

---

**Analysis of cloned cDNA and genomic sequences for phytochrome: complete amino acid sequences for two gene products expressed in etiolated *Avena***

---

Howard P.Hershey<sup>+</sup>, Richard F.Barker<sup>1\*</sup>, Kenneth B.Idler<sup>1</sup>, James L.Lissemore and Peter H.Quail

---

Department of Botany, University of Wisconsin, Madison, WI 53706 and <sup>1</sup>Agrigenetics Advanced Research Laboratory, Madison, WI 53716, USA

---

Received 5 September 1985; Revised 25 October 1985; Accepted 29 October 1985

---

**ABSTRACT**

Cloned cDNA and genomic sequences have been analyzed to deduce the amino acid sequence of phytochrome from etiolated *Avena*. Restriction endonuclease site polymorphism between clones indicates that at least four phytochrome genes are expressed in this tissue. Sequence analysis of two complete and one partial coding region shows ~98% homology at both the nucleotide and amino acid levels, with the majority of amino acid changes being conservative. High sequence homology is also found in the 5'-untranslated region but significant divergence occurs in the 3'-untranslated region. The phytochrome polypeptides are 1128 amino acid residues long corresponding to a molecular mass of 125 kdaltons. The known protein sequence at the chromophore attachment site occurs only once in the polypeptide, establishing that phytochrome has a single chromophore per monomer covalently linked to Cys-321. Computer analyses of the amino acid sequences have provided predictions regarding a number of structural features of the phytochrome molecule.

**INTRODUCTION**

Phytochrome is the best-characterized of the regulatory photoreceptors that control plant development in response to light (1). The native molecule is a dimer, with each subunit composed of a linear tetrapyrrole chromophore covalently linked to a polypeptide that varies from a Mr of 120,000 to 127,000 among various plant species (2,3,4). The photoreceptor has two forms that are reversibly interconvertible by light: the Pr form that absorbs maximally in the red ( $\lambda_{max} = 666$  nm) region of the spectrum and the Pfr form that absorbs maximally in the far-red ( $\lambda_{max} = 730$  nm) region. Photoconversion of Pr to Pfr *in vivo* induces an array of morphogenic responses, whereas reconversion of Pfr to Pr cancels the induction of those responses. It is this property of photointerconvertibility that allows phytochrome to function as a regulatory molecule, with Pr considered the inactive form and Pfr the active form.

Despite considerable research effort, the primary molecular mechanism by which Pfr induces the observed responses remains unknown. Recent

evidence has substantiated the prevailing expectation that the morphogenic changes induced by Pfr result from phytochrome-regulated gene expression. The photoreceptor has been shown to control the expression of a number of nuclear genes including those encoding the small subunit of ribulose-bisphosphate carboxylase (5,6,7), chlorophyll a/b binding protein (5,6,7), protochlorophyllide reductase (8), rRNA (9) and several as yet unidentified gene products (6). For those genes thus far examined, changes in expression have been shown to occur at the level of transcription (10,11,12). Moreover, it has been found that phytochrome also controls the expression of its own gene(s) in a negative feedback fashion (13,14). The phytochrome system therefore simultaneously allows study of the mechanism of action of phytochrome itself and offers an attractive model for the study of phytochrome-regulated nuclear gene expression.

As part of our approach to understanding the mechanism of phytochrome action we have been examining various structural properties of the molecule (2,4,15,16). Here we report the complete amino acid sequence of Avena phytochrome deduced from analysis of cloned cDNA and genomic sequences. We present evidence that at least four different phytochrome genes are expressed in Avena and discuss some of the predicted structural features of the protein obtained from computer analysis of the sequence data.

### MATERIALS AND METHODS

#### Isolation of phytochrome cDNA clones

The isolation and characterization of phytochrome cDNA clones pAP-2 and pAP-3 (redefined here as pAP-3.1 and pAP-3.2 respectively) have been described in detail elsewhere (17). Additional cDNA libraries were prepared from mRNA enriched for phytochrome sequences as described before (17,18). A 3.0 kbp KpnI/SacI fragment of the pAP-3.2 insert (Fig. 1) was excised by restriction endonuclease digestion, isolated by preparative agarose gel electrophoresis and <sup>32</sup>P-labeled to >10<sup>8</sup> cpm/μg by nick translation (19). Additional phytochrome cDNA clones were identified by screening the cDNA libraries with this labeled phytochrome-specific probe at high stringency as described by Hanahan and Meselson (20). Plasmid DNA was isolated from cultures using the alkaline lysis method (21) and purified as described (17).

#### Isolation of genomic clones

High molecular weight Avena DNA was subjected to a non-limit digest with EcoRI and fragments of suitable length were cloned into λCh 32

essentially as described by Murray et al. (22). A  $1.5 \times 10^6$  pfu library was screened for recombinants containing phytochrome sequences by in situ plaque hybridization (23) using the same 3.0 kbp cDNA fragment used for cDNA library screening.

#### DNA sequence analysis

Sequences were determined by the method of Maxam and Gilbert (24) with modifications described by Barker et al. (25). Computer analysis of DNA sequences were performed by using programs made available by Drs. O. Smithies and F. Blattner (University of Wisconsin, Madison WI USA).

#### Computer analysis of derived amino acid sequences

Searches for protein sequences with similarities to phytochrome in the National Biomedical Research Foundation (NBRF) database were conducted using the programs developed by Lipman and Pearson (26) and by the University of Wisconsin Genetics Computer Group (UWGCG) (27). Hydropathy analysis was performed using programs supplied by J. Pustell, Harvard University. Chou-Fassman (28) secondary structure analysis was carried out using UWGCG programs. Secondary structure analysis according to Garnier et al. (29), amino acid composition tabulation, charged residue distribution profiles, isoelectric point estimation and antigenic site prediction according to Hopp and Woods (30) were kindly performed by F. Blattner (University of Wisconsin, Madison) using programs developed by DNASTAR, Madison WI USA.

### RESULTS

#### cDNA clone isolation and characterization

Analysis of previously isolated phytochrome cDNA clones showed them to represent >80% of the mRNA sequence based upon the length of the mature phytochrome mRNA (17). Additional cDNA libraries were constructed with the objective of isolating those cDNA clones required to provide the remaining sequence. Approximately 125,000 cDNA transformants from these libraries were screened with a nick translated 3.0 kbp insert derived from the phytochrome cDNA clone pAP-3.2. This screening yielded 47 positive colonies containing inserts ranging in size from 1.1 kbp to 3.4 kbp. This result indicated that no individual clone represented full length phytochrome sequence.

The inserts of plasmids from all the positive clones were mapped by digestion with a variety of restriction endonucleases. While the inserts showed strong similarity to one another and to the original phytochrome

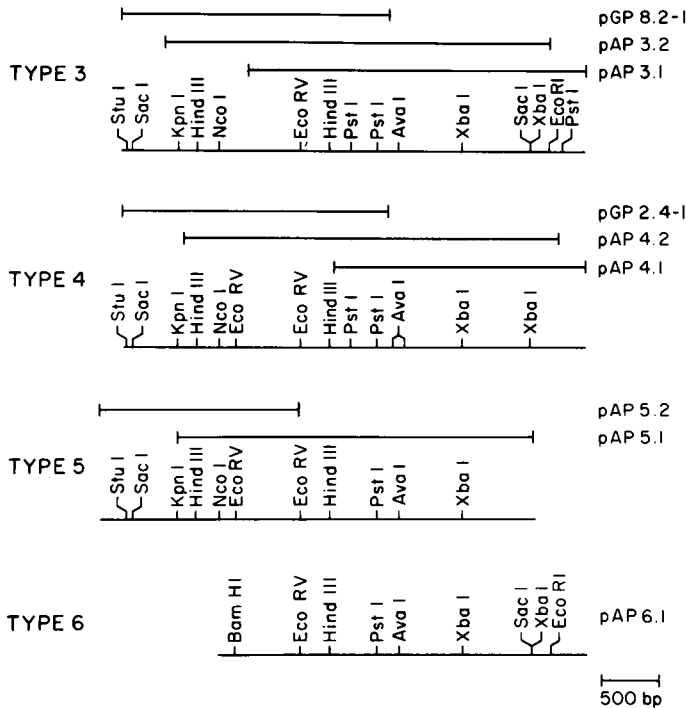


Figure 1. Composite restriction maps of cDNA and genomic clones representing four different expressed genes found in etiolated *Avena*. The overlapping cDNA clones (pAP-3.1, pAP-3.2, pAP-4.1, pAP-4.2, pAP-5.1, pAP-5.2, pAP-6) and/or subcloned genomic fragments (pGP8.2-1, pGP2.4-1) used to construct each restriction map for the various phytochrome types are shown above each map (designated as Types 3,4,5 or 6). The clone initially designated Type 1 has been insufficiently characterized for inclusion here and the clone initially designated as Type 2 has since been shown to have identical phytochrome sequence to Type 3 (17).

clone pAP-3.2 (17), restriction site polymorphisms among the cDNAs allowed the identification of four distinct classes of phytochrome clones. The overlapping clones that provide the longest total sequence for each of these four types are summarized in Fig. 1. In addition to the previously characterized cDNA clones pAP-3.1 and pAP-3.2 (17), two overlapping cDNA clones of each class of phytochrome provide a total of 3.5 kbp of type 4 phytochrome sequence, 3.35 kbp of type 5 sequence, and a single cDNA clone provides 2.9 kbp of type 6 sequence.

Genomic clone isolation and DNA sequence analysis

DNA sequence analysis was performed on the cDNA clones that comprise sequences for phytochrome types 3,4 and 5. Clone pAP-5.2 was the only

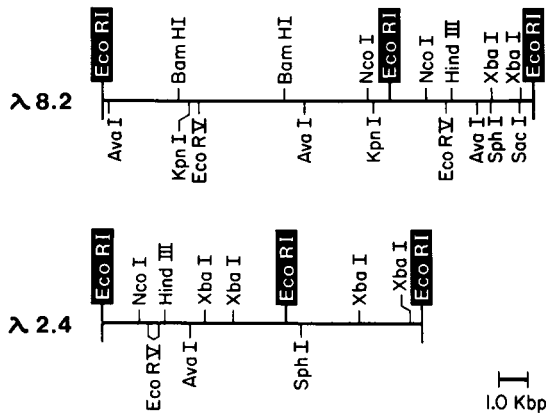


Figure 2. Restriction maps for the cloned *Avena* genomic fragments  $\lambda$ 8.2 and  $\lambda$ 2.4.

cDNA clone found to contain uninterrupted sequence representing both the 5' untranslated region and coding sequence for the NH<sub>2</sub>-terminus of the protein. NH<sub>2</sub>-terminal coding sequences shown for clone types 3 and 4 were therefore derived from sequence analysis of genomic clones.

A number of genomic clones were isolated from an *Avena* library that was screened with the same nick translated 3.0 kbp phytochrome cDNA fragment from pAP-3.2 used to screen the cDNA libraries. Restriction endonuclease analysis of these clones showed one, designated  $\lambda$ 8.2, to be a likely candidate to contain the type 3 phytochrome gene and  $\lambda$ 2.4 to be a strong candidate to contain the type 4 gene (Fig. 2). Southern blots of Eco RI digests of these two clones probed with <sup>32</sup>P-labeled pAP-3.2 showed only the 5.2 kbp Eco RI fragment of  $\lambda$ 8.2 (Hershey et al. 1985) and the 6.6 kbp fragment of  $\lambda$ 2.4 (data not shown) to be homologous to our phytochrome cDNA sequences.

The Eco RI fragment from each genomic clone showing homology to phytochrome cDNA clones was subcloned into pBR322 to yield the plasmids pGP8.2-1 containing the 5.2 kbp Eco RI fragment from  $\lambda$ 8.2 and pGP2.4-1 containing the 6.6 kbp Eco RI fragment from  $\lambda$ 2.4. Sequence analysis of these subcloned fragments confirmed the hypothesis that  $\lambda$ 8.2 contained at least part of the gene for type 3 phytochrome and  $\lambda$ 2.4 contained at least part of the gene for type 4 phytochrome. A 2.12 kbp region of each Eco RI fragment was found to be an exon containing sufficient nucleotide sequence to complete the coding regions for the NH<sub>2</sub>-terminal portion of their respec-

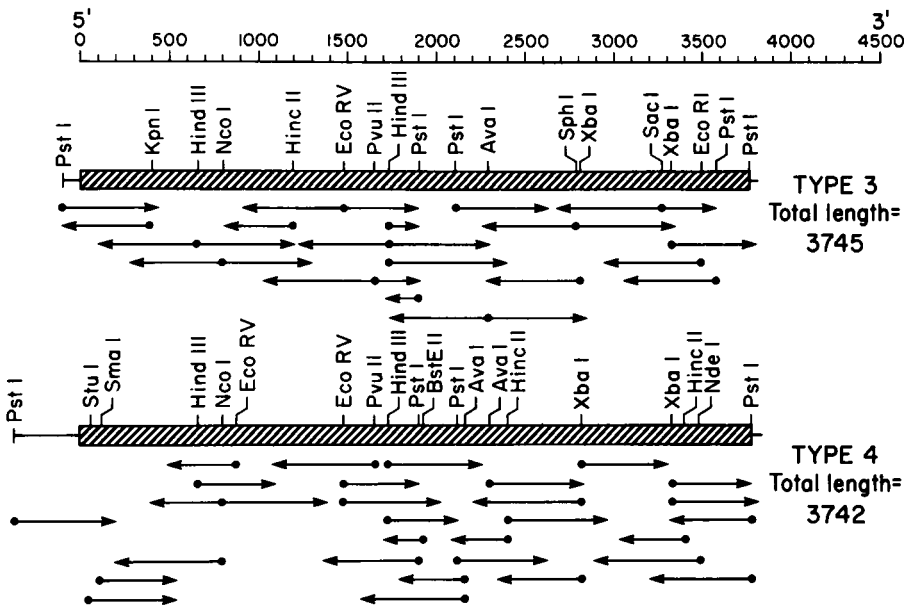


Figure 3. Sequencing strategies used to determine type 3 and 4 phytochrome nucleotide sequences. The arrows indicate individual sequence readings from the labeled sites. A similar strategy was used for type 5 sequencing.

tive phytochrome types (Fig. 1). The full sequence and organization of these genes will be presented elsewhere.

The strategies employed for the sequencing of these clones are shown in Figure 3 and the composite nucleotide sequences in Figure 4. Type 3 and 4 sequences both have an open reading frame of 3384 bp, terminated by two TGA stop codons in tandem. The nucleotide sequence homology between these two sequences and the partial type 5 sequence is 97.4% in the coding and immediately adjacent 5' untranslated regions. Analysis of the codon usage frequency within the open reading frame indicates no substantial preference for any individual codons except for codons containing the dinucleotide CpG which occurs at a significantly lower frequency than all other dinucleotides (data not shown). The 3' untranslated region of the type 3 sequence is 252 bp while that of type 4 is 261 bp. A putative poly(A)<sup>+</sup> attachment signal AATAAA is present starting 35 bp and 36 bp upstream of the poly(A)<sup>+</sup> tract in the type 3 and 4 sequences respectively. The apparent low homology in the 3' untranslated regions of the two clones can be increased from 34% to 72% if realignment of these sequences is permitted to account for a

AP 5

- 138 AGTCA GACCCCTAC TCCACTCCAC - 111

AP 5	AP 4	AP 3	AP 2	AP 1	AP 0	AP -1	AP -2	AP -3	AP -4	AP -5	AP -6	AP -7	AP -8	AP -9	AP -10	AP -11	AP -12	AP -13	AP -14	AP -15	AP -16	AP -17	AP -18	AP -19	AP -20	AP -21	AP -22	AP -23	AP -24	AP -25	AP -26	AP -27	AP -28	AP -29	AP -30	AP -31	AP -32	AP -33	AP -34	AP -35	AP -36	AP -37	AP -38	AP -39	AP -40	AP -41	AP -42	AP -43	AP -44	AP -45	AP -46	AP -47	AP -48	AP -49	AP -50	AP -51	AP -52	AP -53	AP -54	AP -55	AP -56	AP -57	AP -58	AP -59	AP -60	AP -61	AP -62	AP -63	AP -64	AP -65	AP -66	AP -67	AP -68	AP -69	AP -70	AP -71	AP -72	AP -73	AP -74	AP -75	AP -76	AP -77	AP -78	AP -79	AP -80	AP -81	AP -82	AP -83	AP -84	AP -85	AP -86	AP -87	AP -88	AP -89	AP -90	AP -91	AP -92	AP -93	AP -94	AP -95	AP -96	AP -97	AP -98	AP -99	AP -100	AP -101	AP -102	AP -103	AP -104	AP -105	AP -106	AP -107	AP -108	AP -109	AP -110	AP -111
AP 5	AP 4	AP 3	AP 2	AP 1	AP 0	AP -1	AP -2	AP -3	AP -4	AP -5	AP -6	AP -7	AP -8	AP -9	AP -10	AP -11	AP -12	AP -13	AP -14	AP -15	AP -16	AP -17	AP -18	AP -19	AP -20	AP -21	AP -22	AP -23	AP -24	AP -25	AP -26	AP -27	AP -28	AP -29	AP -30	AP -31	AP -32	AP -33	AP -34	AP -35	AP -36	AP -37	AP -38	AP -39	AP -40	AP -41	AP -42	AP -43	AP -44	AP -45	AP -46	AP -47	AP -48	AP -49	AP -50	AP -51	AP -52	AP -53	AP -54	AP -55	AP -56	AP -57	AP -58	AP -59	AP -60	AP -61	AP -62	AP -63	AP -64	AP -65	AP -66	AP -67	AP -68	AP -69	AP -70	AP -71	AP -72	AP -73	AP -74	AP -75	AP -76	AP -77	AP -78	AP -79	AP -80	AP -81	AP -82	AP -83	AP -84	AP -85	AP -86	AP -87	AP -88	AP -89	AP -90	AP -91	AP -92	AP -93	AP -94	AP -95	AP -96	AP -97	AP -98	AP -99	AP -100	AP -101	AP -102	AP -103	AP -104	AP -105	AP -106	AP -107	AP -108	AP -109	AP -110	AP -111
AP 5	AP 4	AP 3	AP 2	AP 1	AP 0	AP -1	AP -2	AP -3	AP -4	AP -5	AP -6	AP -7	AP -8	AP -9	AP -10	AP -11	AP -12	AP -13	AP -14	AP -15	AP -16	AP -17	AP -18	AP -19	AP -20	AP -21	AP -22	AP -23	AP -24	AP -25	AP -26	AP -27	AP -28	AP -29	AP -30	AP -31	AP -32	AP -33	AP -34	AP -35	AP -36	AP -37	AP -38	AP -39	AP -40	AP -41	AP -42	AP -43	AP -44	AP -45	AP -46	AP -47	AP -48	AP -49	AP -50	AP -51	AP -52	AP -53	AP -54	AP -55	AP -56	AP -57	AP -58	AP -59	AP -60	AP -61	AP -62	AP -63	AP -64	AP -65	AP -66	AP -67	AP -68	AP -69	AP -70	AP -71	AP -72	AP -73	AP -74	AP -75	AP -76	AP -77	AP -78	AP -79	AP -80	AP -81	AP -82	AP -83	AP -84	AP -85	AP -86	AP -87	AP -88	AP -89	AP -90	AP -91	AP -92	AP -93	AP -94	AP -95	AP -96	AP -97	AP -98	AP -99	AP -100	AP -101	AP -102	AP -103	AP -104	AP -105	AP -106	AP -107	AP -108	AP -109	AP -110	AP -111

Nucleic Acids Research

Table with 3 columns: sequence identifier (AP5/AP4/AP3/AP2/AP4), DNA sequence (lines of nucleotide bases), and a numerical value (e.g., 1530, 610, 1620, 540, 1710, 570, 1800, 600, 1800, 630, 1900, 660, 2070, 690, 2180, 720, 2280, 750, 2340, 780, 2430, 810, 2520, 840, 2610, 870, 2700, 900, 2790, 930, 2880, 960, 2970, 990, 3060, 1020, 3180, 1050, 3240, 1080, 3330, 1110, 3426, 1129).



number of deletions and insertions (not shown). Nevertheless, sufficiently low homology remains within several tracts of 3' sequence to indicate that appropriate sequences from these regions could serve as specific probes for transcripts derived from type 3 and type 4 genes.

#### Amino acid sequence analysis

The complete, derived amino acid sequences for phytochrome isoforms 3 and 4, together with the NH<sub>2</sub>-terminal 492 residues of the type 5 isoform, are shown in figure 4. The two full-length polypeptides are both 1128 amino acid residues long (excluding the initiator methionine) with molecular masses of 124,870 daltons and 124,949 daltons, comparing favorably with estimates of 1127 residues and 124,534 daltons from amino acid analysis and SDS polyacrylamide gel electrophoresis of the protein (31). The two proteins are 97.8% homologous with most of the amino acid differences involving conservative substitutions. The resultant differences in amino acid composition of type 3 and 4 sequences, are shown in Table 1, together with the previously determined composition of 124 kd Avena phytochrome (31). A computerized analysis of the isoform 3 sequence predicts a pI of 5.9, in precise agreement with that determined experimentally for purified Avena phytochrome (15). One interesting feature, of as yet undetermined significance, is that 8 of the first 11 residues at the NH<sub>2</sub>-terminus of all sequence isoforms are serines (Fig. 4).

Two different studies have identified a single, identical undecapeptide which carries the chromophore in proteolytic digests of Avena phytochrome (32,33). This 11-amino acid sequence is LRAPHSCHLQY, with the chromophore covalently attached to the cysteine residue. This same sequence was found to be invariant for all chromopeptides isolated from Avena phytochrome (32), indicating that the amino acid sequence at the chromophore attachment site is identical regardless of the location or number of chromophores. However, the number of times this sequence occurs in the polypeptide and therefore the number of chromophores per monomer could not be determined in these studies. The present data show that this

Figure 4. Nucleotide and derived amino acid sequences for type 3, 4 and 5 phytochromes. The complete nucleotide sequence of type 3 phytochrome is shown (AP3). Above this sequence are shown the nucleotide sequences for types 4 (AP4) and 5 (AP5) phytochromes. The nucleotides identical to type 3 are indicated by asterisks; those found to differ from the type 3 sequence are listed. The deduced amino acid residues for type 3 phytochrome are denoted below the type 3 nucleotide triplets in standard one letter code. Deduced amino sequences for types 4 and 5 are shown below that for type 3 with only the amino acid differences shown by code letter.

Table 1  
Phytochrome amino acid composition

Amino acid residue	Type-3	From sequence	Type-4	From Vierstra and Quail (31)
Arg	53		54	58
Lys	64		64	65
Asn	45	} 113	43	-
Asp	68		67	-
Gln	49	} 126	49	-
Glu	77		78	-
His	34		35	36
Pro	46		45	48
Tyr	21		21	19
Trp	10		10	10
Ser	89		92	83
Thr	43		40	46
Gly	69		70	79
Ala	96		96	96
Met	34		34	34
Cys	23		23	20
Phe	44		46	44
Leu	124		120	124
Val	83		84	76
Ile	56		57	50
Total	1128		1128	1127
Molecular mass (daltons)	124,870		124,949	124,534

11-amino acid sequence occurs only once in each full-length polypeptide (types 3 and 4), corresponding to amino acids 315-325, with the chromophore attached to Cys 321. Thus these data establish that native phytochrome carries only one chromophore per protein monomer.

#### Sequence similarity search

One approach to obtaining indications to as yet undetermined functions or evolutionary origins of a given protein is to scan protein sequence libraries for extant homologous sequences (26). We searched the current NBRF database for proteins with sequences similar to phytochrome with negative results. Although a variety of proteins with restricted stretches of weak to moderate similarity were identified, none had search scores approaching those considered to represent genuinely related sequences (26).

#### Structural analysis

The full-length phytochrome amino acid sequence has been subjected to

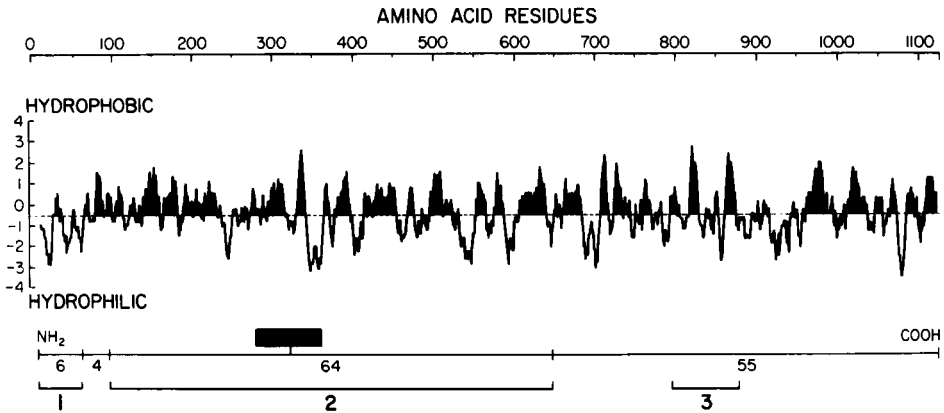


Figure 5. Hydropathy profile of type 3 phytochrome. Solid peaks indicate hydrophobic regions; unshaded troughs indicate hydrophilic regions. A peptide map indicating positions of the chromophore (solid rectangle) and regions containing epitopes for three classes of monoclonal antibody (square brackets, 1,2,3) are shown below.

analysis with a set of algorithms designed to provide information on various structural features of the molecule. The hydropathy profile (Fig. 5) and secondary structure analysis (data not shown) show the chromophore covalent-attachment site to be at a  $\beta$ -turn in a mildly hydrophilic segment between two strongly hydrophobic regions, one of which is relatively extensive (from about residue 80 to 315 from the NH<sub>2</sub>-terminus). The NH<sub>2</sub>-terminal 6-10 kd segment, known to be critical to fidelity of protein-chromophore interaction (4), is predominantly hydrophilic. One striking feature is the extensive, strongly hydrophilic region between residues 343 and 365. This region, which is predicted as  $\alpha$ -helical, also carries a high density of charged residues, with a cluster of acidic residues separated from a cluster of basic residues by a proline. Hopp-Woods (30) analysis predicts that this polypeptide segment, together with two other hydrophilic regions centered at residues 691 and 1079, may be antigenic determinants. Examination of the complete hydropathy profile shows that the mean hydrophobic index across a moving window of 19 amino acid residues does not exceed 1.06 in any hydrophobic region of the polypeptide chain. This observation indicates the absence of any segment in the polypeptide with the properties of a membrane spanning domain (34).

### DISCUSSION

The data presented here represent the first complete amino acid sequences for the phytochrome apoprotein. The overall picture of the photoreceptor now is that of a molecule that is a dimer (2) of two 1128 amino-acid polypeptides each carrying a single, covalently linked chromophore in a globular, NH<sub>2</sub>-terminal domain. The restriction map polymorphism among the cDNA clones indicates that the protein subunits are the products of at least four genes. This result is consistent with previous Southern blot analysis of Avena genomic DNA which indicated the existence of a small gene family for phytochrome (18).

The 3'-untranslated regions of the type 3 and 4 transcripts appear to be complete, but the transcription start sites have not yet been identified. The nucleotides presumed to be missing from the 5' ends of phytochrome types 3 and 4 sequences by comparison to the sequence of the 5' end of type 5 phytochrome are currently being sought through further analysis of their respective genes. The homology between the different nucleotide sequences is high throughout the 5'-untranslated and coding regions, with the majority of base changes leading to amino acid substitutions (~98% sequence homology at both nucleotide and amino acid levels). In comparison, the 3'-untranslated regions of the type 3 and 4 sequences exhibit a considerably higher level of divergence, apparently the result of drift as well as a number of insertions and/or deletions. This divergence offers the opportunity to generate gene specific probes from the 3'-untranslated sequences. It will be of interest to quantitate the relative proportions of the various phytochrome transcripts in the cellular mRNA pool and to determine whether all are coordinately under negative feedback regulation by the photoreceptor (13,14). The relative frequency of cDNA clones representing the different isoforms in our libraries suggests that there may be substantial differences in the levels of expression of each, since more than 80% of the cDNA clones isolated were either of type 4 or 5.

The consequences of the existence of phytochrome isoforms are presently uncertain. The level of protein sequence conservation is clearly very high and, as expected from the fact that the majority of observed amino acid changes are conservative substitutions, the sequence heterogeneity is found to have negligible effects on various structural features such as the hydropathy profile and predicted secondary structure (data not shown). Nevertheless, the possibility cannot be excluded that localized

changes in structure and/or charge distribution may affect such parameters as subunit-subunit interaction or contribute to functional heterogeneity in the phytochrome population. It is not yet possible to determine whether phytochrome exists as a heterogeneous pool of homodimers or a mixture of heterodimers assembled from a common pool of subunit isoforms. The heterogeneity could affect protein-chromophore interaction and contribute to the multiple kinetic populations of phytochrome that have been reported in flash photolysis (35) and dark reversion (31) studies. The presence of a D-P linkage in the Type 3 sequence at positions 539-540 and its absence from the Type 4 sequence will permit these two polypeptides to be distinguished by chemical cleavage methods. As the full-length sequences of the other isoforms are determined other methods of distinguishing between them such as distinctive proteolytic or chemical-cleavage peptide maps may become apparent. Possible differential recognition of the various isoforms by different members of our monoclonal antibody library (39) is also worthy of investigation.

The phytochrome amino acid sequence presented here establishes two principle features of the molecular structure of the photoreceptor. First, the sequence provides direct evidence that phytochrome has one chromophore per 124 kd monomer, thereby substantiating the conclusion reached previously on the basis of indirect data from quantitative amino acid analysis and spectroscopic studies (36,37). Second, the localization of the chromophore attachment site to the NH<sub>2</sub>-terminal domain of the molecule has permitted correct orientation of proteolytic cleavage maps of the polypeptide (4,38), and identification of the locations of spatially separate, primary-sequence epitopes for different monoclonal antibodies (39).

Although the various algorithms used here for structural analysis of the phytochrome molecule have their limitations (40), some points are nevertheless worth noting. The chromophore has been postulated to be housed in the interior of the phytochrome molecule, relatively inaccessible to the external hydrophilic medium (41,42). It seems possible that the relatively extensive hydrophobic region of the polypeptide revealed by hydropathy analysis to reside between residue 80 and 315 could provide a cavity which contains the tetrapyrrole. In contrast, the NH<sub>2</sub>-terminal 6 kd polypeptide segment, known to be critical to fidelity of protein-chromophore interaction (15,31), is predominantly hydrophilic and therefore likely to be located at the surface of the molecule. This feature is consistent with chemical probe data that indicate that this segment shields

the chromophore from the external aqueous environment (16,41,42).

Hopp-Woods (30) analysis of the phytochrome amino acid sequence predicts that three strongly hydrophilic polypeptide segments, centered at residues 347, 691 and 1079, should carry antigenic determinants. However, while the epitope for one class of monoclonal antibodies has been mapped to the chromophore bearing region (39) and may be located near the 347 peak, neither the 691 nor 1079 region corresponds to an epitope for any of the monoclonal antibodies mapped to the COOH-terminal domain of the polypeptide (39). Moreover, the Hopp-Woods algorithm does not predict an antigenic determinant in the NH<sub>2</sub>-terminal 6 kd, a region we have found to be an apparently immunodominant domain carrying epitopes for one class of monoclonal antibodies (39).

The hydropathy profile is consistent with the peptide fragmentation pattern obtained with limited proteolytic digestion of phytochrome. Two regions of the molecule, one at the NH<sub>2</sub>-terminus and the other in the center, are selectively vulnerable to attack by serine protease-like activities present in *Avena* extracts (4,15,39) as well as to other proteases (2,38). Arg-53, Lys-91 and Lys-648, located in strongly hydrophilic to neutral segments, could be candidates for attack by these proteases, which would lead to cleavage of 6 and 10 kd from the NH<sub>2</sub>-terminus and 55 kd from the COOH-terminus. These presumably surface-located regions of the polypeptide are known to undergo conformational changes upon photoconversion, altering their accessibility to proteolytic attack (4,15,31,38). In this context it is of interest to consider the NH<sub>2</sub>-terminal amino acid sequences KAL and ALV reported by Hunt and Pratt (43) for presumptively 118/114 kd phytochrome. The sequence ALV is not detected anywhere in the polypeptide. The sequence KAL on the other hand, while it occurs at residues 142-144, would correspond to the NH<sub>2</sub>-terminus of a polypeptide of only 106 kd. The reason for this discrepancy is unknown.

A popular hypothesis of phytochrome action based on physiological experiments has been that the active photoreceptor functions by virtue of its location in cellular membranes (44), with some authors postulating that the molecule forms ionic channels through the bilayer (45). However, the hydropathic properties of the polypeptide make at least the latter possibility remote. Kyte and Doolittle (34) surveyed the hydropathic properties of the transmembrane segments of an array of established intrinsic membrane proteins. The mean hydropathic index across a window of 19 amino acids, averaged ~1.9 for all transmembrane peptide segments, whereas soluble glo-

bular proteins had comparable mean indices of  $\sim 1.1$ . No region of the phytochrome molecule has a mean index exceeding 1.06, more consistent with a soluble globular protein than an intrinsic membrane protein.

The search for proteins in the NBRF database with sequences similar to that of phytochrome is of interest for two main reasons. First, because the molecular mechanism of phytochrome action is unknown, the identification of proteins with homologies to the photoreceptor might provide some insight into that mechanism (26). Unfortunately this approach provided no immediately useful information in the case of phytochrome. Second, because the tetrapyrrole chromophore of phytochrome is very closely related to the chromophores of the phycobiliproteins of cyanobacteria and red algae, the possible evolutionary relationship between the two groups of chromoproteins has long been a matter of speculation (46). The absence of any substantial sequence similarity between phytochrome and the phycobiliproteins currently available in the NBRF database thus far speaks against a common evolutionary origin for the apoproteins of these photoreceptors.

#### ACKNOWLEDGEMENTS

We are grateful to Dr. Michael G. Murray for generously providing the Avena genomic library used in this work and M. Speckhard for technical assistance. We thank J. Devereux, M. Gribskof, F. Blattner and W. Fitch for helpful discussions and assistance in performing computer analysis of nucleotide and amino acid sequences. Work supported by National Science Foundation Grants PCM 8003792 and PCM 8302206 and U.S. Department of Agriculture, Science and Education Administration Grants 59-2551-1-1-744-0 and 81-CRCR-1-744-0. J.L.L. is the recipient of a National Science Foundation Graduate Fellowship.

<sup>+</sup>Present Address: E.I. du Pont de Nemours & Company, Central Research and Development Department, Experimental Station, Wilmington, DE 19898

<sup>\*</sup>Present Address: Plant Breeding Institute, Trumpington, Cambridge, CB2 2LQ England

#### REFERENCES

1. Shropshire, W. Jr. and Mohr, H., eds. (1983) Encyclopedia of plant physiology, New series, Vols. 16A and 16B, Springer-Verlag, Heidelberg, Berlin.
2. Jones, A.M., Vierstra, R.D., Daniels, S.M. and Quail, P.H. (1984) Plant Physiol. **75**, S73.
3. Pratt, L.H. (1979) Ann. Rev. Plant Physiol. **33**, 557-582.

4. Vierstra, R.D., Cordonnier, M.-M, Pratt, L.H. and Quail, P.H. (1984) *Planta* 160, 521-528.
5. Stiekema, W.J., Wimpee, C.F., Silverthorne, J. and Tobin, E.M. (1983) *Plant Physiol.* 72, 717-724.
6. Thompson, W.F., Everett, M., Polans, N.O., Jorgensen, R.A. and Palmer, J.D. (1983) *Planta* 158, 487-500.
7. Kaufman, L.S., Thompson, W.S. and Briggs, W.R. (1984) *Science* 226, 1447-1449.
8. Batschauer, A. and Apel, K. (1984) *Eur. J. Biochem.* 132, 593-591.
9. Thien, W. and Schopfer, P. (1982) *Plant Physiol.* 69, 1156-1160.
10. Silverthorne, J. and Tobin, E.M. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1112-1116.
11. Mösinger, E., Batschauer, A., Schäfer, E. and Apel, K. (1985) *Eur. J. Biochem.* 147, 137-142.
12. Gallagher, T.F. and Ellis, R.J. (1982) *EMBO J.* 1, 1493-1498.
13. Colbert, J.T., Hershey, H.P. and Quail, P.H. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2248-2252.
14. Colbert, J.T., Hershey, H.P. and Quail, P.H. (1985) *Plant Mol. Biol.*, in press.
15. Vierstra, R.D. and Quail, P.H. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5272-5276.
16. Hahn, T.R., Song, P.S., Quail, P.H. and Vierstra, R.D. (1984) *Plant Physiol.* 74, 755-758.
17. Hershey, H.P., Colbert, J.T., Lissemore, J.L., Barker, R.F. and Quail, P.H. (1984) *Proc. Natl. Acad. Sci. USA* 81: 2332-2336.
18. Hershey, H.P., Barker, R.F., Colbert, J.T., Lissemore, J.L. and Quail, P.H. (1985) In *Molecular Form and Function of the Plant Genome*, van Vloten Doting L., Groot, G.S.P. and Hall, T.C. Eds., pp. 101-111, Plenum, New York.
19. Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
20. Hanahan, D. and Meselson, M. (1983) *Methods Enzymol.* 100, 333-342.
21. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1523.
22. Murray, M.G., Kennard, W.C., Drong, R.F. and Slightom, J.L. (1984) *Gene* 30, 237-240.
23. Benton, W.D. and Davis, R.W. (1977) *Science* 196, 180-183.
24. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-512.
25. Barker, R.F., Idler, K.B., Thompson, D.V. and Kemp, J.D. (1983) *Plant Mol. Biol.* 2, 335-350.
26. Lipmann, D.J. and Pearson, W.R. *Science* 227, 1435-1441.
27. Devereux, J. and Haerberli, P. (1985) *Program Library of the University of Wisconsin Genetics Computer Group, Version 2.1.*
28. Chou, P.Y. and Fasman, G.D. (1978) *Ann. Rev. Biochem.* 47, 251-276.
29. Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.
30. Hopp, T.P. and Woods, K.R. (1983) *Mol. Immunol.* 20, 483-489.
31. Vierstra, R.D. and Quail, P.H. (1983) *Biochemistry* 22, 2498-2505.
32. Lagarias, J.C. and Rapoport, H. (1980) *J. Am. Chem. Soc.* 102, 4821-4828.
33. Fry, K.T. and Mumford, F.E. (1971) *Biochem. Biophys. Res. Commun.* 45, 1466-1473.
34. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
35. Cordonnier, M.M., Mathis, P. and Pratt, L.H. (1981) *Photochem. Photobiol.* 34, 733-740.
36. Roux, S.J., McEntire, K. and Brown, W.E. (1982) *Photochem. Photobiol.* 35, 537-543.
37. Brandlmeier, T., Scheer, H. and Rüdiger, W. (1981) *Z. Naturforsch.* 36c, 431-439.



- 
38. Lagarias, J.C. and Mercurio, F.M. (1985) *J. Biol. Chem.* **260**, 2415-2423.
  39. Daniels, S.M. and Quail, P.H. (1984) *Plant Physiol.* **76**, 622-626.
  40. Kabsch, W. and Sander, C. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1075-1078.
  41. Song, P.S. (1983) *Ann. Rev. Biophys. Bioeng.* **12**, 35-68.
  42. Eilfeld, P. and Rüdiger, W. (1984) *Z. Naturforsch.* **39c**, 742-745.
  43. Hunt, R.E. and Pratt, L.H. (1980) *Biochemistry* **19**, 390-394.
  44. Marmé, D. (1977) *Ann. Rev. Plant Physiol.* **28**, 173-198.
  45. Brownlee, C., Roth-Bejerano, N. and Kendrick, R.E. (1979) *Sci. Prog. Oxf.* **66**, 217-223.
  46. Shropshire, W. Jr. (1977) In *The Science of Photobiology*, Smith K.C., Ed., pp. 281-312, Plenum, New York.