustive; there are long stretches of sequence in which no surface-located region can be predicted with any accuracy *e.g.* residues 68–245.

As noted earlier, loops and turns on the surfaces of homologous proteins display greater variation in sequence than those located in the interior. This is reflected in the degree of conservation of the regions listed in Table III, where only 5 of the 15 segments are sequences which show high amino acid similarity. This does not exclude the possibility of structural similarity in the other 10 cases, and the retention of β-turn structure or hydrophilic nature indicates that amino acid substitutions in these regions are mostly conservative. The surface location of some of these regions is confirmed by the presence of protease-labile residues. In particular, cleavage at K-753 demonstrates that region 750–753 is accessible to protease, although this is one of the regions in Table III with a low predictive score.

In addition to computer methods, surface location of residues has been probed by susceptibility of phytochrome to proteolysis, chemical modification, and antibody binding. The identification of areas of phytochrome which change conformation on phototransformation has been aided by isolation of antibodies with differing affinities to Pr and Pfr [3, 6]. If these antibody probes interact with the native protein, the areas to which they bind must be located on the surface of the protein. Where such antibodies bind with greater affinity to one photoisomer than the other, a conformational change at the binding site probably occurs during phototransformation. As transmission of the light

Table III. Surface prediction of amino acid segments. B-turn score was predicted according to [15]. Hydrophilic score was calculated according to [19]. Flexibility was calculated by the method of [20]. Surface probability is according to [17]. All values were scored over a moving window of seven residues. The number of crosses denotes the magnitude of score. B-turn score: +, > 1.1; ++, > 1.2; ++, > 1.3; ++++, > 1.4. Hydrophilic score: +, > 0.5; ++, > 1.0; +++, > 1.5; ++++, > 2.0. Flexibility: +, > 1.0; ++, > 1.05; +++, > 1.1; ++++, > 1.15. Surface probability: +, > 10; ++, > 20; +++, > 30; ++++, > 50.

Residues	β-turn score	Hydrophilic score	Flexibility	Surface	Conserved	Cleaved
1–19	++++	++	++++	+	N	
29–68	+++	++	+++	++	N	K-65
245–251	++	+++	+	+	Y	
347–366	++	++++	+++	++++	N	E-354
535–555	++	++++	+++	+++	Y	
589–614	++	++	++	++	N	E-597
698–709	+++	+++	++	++	N	K-698
718–724	+	++	++	+	N	
750–753	–	+	++	+	Y	K-753
788–794	+	+++	+	+	Y	
811–819	+	++	++	++	N	
921–928	+	++	++	+++	N	
952–964	++	++	++	+	N	
1080–1087	+	+++	+++	+++	Y	
1104–1108	+	++	+	+	N	

1. Residue numbers refer to *Avena* sequence.

stimulus must involve recognition of the active Pfr form at the molecular level, these regions are candidates for an "active site". One such region is in the N-terminal 10 kDa, where antibodies have been raised to the first 18 amino acid residues [12], and to at least one other site [48, 50] between 6 and 10 kDa from the N-terminus. Both of these antibodies, however, interact more strongly with the inactive Pr form. That the site responsible for signal transduction is conserved (at least in structure), is shown by phenotypic changes induced by *Avena* phytochrome in transgenic tobacco [49].

Two monoclonal antibodies have been shown to exhibit wide cross-reactivity. One of these (Z-3 B 1, [24]) binds to a region between residues 210 and 426 [14]; two of the segments listed in Table III could contain the binding site. Of these, only one (residues 245–251) shows conservation of primary sequence. However, since the antibody shows preferential binding to the Pfr form, and Pfr-specific cleavage by endopeptidase Glu-C occurs at residue E-354, binding to the region 347–366 cannot be excluded. Discrimination between the two possibilities will require finer mapping of the binding site. Antibodies Oat-9 and Oat-16, which interact preferentially with Pfr, have also been shown to bind to epitopes located between residues 200 and 450 [50].

The second antibody displaying good cross-reactivity is Pea-25 [23]. The binding site for this antibody lies between residues 765 and 771 [51], and as such does not correspond to one of the predicted surface regions in Table III. Conservation of binding was screened by immunoblotting following SDS-gel electrophoresis of phytochrome from a wide variety of species, and therefore is to partially denatured protein. Phytochrome in solution has been shown to exist as a dimer, and the site of self-association has been shown to be located in the C-terminal 40 kDa [52]. Conservation of sequence is low in this area, and only two blocks of 20–30 amino acids show homology. One of these is the amphiphilic region previously described, and the other lies between residues 765 and 800. In this latter region, only one segment around an invariant arginine at 788–790 is very hydrophilic, indicating its probable location on the surface of the monomer. The close proximity of this site to the

highly conserved region identified by Thompson *et al.* [51] indicates that this may be a structurally important feature.

Chromophore-protein interaction

Some clues to the amino acid residues involved in protein-chromophore interaction can be obtained by comparison with other biliproteins. The cyanobacterial electron-transfer protein C-phycocyanin contains a chromophore which differs from phytochromobilin by only a few atoms. The chromophore-protein interaction in this protein has been determined by X-ray crystallography [53]. The sequences around the cysteine at which the bilin chromophores are bound show only a few points of identity between the biliproteins, and these are probably significant (Fig. 4). Of particular interest are the invariant arginine lying five residues N-terminal to the cysteine; by analogy to phycocyanin, this is in a position to form hydrogen bonds with the propionate of ring B on the phytochrome chromophore. In phycocyanin, the aspartate two residues C-terminal to the cysteine interacts with the nitrogen of rings B and C; this position is occupied in phytochrome by an invariant glutamine which is capable of fulfilling the same function. Lastly, the serine/threonine at twelve residues N-terminal to the cysteine is in a suitable position to interact with the nitrogen of ring A.

A broad localization of the other regions of the apoprotein which interact with the chromophore can be achieved by reference to the spectra and phototransformation characteristics of fragments of phytochrome produced by proteolysis. Pre-eminent among these is a 118 kDa polypeptide, generated by preferential cleavage of the N-terminus in the Pr form, and which exhibits decreased absorbance in the far-red and an increased rate of dark reversion from Pfr to Pr. Antibodies to this region show higher affinity for the Pr form over the Pfr form, and also increase the rate of dark reversion. Chromophore oxidation studies suggest that in the Pr form, this region is more exposed, but is in contact with the chromophore in the Pfr form [70, 71]. Anti-phycocyanin antibodies have been shown to cross-react with *Avena* phytochrome [54], and the binding site for these has been also been localized in the N-terminus. There is little conservation of sequence (Fig. 3) in this region, and no discernable similarity with the phycocyanin sequence. As only one or two residues would be involved in chromo-

```
                 301                          ↓          335
        Avena3   KVIEAEALPF DISLCGSALR APHSCHLQYM ENMNS
        Avena4   KVIEAEALPF DISLCGSALR APHSCHLQYM ENMNS
     Cucurbita   KVLQDEKLQF DLTLCGSTLR APHSCHLQYM ENMNS
        Pisum    KVLQDEKLPF DLTLCGSTLR APHSCHLQYM ANMDS
        Oryza    KIIEDESLHL DISLCGSTLR APHSCHLQYM ENMNS
  Arabidopsis A  RVLQDEKLSF DLTLCGSTLR APHSCHLQYM ANMDS
  Arabidopsis B  LVVQDDRLTQ SMCLVGSTLR APHGCHSQYM ANMGS
  Arabidopsis C  KVVQDKSLSQ PISLSGSTLR APHGCHAQYM SNMGS
      Consensus  kviqdekLpf disLcGStLR APHsCHlQYM eNMnS

                          * *   * * *  **
C-Phycocyanin α           PNYAADAR GKSKCARDIG HYL
               β          S..A...R RMAACLRDM
                                   *      *

Allophycocyanin                 TR RYAACIRDLD YYL
Phycoerythrocyanin             FHHR NQAACIRDLG FIL
```

Fig. 4. Chromophore-binding region of biliproteins. Sequences were aligned on the cysteine residue to which the bilin chromophore is bound, which is marked with an arrow. Asterisks denote the residues identified in phycocyanin chromophore-apoprotein interactions by X-ray crystallography [53]. Phycocyanin, allophycocyanin and phycoerythrocyanin sequences are all for *Mastigocladus laminosus*.

	874							882
	↓	↓			↓			
Avena 3	A	S	H	E	L	Q	H	A L
Avena 4	A	S	H	E	L	Q	H	A L
Oryza	P	S	H	E	L	Q	H	A L
Cucurbita	P	S	H	E	L	Q	Q	A L
Pisum	A	S	P	E	L	Q	Q	A L
Arabidopsis A	A	S	H	E	L	Q	Q	A L
Thermolysin	V	G	H	E	L	T	H	A V
Stromelysin	A	A	H	E	L	G	H	S L

Fig. 5. Conserved consensus sequence of zinc-dependent peptidases. Arrows indicate the signature H E X X H.

phore interaction, conservation of the appropriate conformation is probably more important than identity of sequence. It is significant that this N-terminal region in all four of the sequences shows a strong tendency to adopt a β-turn structure. An analysis of CD spectra of Pr and Pfr has led Song [5] to propose that a β-turn to α-helix transformation in the N-terminus follows photoisomerization of the chromophore.

All of the amino acid sequence required for phototransformation lie in the region bounded by residues 200 and 400 [55]. This portion of the protein is well conserved, with the exception of residues 346–366, where amino acid insertions in both monocots and dicots lead to low homology. This region is noteworthy in being the most hydrophilic segment in all four sequences, and as such should be a surface-located loop in which sequence variation might be expected to occur. This is supported by the high scores for chain flexibility [20] and surface probability [17] of these residues. Grimm *et al.* [45] find that Pfr-specific proteolysis by endoproteinase Glu-C occurs at E-354 (E-352 in [72]), and suggest that this site is exposed only in the Pfr form. An interior location for this residue is of low probability given the extremely hydrophilic nature of the segment, but the flexibility of the region makes it a good candidate for a conformational change allowing the antibody to bind. There are several hydrophobic segments in the region 200–400 which should be buried within the protein, and which have the potential of forming a hydrophobic cleft with which the chromophore may be associated. The two major regions are residues 333–344 lying between the site of covalent attachment of chromophore and the hydrophilic loop, and residues 308–316 which lie N-terminal to be chromophore site.

Phytochrome as an enzyme

The hypothesis that phytochrome participates in signal transduction *via* an intrinsic enzymic activity has a venerable history. Apparent support for this hypothesis was provided most recently by the results of Wong *et al.* [38], who found protein kinase activity in highly purified phytochrome preparations. Lagarias *et al.* [56] identified some local homology between *Avena* phytochrome (residues 403–415) and ten protein kinases. Extensive homology with protein kinases is not found, nor do the phytochrome sequences contain the consensus sequences generally employed as signatures for kinase activity [57]. Moreover, Grimm *et al.* [57] were able to separate the kinase activity from phytochrome by electrophoresis on non-denaturing gels.

Jongeneel *et al.* [58] have screened the Swiss-Prot database for the occurrence of the H E X X H motif which is an unique signature of zinc-dependent metallo-peptidases such as thermolysin [59] or stromelysin [60]. In order to refine the alignment, they have redefined the putative zinc signature as (uncharged) – (uncharged) – H – E – (uncharged) – (uncharged) – H – (uncharged) – (hydrophobic). Surprisingly, they found a matching sequence in *Avena* phytochrome, but this was absent in *Cucurbita* phytochrome. Furthermore, had the sequences for other phytochromes been in the database, a similar sequence would only have been found for *Oryza* but not for *Pisum* or *Arabidopsis;* whether binding of zinc ions to monocot phytochrome elicits protease activity remains to be investigated.

Potential cleavage points: pairs of basic residues

It is generally assumed that the active form of phytochrome comprises the entire molecule. The susceptibility of phytochromes to proteolytic degradation is well documented; cleavage with endogenous and a variety of specific proteases has been reported in structure/function studies [44, 45] and

in epitope mapping [24, 61–63]. Several cleavage points are more accessible to attack in one or other of the photoisomers. In animal cells, pairs of basic residues are preferentially cleaved in precursor proteins to generate a variety of neuropeptides, *e.g.* the precursor pro-opiomelanocortin is cleaved into corticotropin, β-lipotropin, γ-lipotrophin, β-endorphin and enkephalin [64]. In addition to the four basic residues mentioned, the phytochrome sequences contain one conserved pair of basic residues around position 546. Furthermore, there are 4–5 pairs of basic residues distributed on the carboxy-terminal half of the protein. The majority of these residue pairs are conserved. Although peptide hormones are unknown in the plant kingdom, comparison of sequences of animal origin and of known significance, can generate novel prospects for investigation of the mechanism of signal transduction.

Conclusion

Computer analysis of amino acid sequence often raises as many questions as it answers. It does however, indicate areas of interest and generates ideas for future experimentation. Many of the views expressed here are speculative, but should be experimentally verifiable. In particular, site-directed antibodies to possible dimerization sites, putative membrane-active segments and "active sites" should prove to be valuable probes of phytochrome structure and function. The expression of functional *Avena* phytochrome in transgenic tobacco [54] opens up the possibility of site-directed mutagenesis to investigate these areas.

Acknowledgements

R. G. is supported by Deutsche Forschungsgesellschaft (SFB 206).

[1] W. Rudiger, Phytochrome: The chromophore, in: Photomorphogenesis of Plants (R. E. Kendrick and G. H. M. Kronenberg, eds.), pp. 17–34, Nihoff/Junk, Dordrecht, The Netherlands 1986.
[2] B. R. Jordan, M. D. Partis, and B. Thomas, The biology and molecular biology of phytochrome. Oxford surveys of plant molecular and cell biology **3,** 315–362 (1986).
[3] B. Thomas, M. D. Partis, and B. R. Jordan, Immunological approaches to the study of phytochrome, in: Immunology in Plant Science (T. L. Wang, ed.), pp. 171–195, Cambridge University Press, Cambridge 1986.
[4] R. D. Vierstra and P. H. Quail, Phytochrome: the protein, in: Photomorphogenesis in Plants (R. E. Kendrick and G. H. M. Kronenberg, eds.), pp. 35–60, Nihoff/Junk, Dordrecht, The Netherlands 1986.
[5] P.-S. Song, J. Photochem. Photobiol. **2,** 43–57 (1988).
[6] M. D. Partis and B. Thomas, Phytochromes, in: Methods in Plant Biochemistry (L. Rogers, ed.), **Vol. 5,** pp. 235–259, Academic Press, London 1990.
[7] H. P. Hershey, R. F. Barker, K. B. Idler, J. L. Lissemore, and P. H. Quail, Avena. Nuc. Acids Res. **13,** 8543–8559 (1985).
[8] R. A. Sharrock, J. L. Lissemore, and P. H. Quail, Gene **47,** 287–295 (1986).
[9] S. A. Kay, B. Keith, K. Shinozaki, and N.-H. Chua, Nuc. Acids Res. **17,** 2865–2866 (1989).
[10] N. Sato, Plant Mol. Biol. **11,** 697–710 (1988).
[11] R. A. Sharrock and P. H. Quail, Gene Develop. **3,** 1745–1757 (1989).
[12] C. Plumpton, M. D. Partis, B. Thomas, N. S. Huskisson, and G. W. Butcher, Plant Cell Environ. **11,** 487–491 (1988).
[13] G. von Heijne, Biochim. Biophys. Acta **947,** 307–333 (1988).
[14] H. A. W. Schneider-Poetsch, B. Braun, and W. Rudiger, Planta **177,** 511–514 (1989).
[15] P. Y. Chou and G. D. Fasman, Ann. Rev. Biochem. **47,** 251–276 (1978).
[16] J. Garnier, D. J. Osguthorpe, and B. Robson, J. Mol. Biol. **120,** 97–120 (1978).
[17] E. A. Emini, J. V. Hughes, D. S. Perlow, and J. Boger, J. Virol. **55,** 836–839 (1985).
[18] D. Eisenberg, E. Schwarz, M. Komaromy, and R. Wall, J. Mol. Biol. **179,** 125–142 (1984).
[19] T. P. Hopp and K. R. Woods, Proc. Natl. Acad. Sci. U.S.A. **78,** 3824–3828 (1981).
[20] P. A. Karplus and G. E. Schulz, Naturwissenschaften **72,** 212–213 (1985).
[21] D. Eisenberg, Ann. Rev. Biochem. **53,** 595–623 (1984).
[22] T. E. Creighton, Proteins; structures and molecular properties, Freeman & Co, New York 1983.
[23] M.-M. Cordonnier, H. Greppin, and L. H. Pratt, Plant Physiol. **80,** 982–987 (1986).
[24] H. A. W. Schneider-Poetsch, H. Schwarz, R. Grimm, and W. Rudiger, Planta **173,** 61–72 (1988).
[25] W. C. Johnson, Ann. Rev. Biophys., Biophys. Chem. **17,** 145–166 (1988).
[26] S. J. Roux, Phytochrome; phytochrome and membranes, in: Photomorphogenesis in Plants (R. E. Kendrick and G. H. M. Kronenberg, eds.), pp. 115–134, Nihoff/Junk, Dordrecht, The Netherlands 1986.
[27] E. T. Kaiser and F. J. Kezdy, Proc. Natl. Acad. Sci. U.S.A. **80,** 1137–1143 (1983).
[28] G. von Heijne, EMBO J. **5,** 1335–1342 (1986).

[29] R. Grimm, F. Lottspeich, and W. Rudiger, FEBS Lett. **225,** 215–217 (1987).
[30] G. A. Karlin-Neumann and E. M. Tobin, EMBO J. **5,** 9–13 (1986); Chem. **17,** 145–166.
[31] B. R. Jordan, M. D. Partis, and B. Thomas, Physiol. Plant. **60,** 416–421 (1984).
[32] I.-S. Kim, E. S. Kim, and P.-S. Song, Biochim. Biophys. Acta **747,** 55–64 (1983).
[33] P. Borst, Biochim. Biophys. Acta **866,** 179–203 (1986).
[34] P. G. Bruinenberg, M. Evers, H. R. Waterham, J. Kuipers, A. C. Arnberg, and A. B. Geert, Biochim. Biophys. Acta. **1008,** 157–167 (1989).
[35] B. Roberts, Biochim. Biophys. Acta **1008,** 263–280 (1989).
[36] A. Nagatani, G. I. Jenkins, and M. Furuya, Plant Cell Physiol. **29,** 1141–1145 (1988).
[37] J. C. Lagarias and H. Rapoport, J. Am. Chem. Soc. **102,** 4821–4828 (1980).
[38] Y.-S. Wong, H. C. Cheng, D. A. Walsh, and J. C. Lagarias, J. Biol. Chem. **261,** 12089–12097 (1986).
[39] R. Kornfeld and S. Kornfeld, Ann. Rev. Biochem. **54,** 631–664 (1985).
[40] S. J. Roux, S. G. Linsansky, and B. M. Stoker, Physiol. Plant. **35,** 85–90 (1975).
[41] M. L. Boeshore and L. H. Pratt, Plant Physiol. **68,** 789–797 (1981).
[42] L. H. Pratt, Ann. Rev. Plant Physiol. **33,** 557–582 (1982).
[43] G. Hortin, R. Folz, J. I. Gordon, and A. W. Strauss, Biochem. Biophys. Res. Commun. **141,** 326–333 (1986).
[44] J. C. Lagarias and F. M. Mercurio, J. Biol. Chem. **260,** 2415–2423 (1985).
[45] R. Grimm, C. Eckerskorn, F. Lottspeich, C. Zenger, and W. Rudiger, Planta **174,** 396–401 (1988).
[46] W. B. Huttner, TIBS **12,** 361–363 (1987).
[47] V. Krchnak, O. Mach, and A. Maly, Anal. Biochem. **165,** 200–207 (1987).
[48] M. Holdsworth and G. C. Whitelam, Planta **172,** 539–547 (1987).
[49] J. M. Keller, J. Shanklin, R. D. Vierstra, and H. P. Hershey, EMBO J. **8,** 1005–1012 (1989).
[50] L. H. Pratt, M.-M. Cordonnier, and J. C. Lagarias, Arch. Biochem. Biophys. **267,** 723–735 (1988).
[51] L. K. Thompson, L. H. Pratt, M.-M. Cordonnier, S. Kadwell, J.-L. Darlix, and L. Crossland, J. Biol. Chem. **264,** 12426–12431 (1989).
[52] A. M. Jones and P. H. Quail, Biochemistry **25,** 2987–2995 (1986).
[53] T. Schirmer, W. Bode, and R. Huber, J. Mol. Biol. **196,** 677–695 (1987).
[54] D. R. Keiller, G. C. Whitelam, and H. Smith, Planta **176,** 391–398 (1988).
[55] A. M. Jones and P. H. Quail, Planta **178,** 147–156 (1989).
[56] J. C. Lagarias, Y.-S. Wong, T. R. Bekelman, D. G. Kidd, and R. W. McMichael Jr., Structure-function studies on *Avena* phytochrome, in: Phytochrome and photoregulation in plants (M. Furuya, ed.), pp. 51–61, Academic Press, Tokyo, New York 1987.
[57] R. Grimm, D. Gast, and W. Rudiger, Planta **174,** 199–206 (1989).
[58] C. U. Jongeneel, J. Bouvier, and A. Bairoch, FEBS Lett. **242,** 211–214 (1989).
[59] K. Titani, M. A. Hermodson, L. H. Ericsson, K. A. Walsh, and H. Neurath, Nature New Biol. **238,** 35–37 (1972).
[60] R. Breathnath, L. M. Matrisian, M. C. Gesnel, A. Staub, and P. Leroy, Nuc. Acid. Res. **15,** 1139–1152 (1987).
[61] S. M. Daniels and P. H. Quail, Plant Physiol. **76,** 622–626 (1984).
[62] R. D. Vierstra, M. M. Cordonnier, L. H. Pratt, and P. H. Quail, Planta **160,** 521–528 (1984).
[63] R. Grimm, F. Lottspeich, H. A. W. Schneider, and W. Rudiger, Z. Naturforsch. **41c,** 993–1000 (1986).
[64] F. E. Bloom, Neuropeptide – Botenstoffe in Gehirn und Körper, Scientific American German edition **1981,** 72–83.
[65] Y. Gavel and G. von Heijne, FEBS Lett. **261,** 455–458 (1990).
[66] B. R. Singh, J.-K. Choi, I.-S. Kim, and P.-S. Song, Physiol. Plant. **76,** 319–325 (1989).
[67] B. S. Serlin and S. J. Roux, Biochim. Biophys. Acta **848,** 372–377 (1986).
[68] T. Sollner, G. Griffiths, R. Pfaller, N. Pfanner, and W. Neupert, Cell. **59,** 1061–1070 (1989).
[69] G. E. Schulz, Ann. Rev. Biophys, Biophys. Chem. **17,** 1–21 (1988).
[70] T.-R. Hahn, P.-S. Song, P. H. Quail, and R. D. Vierstra, Plant Physiol. **74,** 755–758 (1984).
[71] F. Thummler, P. Eilfeld, W. Rudiger, D.-K. Moon, and P.-S. Song, Z. Naturforsch. **40c,** 215–218 (1985).
[72] R. Schendel, Z. Tong, and W. Rudiger, Z. Naturforsch. **44c,** 757–764 (1989).