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

Howard P.Hershey[+], Richard F.Barker[1]*, Kenneth B.Idler[1], James L.Lissemore and Peter H.Quail

Department of Botany, University of Wisconsin, Madison, WI 53706 and [1]Agrigenetics Advanced Research Laboratory, Madison, WI 53716, USA

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 mole-
cule is a dimer, with each subunit composed of a linear tetrapyrrole chro-
mophore 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 (λmax = 666 nm) region of the spectrum and the
Pfr form that absorbs maximally in the far-red (λmax = 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 <u>Avena</u> phytochrome deduced from analysis of cloned cDNA and genomic sequences. We present evidence that at least four different phytochrome genes are expressed in <u>Avena</u> 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 ^{32}P-labeled to >10^8 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 weigh <u>Avena</u> 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 1.5x10^6 pfu library was
screened for recombinants containing phytochrome sequences by <u>in situ</u>
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 <u>et al</u>. (25). Computer analysis of DNA
sequences were performed by using programs made available by Drs. 0.
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 phy-
tochrome 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
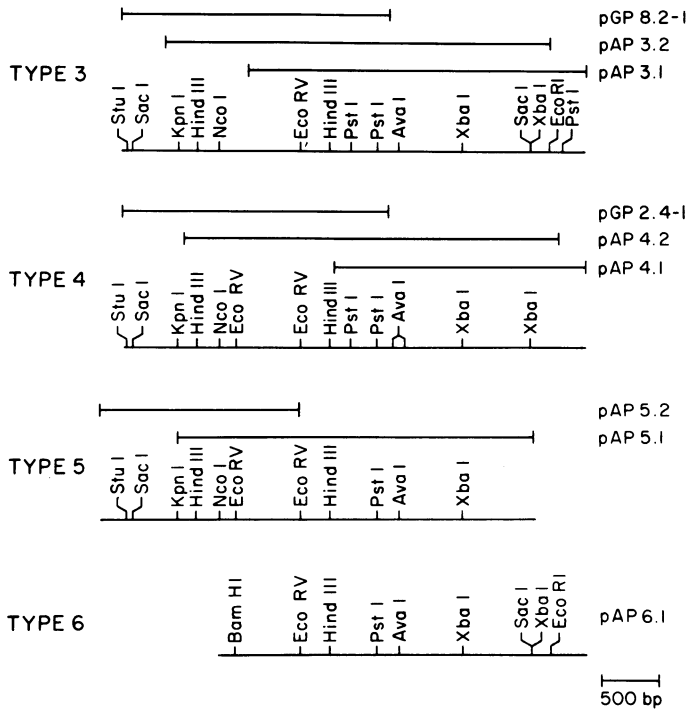
Figure 1. Composite restriction maps of cDNA and genomic clones representing four different expressed genes found in etiolated Avena. The overlaping cDNA clones (pAP-3.1, pAP-3.2, pAP-4.1, pAP-4.2, pAP-5.1, pAP-5.2, pAP-6) and/or subcloned genomic fragments (pGP8.2-1, pGP2.4-1) used to construct each 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
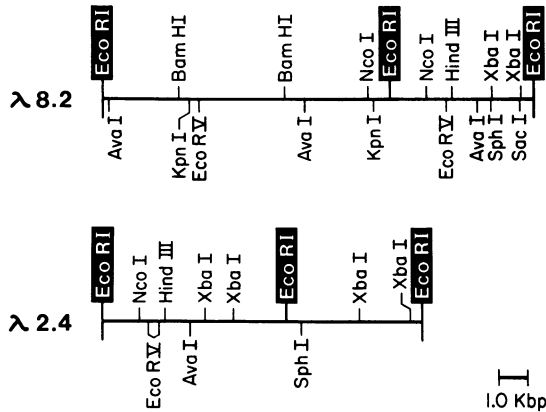
Figure 2. Restriction maps for the cloned _Avena_ genomic fragments λ8.2 and λ2.4.

cDNA clone found to contain uninterrupted sequence representing both the 5' untranslated region and coding sequence for the NH_2-terminus of the protein. NH_2-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 λ8.2, to be a likely candidate to contain the type 3 phytochrome gene and λ2.4 to be a strong candidate to contain the type 4 gene (Fig. 2). Southern blots of Eco R1 digests of these two clones probed with ^{32}P-labeled pAP-3.2 showed only the 5.2 kbp Eco RI fragment of λ8.2 (Hershey et al. 1985) and the 6.6 kbp fragment of λ2.4 (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 λ8.2 and pGP2.4-1 containing the 6.6 kbp Eco RI fragment from λ2.4. Sequence analysis of these subcloned fragments confirmed the hypothesis that λ8.2 contained at least part of the gene for type 3 phytochrome and λ2.4 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 nucletoide sequence to complete the coding regions for the NH_2-terminal portion of their respec-
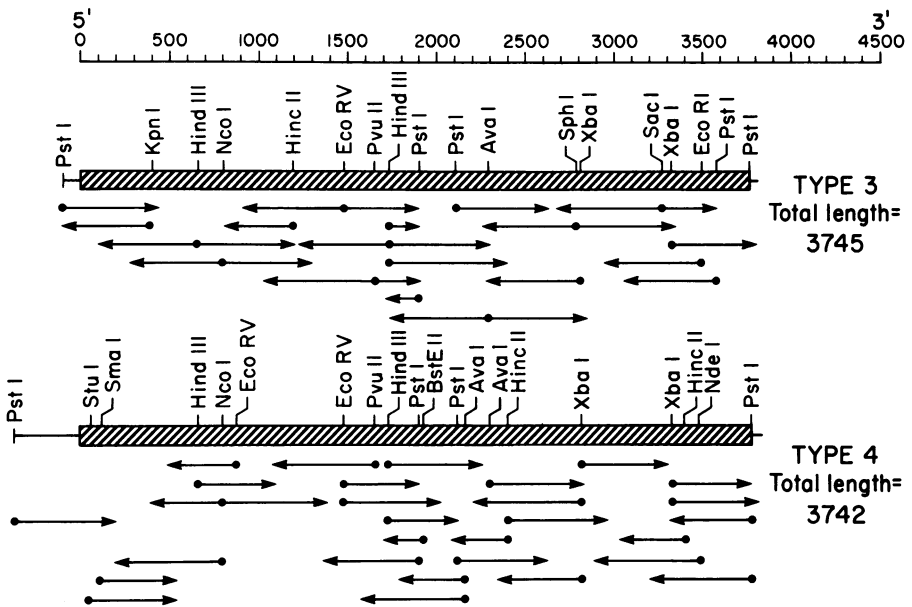
Figure 3. Sequencing strategies used to determine type 3 and 4 phytochrome nucleotide sequences. The arrows indicate individual sequence readings from the labeled sites. A similar strategy was used for type 5 sequencing.

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 dinucleo-tides (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)$^+$ attachment signal AATAAA is present starting 35 bp and 36 bp upstream of the poly(A)$^+$ tract in the type 3 and 4 sequences respectively. The apparent low homo-logy 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

AP5

-136   AGTCCA GACCCCTCAC TCCACTCCAC   -111

AP5   TCCACTCCTG CTCCTCTCCG CCTCCACCTC TCCATTGCCG CCGCCGGTAG CAG•••••• •••••••••• •••••••••• •••••••••• •••••••••• •••••••••
AP4
AP3   -57 CAGGAGC GATACACGGG GTATACGACC GTTGAGTGGT TCAATTACTT GAGGCAGGCG   -1

AP5
AP4
AP3   ATG TCT TCC TCA AGG CCT GCT TCC AGT TCT TCC AGC AGG AAC CGC CAG AGC TCC CAA GCA AGG GTG TTA GCA CAG ACA ACC CTT GAT GCC   90
AP3   M   S   S   S   R   P   A   S   S   S   S   S   R   N   R   Q   S   S   Q   A   R   V   L   A   Q   T   T   L   D   A   30
AP4
AP5

AP5
AP4
AP3   GAG CTC AAT GCT GAA TAT GAA GAA TCT GGT GAC TCC TTT GAC TAC TCC AAG CTG GTT GAA GCC CAG CGG GAT GGT CCA CCT GTG CAG CAA   180
AP3   E   L   N   A   E   Y   E   E   S   G   D   S   F   D   Y   S   K   L   V   E   A   Q   R   D   G   P   P   V   Q   Q   60
AP4
AP5

AP5
AP4
AP3   GGG CGG TCG GAG AAG GTC ATA GCC TAC TTA CAG CAC ATT CAG AAA GGA AAG CTA ATC CAA ACA TTT GGT TGC CTG TTG GCC CTT GAT GAG   270
AP3   G   R   S   E   K   V   I   A   Y   L   Q   H   I   Q   K   G   K   L   I   Q   T   F   G   C   L   L   A   L   D   E   90
AP4
AP5

AP5
AP4
AP3   AAG AGC TTC AAT GTC ATC GCG TTC AGT GAG AAC GCG CCA GAA ATG CTT ACA ACG GTC AGC CAT GCG GTA CCC AGT GTT GAT GAT CCC CCA   360
AP3   K   S   F   N   V   I   A   F   S   E   N   A   P   E   M   L   T   T   Y   S   H   A   V   P   S   V   D   D   P   P   120
AP4
AP5

AP5
AP4
AP3   AGG CTG GGG ATT GGC ACC AAT GTA CGG TCT CTT TTC AGT GAC CAA GGT GCC ACA GCA CTG CAT AAG GCA CTA GGA TTT GCT GAT GTG TCT   450
AP3   R   L   G   I   G   T   N   V   R   S   L   F   S   D   Q   G   A   T   A   L   H   K   A   L   G   F   A   D   V   S   150
AP4
AP5

AP5
AP4
AP3   TTG CTG AAT CCT ATC CTG GTT CAG TGC AAG ACA TCA GGC AAG CCT TTC TAT GCC ATT GTT CAT CGA GCA ACT GGT TGT TTG GTG GTA GAC   540
AP3   L   L   N   P   I   L   V   Q   C   K   T   S   G   K   P   F   Y   A   I   V   H   R   A   T   G   C   L   V   Y   D   180
AP4
AP5

AP5
AP4
AP3   TTT GAG CCT GTA AAG CCT ACA GAA TTT CCT GCC ACT GCT GCT GGG GCT TTG CAG TCC TAC AAG CTT GCT GCC AAG GCA ATA TCC AAG ATC   630
AP3   F   E   P   V   K   P   T   E   F   P   A   T   A   A   G   A   L   Q   S   Y   K   L   A   A   K   A   I   S   K   I   210
AP4
AP5

AP5
AP4
AP3   CAG TCA TTG CCA GGT GGA AGC ATG GAG GTG CTA TGC AAT ACT GTG GTG AAG GAA GTC TTT GAC CTT ACC GGG TAT GAC AGG GTT ATG GCT   720
AP3   Q   S   L   P   G   G   S   M   E   V   L   C   N   T   V   V   K   E   V   F   D   L   T   G   Y   D   R   V   M   A   240
AP4
AP5

AP5
AP4
AP3   TAC AAG TTT CAT GAA GAT GAC CAT GGT GAG GTA TTC TCC GAA ATC ACA AAG CCT GGT CTT GAG CCT TAT CTA GGC CTG CAC TAT CCA GCC   810
AP3   Y   K   F   H   E   D   D   H   G   E   V   F   S   E   I   T   K   P   G   L   E   P   Y   L   G   L   H   Y   P   A   270
AP4
AP5

AP5
AP4
AP3   ACT GAC ATC CCT CAA GCA GCC AGG CTT CTT TTC ATG AAG AAC AAA GTA CGG ATG ATT TGT GAT TGC CGT GCG AGA TCC ATA AAG GTC ATT   900
AP3   T   D   I   P   Q   A   A   R   L   L   F   M   K   N   K   V   R   M   I   C   D   C   R   A   R   S   I   K   V   I   300
AP4
AP5

AP5
AP4
AP3   GAG GCT GAG GCA CTC CCC TTT GAT ATT AGC CTA TGT GGT TCA GCA CTC AGG GCA CCA CAC AGT TGT CAC CTT CAG TAT ATG GAG AAC ATG   990
AP3   E   A   E   A   L   P   F   D   I   S   L   C   G   S   A   L   R   A   P   H   S   C   H   L   Q   Y   M   E   N   M   330
AP4
AP5

AP5
AP4
AP3   AAC TCG ATT GCA TCC CTT GTC ATG GCT GTT GTG GTT AAT GAG AAT GAA GAG GAT GAT GAA GCT GAG TCT GAA CAA CCA GCA CAG CAG CAG   1080
AP3   N   S   I   A   S   L   V   M   A   V   V   V   N   E   N   E   E   D   D   E   A   E   S   E   Q   P   A   Q   Q   Q   360
AP4
AP5

AP5
AP4
AP3   C•G
AP3   AAA AAG AAG AAA CTA TGG GGC CTC CTT GTT TGC CAC CAT GAG AGC CCC AGA TAT GTC CCT TTT CCG CTG CGT TAT GCT TGT GAG TTC TTA   1170
AP3   K   K   K   K   L   W   G   L   L   V   C   H   H   E   S   P   R   Y   V   P   F   P   L   R   Y   A   C   E   F   L   390
AP4
AP5

AP5
AP4
AP3   GCA CAG GTG TTT GCT GTC CAT GTC AAC AGG GAG TTT GAA TTA GAG AAA CAG TTG CGT GAG AAG AAC ATA CTG AAG ATG CAA ACA ATG CTC   1260
AP3   A   Q   V   F   A   V   H   V   N   R   E   F   E   L   E   K   Q   L   R   E   K   N   I   L   K   M   Q   T   M   L   420
AP4
AP5

AP5
AP4
AP3   TCT GAT ATG TTG TTC CGA GAA GCC TCT CCC CTG ACT ATC GTA TCA GGG ACC CCC AAT ATC ATG GAC CTA GTC AAA TGT GAT GGT GCT GCT   1350
AP3   S   D   M   L   F   R   E   A   S   P   L   T   I   V   S   G   T   P   N   I   M   D   L   V   K   C   D   G   A   A   450
AP4
AP5

AP5
AP4
AP3   CTT CTG TAT GGG GGA AAA GTA TGG CGT CTG CGT AAT GCT CCA ACG GAG TCT CAG ATA CAT GAT ATC GCC TTC TGG CTA TCA GAT GTT CAC   1440
AP3   L   L   Y   G   G   K   V   W   R   L   R   N   A   P   T   E   S   Q   I   H   D   I   A   F   W   L   S   D   V   H   480
AP4
AP5

```
AP5    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •G
AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  ••   •••  •••  •••  T••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    AGG GAT TCC ACT GGC CTG AGT ACT GAC AGC CTC CAT GAT GCT GGC TAT CCA GGA GCT GCT GCT CTT GGT GAT ATG ATT TGT GGA ATG GCA   1530
AP3    R   D   S   T   G   L   S   T   D   S   L   H   D   A   G   Y   P   G   A   A   A   L   G   D   M   I   C   G   M   A       510
AP4    •
AP5    •

AP4    •••  •••  •••  •••  •••  •••  •••  •••  A••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  C••  •••  T•G
AP3    GTG GCT AAG ATC AAC TCC AAG GAT ATT CTT TTT TGG TTC AGG TCA CAT ACA GCT GCT GAA ATC AGA TGG GGA GGT GCA AAG AAT GAT GCA   1620
AP3    V   A   K   I   N   S   K   D   I   L   F   W   F   R   S   H   T   A   A   E   I   R   W   G   G   A   K   N   D   P       540
AP4    •                                                           S

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  C••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •C•
AP3    TCG GAC ATG GAT GAC AGC AGA AGG ATG CAC CCT AGG TTG TCT TTC AAA GCT TTC CTT GAA GTT GTC AAG ATG AAG AGC TTG CCT TGG AGT   1710
AP3    S   D   M   D   D   S   R   R   M   H   P   R   L   S   F   K   A   F   L   E   V   V   K   M   K   S   L   P   W   S       570
AP4    •

AP4    •••  •••  •G   •••  •••  •••  •••  •••  •••  •C   •••  •••  •••  •C   •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    GAC TAT GAA ATG GAT GCT ATT CAT TCA TTG CAA CTT ATA CTG CGA GGG ACA CTA AAT GAT GCC AGC AAG CCA AAG CGG GAA GCT AGT TTA   1800
AP3    D   Y   E   M   D   A   I   H   S   L   Q   L   I   L   R   G   T   L   N   D   A   S   K   P   K   R   E   A   S   L       600
AP4    •

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •G
AP3    GAT AAC CAG ATT GGT GAT CTA AAA CTT GAT GGG CTT GCT GAA CTG CAG GCC GTG ACC AGT GAA ATG GTT CGT CTA ATG GAA ACA GCA ACT   1890
AP3    D   N   Q   I   G   D   L   K   L   D   G   L   A   E   L   Q   A   V   T   S   E   M   V   R   L   M   E   T   A   T       630
AP4    •

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    GTT CCA ATC TTG GCA GTA GAT GGC AAT GGA CTG GTC AAC GGG TGG AAT CAG AAA GCA GCG GAG TTG ACT GGG CTA AGA GTT GAT GAT GCA   1980
AP3    V   P   I   L   A   V   D   G   N   G   L   V   N   G   W   N   Q   K   A   A   E   L   T   G   L   R   V   D   D   A       660
AP4    •

AP4    ••A  AGG  ••A  •••  •••  •••  •••  •••  •••  •G   •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    ATT GGA CAC ATA CTT ACC CTT GTG GAG GAC TCC TCT GTA CCA GTT GTC CAG AGG ATG CTA TAT CTA GCT CTG CAG GGT AAA GAA GAG   2070
AP3    I   G   H   I   L   T   L   V   E   D   S   S   V   P   V   V   Q   R   M   L   Y   L   A   L   Q   G   K   E   E       690
AP4    •                           E

AP4    ••A  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •G•  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    AAG GAA GTT CGA TTT GAG GTA AAG ACT CAT GGC CCG AGG AGG GAT GAT GGT CCA GTT ATC TTG GTT GTG AAT GCT TGT GCC AGT CGG GAC   2160
AP3    K   E   V   R   F   E   V   K   T   H   G   P   R   R   D   D   G   P   V   I   L   V   V   N   A   C   A   S   R   D       720
AP4    •                                               R

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    CTT CAT GAT CAT GTT GTT GGA GTG TGC TTT GTT GCC CAA GAT ATG ACT GTC CAT AAG TTG GTG ATG GAC AAG TTT ACT CGG GTT GAG GGT   2250
AP3    L   H   D   H   V   V   G   V   C   F   V   A   Q   D   M   T   V   H   K   L   V   M   D   K   F   T   R   V   E   G       750
AP4    •

AP4    •••  •••  •••  •••  •••  ••A  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  ••A  •••  •••  •••
AP3    GAC TAC AAG GCG ATC ATT CAC AAC CCG AAC CCA CTC ATT CCT CCT ATA TTT GGT GCT GAC GAA TTT GGA TGG TGT TCG GAG TGG AAT GCT   2340
AP3    D   Y   K   A   I   I   H   N   P   N   P   L   I   P   P   I   F   G   A   D   E   F   G   W   C   S   E   W   N   A       780
AP4    •

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •C   •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    GCA ATG ACC AAG TTG ACT GGG TGG AAT AGA GAT GAA GTG CTC GAT AAG ATG CTT CTT GGT GAA GTG TTT GAC AGT AGC AAT GCT TCC TGC   2430
AP3    A   M   T   K   L   T   G   W   N   R   D   E   V   L   D   K   M   L   L   G   E   V   F   D   S   S   N   A   S   C       810
AP4    •

AP4    •••  •••  ••A  ••A  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  ••T  •••  •••  •••  •••  •••  •••  •••
AP3    CCT TTG AAG AAC AGA GAT GCA TTT GTA AGT CTT TGT GTT CTT ATC AAC AGT GCA TTA GCC GGG GAA GAA ACA GAA AAG GCT CCA TTT GGC   2520
AP3    P   L   K   N   R   D   A   F   V   S   L   C   V   L   I   N   S   A   L   A   G   E   E   T   E   K   A   P   F   G       840
AP4    •                K   M

AP4    •••  •••  •••  C••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    TTC TTC GAC AGA AGT GGA TAC TAT ATT GAG TGT CTT CTA TCA GCA AAC AGA AAA GAA AAT GAG GGT GGT CTC ATC ACT GGA GTA TTC TGT   2610
AP3    F   F   D   R   S   G   Y   Y   I   E   C   L   L   S   A   N   R   K   E   N   E   G   G   L   I   T   G   V   F   C       870
AP4    •       E

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  ••G  •••  •••  •••  •••  •••  •••
AP3    TTT ATT CAT GTT GCT AGT CAT GAG CTG CAA CAT GCA CTA CAG GTG CAG CAA GCC TCG GAG CAA ACG TCG CTA AAA AGG CTC AAG GCT TTC   2700
AP3    F   I   H   V   A   S   H   E   L   Q   H   A   L   Q   V   Q   Q   A   S   E   Q   T   S   L   K   R   L   K   A   F       900
AP4    •

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •C   •••  •••  •••  ••A  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    TCC TAC ATG AGA CAT GCG ATC AAC AAC CCT CTC TCA GGC ATG CTC TAC TCT AGA AAA GCA TTG AAG AAC ACA GAT TTG AAT GAA GAA CAG   2790
AP3    S   Y   M   R   H   A   I   N   N   P   L   S   G   M   L   Y   S   R   K   A   L   K   N   T   D   L   N   E   E   Q       930
AP4    •

AP4    •••  •••  •••  •••  •C   •••  •••  •C   •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •G   •••  •••  ••G  •••  •G•  •••  •••
AP3    ATG AAG CAG ATT CAT GTT GGA GAT AAT TGT CAC CAC CAG ATA AAC AAG ATA CTT GCA GAC TTG GAT CAA GAT AGC ATC ACC GAA AAA TCT   2880
AP3    M   K   Q   I   H   V   G   D   N   C   H   H   Q   I   N   K   I   L   A   D   L   D   Q   D   S   I   T   E   K   S       960
AP4    •                                                                                               S

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  G••  ••T  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    AGC TGC TTG GAT TTG GAG ATG GCT GAA TTT CTG TTG CAA GAT GTG GTG GTG GCT GCT GTA AGT CAA GTA CTG ATA ACC TGC CAG GGA AAA   2970
AP3    S   C   L   D   L   E   M   A   E   F   L   L   Q   D   V   V   V   A   A   V   S   Q   V   L   I   T   C   Q   G   K       990
AP4    •                                   V   F

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    GGG ATC AGA ATC TCT TGC AAC CTG CCA GAG AGA TTT ATG AAG CAG TCA GTC TAT GGA GAT GGT GTT CGA CTC CAG CAG ATC CTC TCT GAC   3060
AP3    G   I   R   I   S   C   N   L   P   E   R   F   M   K   Q   S   V   Y   G   D   G   V   R   L   Q   Q   I   L   S   D       1020
AP4    •

AP4    •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    TTC CTG TTT ATT TCA GTG AAG TTC TCT CCT GTT GGA GGT TCT GTT GAG ATT TCT TCC AAG CTG ACA AAG AAC AGC ATC GGA GAA AAC CTT   3150
AP3    F   L   F   I   S   V   K   F   S   P   V   G   G   S   V   E   I   S   S   K   L   T   K   N   S   I   G   E   N   L       1050
AP4    •

AP4    •••  •••  •••  •••  ••A  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  ••G  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••
AP3    CAT CTT ATT GAC CTT GAA CTT AGG ATC AAG CAC CAA GGA TTA GGA GTC CCA GCA GAG CTC ATG GCA CAA ATG TTT GAG GAG GAC AAC AAG   3240
AP3    H   L   I   D   L   E   L   R   I   K   H   Q   G   L   G   Y   P   A   E   L   M   A   Q   M   F   E   E   D   N   K       1080
AP4    •                                                                   E

AP4    •••  •••  ••T  •••  •••  •••  •••  •••  •••  G••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •••  •C
AP3    GAG CAG TCA GAG GAG GGC TTG AGC CTC CTA GTT TCT AGA AAC CTG CTG AGG CTC ATG AAT GGT GAT GTT CGG CAT CTA AGG GAA GCT GGT   3330
AP3    E   Q   S   E   E   G   L   S   L   L   V   S   R   N   L   L   R   L   M   N   G   D   V   R   H   L   R   E   A   G       1110
AP4    •                       D                           K

AP4    •••  •••  •••  •••  C••  •••  •••  •••  •••  •••  •••  •••  •C•  •••  •••  ••G  •••  •••  •••  ••••  •C••CA  TATGGT•A•C  A•ACGG•TTG
AP3    GTG TCA ACC TTC ATC ATC ACC GCT GAA CTT GCT TCC GCT CCA ACA GCA ATG GGA CAA TGA TGAAGC CAGTGGAAGT GTACAACTTA TGGTCATCAA   3420
AP3    V   S   T   F   I   I   T   A   E   L   A   S   A   P   T   A   M   G   Q   STOP                                           1129
AP4    •                                                                       STOP

AP4    T•TCAA•T•C GAGGA•AGCC ••G•GC•AT• GT•AAGTTGG ••ACATT••C CGA•G•G•A• GGAAT•TG•T GCA•CCTGT• •C•GC•TCAG GGTG•AAAGT TGTGC•GCA•
AP3    ATGTTCTGTT TGAATTCCAA GTCACGATAG CCGTGAGCAT TGGTGAGTTG GTGGCATGTT CCTGGGGACG AGGATGGAAT GTGCTGCAGC CTGTAGTGTA GTCTTGCAGC   3547

AP4    AGCT•GG••A T•TAC•••G C•••G•TACG TT••TGCAAT CC••CT••GT •••G••C•C ••TAAT•TTT ••C•TGAATA •AG•••G•AG ATGTACCT•C AC•GTTGGTC
AP3    TTGGTACTTC CGCTGTTGTT ATGTTTCTAT CATCCTACTG TGTAAGGTTC AAGTTTGAAT TTACCATGAA TAAAGTGTGC AGAGTGTACCT GCACTTTGGT CTCAAAAAAA   3668

AP4    TCAAAAAAAA A
AP3    AAAAAAAAAA AAAAAAAAAA AAAAAA
```

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_2-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 molecu-
lar 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_2-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	From sequence Type-3			Type-4			From Vierstra and Quail (31)		
Arg	53			54			58		
Lys	64			64			65		
Asn	45	}	113	43	}	110	-	}	113
Asp	68			67			-		
Gln	49	}	126	49	}	127	-	}	126
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

AMINO ACID RESIDUES
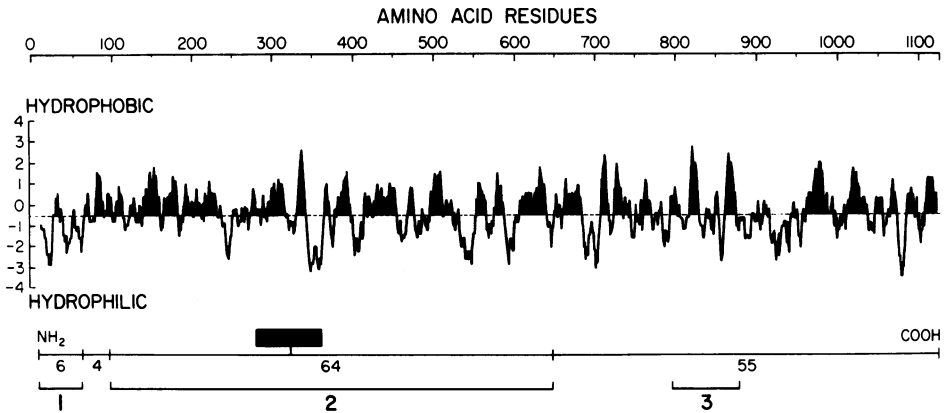
Figure 5. Hydropathy profile of type 3 phytochrome. Solid peaks indicate hydrophobic regions; unshaded troughs indicate hydrophilic regions. A peptide map indicating positions of the chromophore (solid rectangle) and regions containing epitopes for three classes of monoclonal antibody (square brackets, 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 β-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_2-terminus). The NH_2-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 α-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 chro-
mophore in a globular, NH_2-terminal domain. The restriction map polymor-
phism 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 iden-
tified. The nucleotides presumed to be missing from the 5' ends of phy-
tochrome types 3 and 4 sequences by comparison to the sequence of the 5'
end of type 5 phytochrome are currently being sought through further analy-
sis 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 substitu-
tions (~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 rela-
tive 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 heteroge-
neity 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 heterogenous pool of homodimers or a mixture of
heterodimers assembled from a common pool of subunit isoforms. The hetero-
geneity 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 pre-
viously on the basis of indirect data from quantitative amino acid analysis
and spectroscopic studies (36,37). Second, the localization of the chro-
mophore attachment site to the NH_2-terminal domain of the molecule has per-
mitted 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_2-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 con-
sistent 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 pre-
dicts 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_2-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_2-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 pro-
teases (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_2-terminus and 55 kd
from the COOH-terminus. These presumably surface-located regions of the
polypeptide are known to undergo conformational changes upon photoconver-
sion, altering their accessibility to proteolytic attack (4,15,31,38). In
this context it is of interest to consider the NH_2-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 resi-
dues 142-144, would correspond to the NH_2-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 possibi-
lity 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 ~1.1. No region of the phy-
tochrome 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 phycobi-
liproteins currently available in the NBRF database thus far speaks against
a common evolutionary origin for the apoproteins of these photoreceptors.

[+]Present Address: E.I. du Pont de Nemours & Company, Central Research and
Development Department, Experimental Station, Wilmington, DE 19898
[*]Present Address: Plant Breeding Institute, Trumpington, Cambridge, CB2
2LQ England

REFERENCES
1. Shropshire, W. Jr. and Mohr, H., eds. (1983) Encyclopedia of plant phy-
   siology, 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 Haeberli, 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.