Pteridines Vol. 10, 1999, pp. 171 – 177

Isolation and Sequencing of a Plant cDNA Encoding a Bifunctional Methylenetetrahydrofolate Dehydrogenase:Methenyltetrahydrofolate Cyclohydrolase Protein

Liangfu Chen*, Frank E. Nargang and Edwin A. Cossins[§]

Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada

(Received September 7, 1999)

Summary

In plant cells, the interconversion of formyl- and methylene-tetrahydrofolates is catalyzed by a bifunctional protein possessing methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9) and methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5) activities. The present work reports the isolation and sequencing of a cDNA that encodes this protein. Polyclonal antibodies, raised against purified pea cytosolic dehydrogenase:cyclohydrolase, were used to screen a λgt11 cDNA expression library, constructed from leaf extracts of this species. The screen identified a phage containing a cDNA insert with an open reading frame encoding a 294 amino acid protein (M_r 31,344). The deduced primary structure of this protein contained most of the conserved regions found in other dehydrogenase:cyclohydrolase proteins including the corresponding domains of the trifunctional C₁THF synthases of mammalian and yeast origins.

Key words: cDNA, Methylenetetrahydrofolate dehydrogenase, Methenyltetrahydrofolate cyclohydrolase, Bifunctional protein, *Pisum sativum* L.

Introduction

One-carbon metabolism has importance in the formation of purines, thymidylate, serine, methionine and formylmethionyl-tRNA (1, 2). In these syntheses, single-carbon groups are donated by one-

carbon substituted tetrahydrofolate polyglutamates

Studies of the subcellular distribution and struc-

dent DHY-CYC protein (11-14).

⁽H₄PteGlu_n). In yeast and mammalian cells, the formation of 10-HCO-H₄PteGlu and its conversion to 5,10-CH₂-H₄PteGlu are catalyzed by C₁-THF synthase, a trifunctional complex with formyltetrahydrofolate synthetase (SYN, EC 6.3.4.3), methenyltetrahydrofolate cyclohydrolase (CYC), and methylenetetrahydrofolate dehydrogenase (DHY) activities (1, 3). In bacteria these reactions are catalyzed by a mono-functional SYN protein (4, 5) and by a bifunctional DHY-CYC complex (6-9). In eukaryotic cells, C₁-THF synthases occur in the cytosol and mitochondria (1, 10) and the DHY activity is NADP-dependent (1, 3). In contrast, transformed cells and the embryonic and undiffer-entiated tissues of mammals express a unique mitochondrial, NAD-depen-

[§]Address for Correspondence: Dr. Edwin A. Cossins, 99 Fairway Drive, Edmonton, Alberta, T6J 2C2, Canada.

^{*}Present address: Phoenix International Life Sciences Inc., St. Laurent, Quebec, H4R 2N6, Canada.

Enymes: 10-Formyltetrahydrofolate synthetase [Formyltetrahydrofolate ligase],(EC 6.3.4.3); Methylenetetrahydrofolate dehydrogenase (NADP⁺), (EC 1.5.1.5); Methylenetetrahydrofolate dehydrogenase (NAD⁺), (EC 1.5.1.15); Methenyl-tetrahydro-folate cyclohydrolase, (EC 3.5.4.9). Accession numbers: EMBL database: AJ011589; Genbank database: AF030516.

tural organization of plant SYN, CYC and DHY suggest that all three activities are mainly cytosolic with relatively small amounts of each activity being mitochondrial (15). In pea (15, 16) and spinach (17), SYN activity is associated with a protein that lacks DHY and CYC. Furthermore, the SYN protein of spinach leaves has structural homologies with the SYN domains of mammalian and yeast C₁-THF synthases (18). Work in this laboratory (15, 19) also provided evidence that the DHY and CYC activities of pea leaves and cotyledons are associated with a homodimeric protein (subunit M, 38,000) which lacks SYN activity. Attempts to determine the amino acid sequence of this protein were not successful due to an apparent blockage of the N-terminus. In the present work we have therefore isolated and sequenced a cDNA clone from a pea leaf \(\lambda gtl1 \) library that encodes this cytosolic DHY-CYC protein.

Materials and Methods

Chemicals

Reagent-grade chemicals were purchased from Sigma, Fisher Scientific, and BDH Chemicals. Zeta-Probe blotting membranes were supplied by Bio-Rad and E. coli (Y1090r and XL1-Blue) were from Stratagene. ECLTM immunoscreening reagent, HybondTM C-extra nitrocellulose filters, Hybond-N⁺ nvlon filters and ³²P-dCTP were purchased from Amersham Life Science. Promega supplied RNA markers and their RNAgents Total RNA Isolation System. Restriction endonucleases, Taq DNA polymerase and Klenow fragment of DNA polymerase I were from Pharmacia, Bethesda Research Laboratories (BRL) Inc., New England BioLabs, and Boehringer Mannheim. \(\lambda gtll \) expression libraries, constructed from 10 day light-grown (designated L3) or 9 day darkgrown (designated D3) pea leaves were kindly provided by Dr. David Macherel, Centre dEtude Nucleaires, Grenoble, France. The following primers were used: T3 primer, 5'ATTAACCCTCACTAAAG3'; T7 primer, 5'AATACGACTCACTATAG3'; λgtl1 forward primer, 5'GGTGGCGACGACTCCTGGA-GCCCG3'; and \(\lambda\)gtl1 reverse primer, 5'TTGACA-CCAGACCAACTGGTAATG3. Primers were synthesized on a 391 DNA Synthesizer PCR-MATE (ABI Applied Biosystems). Nucleotide sequencing employed an ABI Applied Biosystems 373 DNA Sequencer STRETCH.

Oligolabeling and amplification of DNA fragments

DNA (2-4 µg in 32 µl) was denatured by heating

(5 min at 100°C) and cooling (3 min in a salt-ice bath). To each reaction tube were added 10 μl of oligolabeling buffer (0.25 M Tris-HCl, pH 7.5, 50 mM MgCl₂, 150 mM β-mercaptoethanol, 0.3 mM each of dATP, dGTP and dTTP, 20 U/ml Random Primers) and 5 μl of $^{32}\text{P-dCTP}$ (10 μCi/μl): The labeling reaction was started by addition of 1 μl of Large Fragment of DNA Polymerase I (Klenow fragment, 4.3 U/μl). Incubation was at 37°C for 1-2 hr. Oligolabeled DNA was denatured (as above) before use in hybridization reactions. Restriction enzyme digestion of DNA was according to the suppliers protocol.

Taq PCR was used to amplify DNA fragment inserts in the \(\lambda\gt11\) vector. Phage lysate, (1 \(\mu\l, 50\)-100 ng DNA), was prepared by resuspending a single isolated plaque in 100 µl of suspending medium (100 mM NaCl, 0.2% Mg SO₄·7H₂O, 50 mM Tris-Cl pH 7.5 and 0.01% gelatin), and used as template in a 50 µl reaction volume. The reaction contained 50 mM Tris-HCl (pH 9.0), 1.5 mM MgCl₂, 0.0025% β-mercaptoethanol, 0.01% BSA, 0.2 mM of each of dATP, dCTP, dGTP, dTTP, 0.12 mM of each λgt11 primer, and 2.5 Units of Taq polymerase. PCR was carried out in an AutoGene II Thermal Cycler (Grant Instruments Cambridge Ltd.). The reaction mixture was initially heated to 94°C for 7 min to remove phage protein coats and release DNA, followed by 30 cycles of 1 min at 60°C, 1 min at 72 °C and 1 min at 92°C. A final cycle (1 min at 60°C and 4 min at 72°C) was followed by cooling to 23°C. PCR products were fractionated on 1.2% agarose gels, and DNA bands of interest were excised and purified using NaI and glass milk (20).

Immunoscreening and isolation of cDNA

The L3 Agt11 expression library was initially screened (more than 5×10⁵ pfu, plaque-forming units) using polyclonal antibodies raised against purified pea leaf cytosolic DHY-CYC (15). E. coli (Y1090r⁻) was used as host. Amersham HybondTM C-extra nitrocellulose filters and ECLTM were employed for plaque lifts and immuno-detections. The protocols of Sambrook et al. (21) and those of Amersham Tech Tip 133 were followed. A single positive plaque was identified after tertiary screenings and λDNA from this clone was prepared from a plate lysate (22). After fractionation on a 0.8% agarose gel, \(\lambda\)DNA was purified using glass-milk (20), subjected to EcoRI digestion and the products ligated into pBluescript SK (+/-) (Promega). Ligation products were then transformed into E. coli

AACCACA ACDCCDCAAC

strain XL-1 Blue and grown on agar plates containing 100 μ g/ml ampicillin. Phagemid DNA was isolated from transformants (21) and fractionated on 1% agarose gels after *Eco*RI digestion. A phagemid containing the appropriate insert, designated D λ 2, was sequenced in both directions using T3 and T7 primers.

The cDNA insert from λDNA was ³²P-labeled and used as a probe in further screening of the L3 library (22). Amersham Hybond-N⁺ nylon filters were used in plaque lifts and E. coli strain Y1090r acted as host. The cDNA inserts from the identified clones were then amplified by PCR using \(\lambda gtll \) forward and reverse primers. After fractionation on agarose gels and purification using glass-milk, PCR products were sequenced in both directions using λgtll forward and reverse primers. The longest cDNA insert in one of the positive clones (designated DD7) was subjected to HindIII digestion. One of the fragments contained 377 bp derived from the 5 end. This nucleotide sequence was ³²P-labeled and used to screen the D3 library. The protocols described above for screening, PCR amplification, and sequencing were followed.

RNA Isolation and Northern blot analyses

Pea seedlings were raised in growth chambers for 14 days (23). The leaves were then frozen in liquid N_2 and stored at -70°C until analyzed. Total RNA was isolated using the Promega RNAgents Total RNA Isolation System with the manufacturers technical manual (No. 087) providing the necessary protocol. After ethanol precipitation, RNA pellets were dissolved in nuclease-free water, spectrophotometrically quantified at 260 nm (24) and stored at -20°C. Samples (10-20 μ g) of this RNA were fractionated on HCHO agarose gels and transferred to Bio-Rad Zeta-Probe Blotting membranes (22). ³²P-Labeled restriction fragments of the D λ 2 insert were used as probes to detect mRNA of interest.

Results and Discussion

Isolation of cDNA encoding DHY-CYC protein

Two λgt11 cDNA expression libraries were screened using polyclonal antibodies raised against pea leaf cytosolic DHY-CYC. As DHY-CYC expression is enhanced in greening seedlings (15), the L3 library, constructed from 10 day light-grown seedlings, was screened first. Approximately 5×10⁵ plaques were screened and a single positive plaque was isolated. Phage DNA from this clone was subsequently

| AAGGACA | AGTCGTGAAG 17 | |
|--|--|--|
| TTGTATTTTA ACAGAAAACC ACACCGCACG ATCTCAGTTT | CACCGATCAA 67 | |
| | DD7 | |
| ATG GCC ACC GTA ATC GAC GGC AAA GCC GTT GCA MET Ala Thr Val Ile Asp Gly Lys Ala Val Ala | CAA ACT ATC CGA 112 | |
| TCT GAA ATC GCC GAC GAG GTT CGT CTT CTC TCA Ser Glu Ile Ala Asp Glu Val Arg Leu Leu Ser | | |
| AAG GTT CCT GGA CTT GCA GTG GTG ATA GTA GGA | | |
| Lys Val Pro Gly Leu Ala Val Val Ile Val Gly | Asn Arg Lys Asp | |
| TCT CAA AGC TAC GTT GGA ATG AAG AGA AAA GCG | | |
| Ser Gln Ser Tyr Val Gly MET Lys Arg Lys Ala | | |
| GGA ATC AAA TCC TTC GAT ATC GAC CTT CCG GAG | GAT GCT TCT GAA 292 | |
| Gly Ile Lys Ser Phe Asp Ile Asp Leu Pro Glu | | |
| GCT GAA ATC ATA AAA AAT GTT CAC GAA TTG AAC Ala Glu Ile Ile Lys Asn Val His Glu Leu Asn | | |
| | | |
| GTA CAT GGT ATA TTG GTT CAA CTT CCA TTG CCT Val His Gly Ile Leu Val Gln Leu Pro Leu Pro | | |
| | AAG GAT GTA GAT 427 | |
| GAA GAG AAA GTT TTG ACT GAA ATC AGC ATT TCA Glu Glu Lys Val Leu Thr Glu Ile Ser Ile Ser | Lys Asp Val Asp | |
| GGC TTC CAT CCC TTG AAC ATT GGC AAG CTT GCA | ATG AAA GGA AGA 472 | |
| Gly Phe His Pro Leu Asn Ile Gly Lys Leu Ala | | |
| GAC CCT CTG TTT CTT CCA TGT ACT CCC AAG GCA | TGT CTT GAA CTA 517 | |
| Asp Pro Leu Phe Leu Pro Cys Thr Pro Lys Ala | | |
| TTA TCA CGA AGT GGT GTA AGT ATA AAG GGG AAA | AAG GCT GTT GTG 562 | |
| Leu Ser Arg Ser Gly Val Ser Ile Lys Gly Lys | Lys Ala Val Val | |
| GTT GGT AGA AGC AAC ATA GTT GGA TTA CCA GCT | TCA TTG CTG CTT 607 | |
| Val Gly Arg Ser Asn Ile Val Gly Leu Pro Ala | | |
| CTG AAA GCA GAT GCT ACA GTT ACC ATT GTT CAT | TCA CAC ACA AGT 652 | |
| Leu Lys Ala Asp Ala Thr Val Thr Ile Val His | | |
| CAA CCA GAA ACT ATC ATT CGT GAA GCA GAT ATT | GTT ATT GCA GCA 697 | |
| Gln Pro Glu Thr Ile Ile Arg Glu Ala Asp Ile | | |
| GCA GGA CAG GCA AAG ATG ATC AAG GGA AGC TGG | ATA AAA CCA GGA 742 | |
| Ala Gly Gln Ala Lys MET Ile Lys Gly Ser Trp | | |
| GCT GCA GTG ATA GAT GTT GGC ACA AAT TCT GTG | GAT GAC CCA ACT 787 | |
| Ala Ala Val Ile Asp Val Gly Thr Asn Ser Val | Asp Asp Pro Thr | |
| AGG AAA TCA GGT TAT AGA CTT GTT GGA GAT GTA | GAT TTT GAG GAA 832 | |
| Arg Lys Ser Gly Tyr Arg Leu Val Gly Asp Val | | |
| GCA TCT AAA GTT GCT GGT TGG ATT ACT CCT GTT Ala Ser Lys Val Ala Gly Trp Ile Thr Pro Val | CCT GGT GGT GTG 877 | |
| | | |
| GGT CCA ATG ACA GTC ACA ATG TTG CTG AAG AAT Gly Pro MET Thr Val Thr MET Leu Leu Lys Asn | ACT TTG GAG GGC 922 Thr Leu Glu Gly | |
| | | |
| GCT AAA CGC ACC ATT GAG CAG AAT AAC TAAATAT Ala Lys Arg Thr Ile Glu Gln Asn Asn *** | TCC TTTAGTTTTA 969 | |
| ATAAAAAGAA AGTTACTCTT CGTTTGAGAA TCAGTTTTTG | CTGATAGATC 1019 | |
| AATGTATCTT TATAAAACCG TAAGAATTGG TATTGAAAGT | TGTTTTGGT 1069 | |
| ກາເງິ | | |
| DA2 TCTTCTCGTT CATTATTTTA GTTTCTCCCC CTTATAGTTA | CCTTGTTCTC 1019 | |
| TTTCTTGTAG ACTTATATTG GAATGCCGTC AAGTCTGGTA | AGGCTTCAAA 1169 | |
| | ן דממ | |
| TAGTACACCA AAATAAGGGA ACTATTATGA TGTTAATGTC | | |

Figure 1. Nucleoride and deduced amino acid sequence of a 1.2 kb cDNA (Clone D3-1) encoding cytosolic DHY-CYC. The nucleotide sequence is numbered sequentially from the first nucleotide. The deduced amino acid sequence commences with the first methionine codon. Stop codons are marked by asterisks. This sequence has been assigned accession number AJ011589 (EMBL database) and accession number AF030516 (Genbank). The 5' and 3' ends of the truncated clones, D λ 2 and DD7, are indicated by (Γ) and (Γ) respectively.

isolated and the 830 bp cDNA insert (Dl2) was cloned into pBluescript and its DNA sequence determined (Figure 1). The deduced amino acid sequence encoded by this cDNA was compared to

published sequences of catalytic proteins of other species (8, 25, 26). These comparisons suggested that the cDNA did not correspond to a full length mRNA. The insert of Dλ2 was therefore ³²P-labeled and used as a probe for further screening of the L3 library. Sixteen positive clones were identified and subjected to PCR analyses using \(\lambda gt 11 \) forward and reverse primers. A clone, designated DD7, had the longest cDNA insert. Sequencing of the PCR product revealed a 1105 bp insert (Figure 1). However, the insert in DD7 also appeared to be prematurely truncated. As a result, the 5' region of this clone was used as a probe in further screenings. In this regard, the DD7 insert was amplified by PCR and digested with HindIII to give a 377 bp sequence derived from the 5' end. The D3 library was screened using this sequence. Seven positive plaques were identified and three were found to contain an identical insert of 1219 bp that appeared to encode the full-length amino acid sequence of DHY-CYC. The sequence derived from clone D3-1 is shown in Figure 1.

Clone D3-1 and its deduced amino acid sequence

When Northern blots of pea leaf total RNA were probed with ³²P-labeled Dλ2 insert a single 1.2 kb band was revealed (data not shown). The size of this band was similar to that of the D3-1 insert (1219 bp, Figure 1), supporting the notion that this cDNA represents a full-length clone. Figure 1 also shows that this cDNA encodes a putative amino acid sequence of 294 residues from the first methionine to the downstream stop codon. The molecular weight of this amino acid sequence was calculated to be 31,344 Da with an estimated isoelectric point of 8.46. Our previous studies provided evidence that cytosolic DHY-CYC protein is homodimeric with a native M, of 58,000 (15). It appears likely that the D3-1 insert encodes the cytosolic form of DHY-CYC since the N-terminus of the predicted protein does not contain an obvious mitochondrial or chloroplastic targeting signal. Furthermore, the region immediately following the first met codon corresponds to the mature form of the protein in several other species (Figure 2).

Joshi (27) compared the start sites of 79 plant genes and proposed the consensus sequence (TAAA-CAATGGCT) for initiation of translation. The corresponding sequence around the start codon of the D3-1 insert (Figure 1) is GATCAAATGGCC, with 7 out of 12 bases identical to the proposed consensus initiation sequence. When codon usage in the D3-1 insert was examined it was found that 37% of codons end in either G or C. This is consistent with

```
MATVIDGKAV. AQTIRSEIA. DEVRLLSQRYGKVPGLAVVIVGNRKDSQ
MAPAEILINGKEISAGIRARIKN. QVTQLKEQVPGFTPRLAILQVGNRDDSN
MAPAGIINGKVVSAQIRNLIKT. QFTQMQEQVPGFTPGLAILQVGDRDDSN
SQYHLISGRKI. AQSIREKANDEI ÇAIKLKHEMFKPILKI IQVGAREDSS
MAGQVLDGKAC. AQQFRSNIAMEIKSIQGHVPGFNPALAII IQVGNREDSA
MAAKIIDGKTI. AQQVESEVA. QKVQARIAAGLRAPGLAVVIVGSNPASQ
NEAVVISGRKI. AQQIKQEVR. QEVEEWVASGNRPHLSVILVGENPASQ
NEAVVISGRKI. AQQIKQEVR. QEVEEWVASGNREPHLSVILVGENPASQ
NEAVVISGRKI. AQQIKQEVR. EXPERVASGNREPHLSVILVGENPASH
                                 NEAVVISGRKL.AQQIKQEVQ.QEVEEWVASGNKRPHLSVILVGDNPASH
                                  SYVGMRRKACAELGIKSFDIDLPEDASEAEIIKNVHELNANPDVHGILVÇ
                                SYVGMRRRACAELGIKSFDIOLPEDASEAEIIKNYHELNANPOVHGILVQ
LYINVKLKAAEIGIKATHIKLPRTTESESVWKIYISINDESSYWHGILV
LYINVKLKAAQEIGIKATHIKLPRTSESSVKKYVISINDESSYWHGILV
LYINVKLKASKOSNYDCIIEKLPREITEVELLKKISJINDDSIHOLTU
TYVRMKRAAEEAGIVANFIHLDESATEFEVLRYVDQLINEDPHTHGIIVQ
IYVASKRRACEEVGFVSRSYDLPETTSEAELLELIDTINADNTIDGILVQ
SYVLNKTRAAAVVGINSETIKKPASISEEELLINJIKHNDDNYDGLLVQ
SYVLNKTRAAAVVGINSETIKKPASISEEELLINJIKKDDNYDGLLVQ
                                                                                                                                                                                          100
131
99
98
130
132
PEA DC
HUM SDC
RAT SDC
YMI SDC
YCY SDC
ECO DC
HUM DC
MOU DC
                                 LPLPKH..VNEEKVLTEISISKDVDGFHPLNIGKLAMKGRDPLFLPCTPK
                                                                                                                                                                                          150
150
                                 LPLDSENSINTEEVINAIAPEKDVDGLTSINAGRLARGDLNDCFIPCTPK
                                 LPLDSENSINTEAVINAIAPEKDVDGLTSINAGKLARGDLKDCFIPCTPK
                                LPLDSENSINTEAVINAIAPEKDVDGLTSINAGKLARGDLKOCFIPCTDK
LPLPRH. .LDETTITNAVPKKDVDGFHRVNAGELAKKGGKPYFIPCTDY
LPLPAH. .LDEDRITSRVLAEKDVDGFGPTNIGELNKKNGHPFFLPCTDK
LPLPAH. .IDEDRITSRVLAEKDVDGFHPYNVGRLCQRAPR. .LRPCTDR
LPLPEH. .IDERRICHAVSFDKDVDGFHVINVGRMCLDQYS. .MLPATDW
LPLPEH. .IDERKVCNAVSFDKDVDGFHVINVGRMCLDQYS. .MLPATDW
****
PEA DC
HUM SDC
RAT SDC
YMI SDC
YCY SDC
ECO DC
HUM DC
MOU DC
                                                                                                                                                                                          187
192
192
221
                                 ACLELLSRSGVSIKGKKAVVVGRSNIVGLPASLLLLKA......
GCLELIKETGVPIAGRHAVVVGRSKIVGAPMHDLLLWN.....
                                  GCLELIKETGVOIAGRHAVVVGRSKIVGAPMHDLLLWN...
                                  GCMKLLEEAHVKLDGKNAVVLGRSSIVGNPIASLLKNA...
                                  186
PEA DC
HUM SDC
RAT SDC
YMI SDC
YCY SDC
ECO DC
HUM DC
MOU DC
                                 TIVHSHTS...QPETIIREADIVIAAAGQAKMIKGSWIKP....
TTCHSKTA...HLDEEVNKGDILVVATGQPEMVKGEWIKP....
                                  TTCHSKTA..DLDKEVNKGDILVVATGQPEMVKGEWIKP...
                                  TVCHSHTR. NIAEVVSOADIVIAACGIPOYVKSDWIKE.
                                                                                                                                                                                          261
237
                                 TITHSKTR. DIASYLHDADIVVVAIGQPEFVKGEWFKPRDGTSSDKKTV
TYTHRFYK. NIRHHLENADLLIVAVGKPGFIPGDWIKE. GAI
TISHRYTPKEQLKKHTILADIVISAAGIPNLITADMIKE. GAA
                                  VIDVGTNSVDDPTRKSGYRLVGDVDFEEASKVAGMITEVPGGVGPMTVTM
VIDCGINYVPDDKKPNGRKVVGDVAYDEAKERASFITEVPGGVGPMTVAM
VIDVGINYVPDDTKPNGRKVVGDVAYDEAKEKASFITEVPGGVGPMTVAM
VIDVGINYVPDDTKPNGRKVVGDVAYDEAKEKASFITEVPGGVGPMTVAM
VIDVGTNYVADPSKKSGFKCVGDVEFNEAIKVHLITEVPGGVGPMTVAM
VIDVGTNYVADPSKKSGFKCVGDVEFNEAIKVHLITEVPGGVGPMTVAM
VIDVGINRLENG...KVVGDVYFEDAAKRASYITEVPGGVGPMTVAT
PEA DC
HUM SDC
RAT SDC
YMI SDC
 YCY SDC
                                  VIDVGINRVHDPVTAKP, KLVGDVDFEGVROKAGYITPVPGGVGPMTVAM
                                  VIDVGINRVQDPVTAKP. KLVGDVDFEGVKKKAGYITPVPGGVGPMTVAM
                                  LMONTLIAAKROMEESSKPL.QIPPLPLKLLTP
LIENTLQACVEYHDPQDE!
LMKNTIIAAKKVLRLEEREVLKSKELGVATN!
                                  LMKNTI I AAKKVLRPEELEVFKSKQRGVATN
```

Figure 2. A comparison of the deduced amino acid sequences of eight different proteins exhibiting folate-dependent dehydrogenase and cyclohydrolase activities. Protein sources are: HUM, human; YMI, yeast mitochondria; YCY, yeast cytoplasm; ECO, *E. coli*; and MOU, mouse. Enzyme activities are: S, Formyltetrahydrofolate synthetase; D, Methylenetetrahydrofolate dehydrogenase; C, Methenyltetrahydrofolate cyclohydrolase. Stop codons are indicated by (!). Numerals indicate amino acid positions as reported in the data bank. Amino acids common to all eight sequences are indicated by subscript asterisks. Amino acids present in all seven sequences but absent in the pea sequence are indicated by a subscript (^).

the observation that nuclear genes of dicotyledonous plants have a preference for codons ending in an A or U (28).

In yeast, cytosolic C_1 -THF synthase is encoded by the *ADE-3* gene (25, 29, 30), whereas the mitochondrial isoform is encoded by the *MIS1* gene (1, 31). The subunits of both isoforms are approximately 110 kDa and are immunologically closely related (31). Western blot analyses (15) suggest that

the cytosolic and mitochondrial forms of pea DHY-CYC are of similar subunit size (38 kDa) and both show cross reactivity to the same polyclonal antibodies. Conceivably, the mitochondrial activity is also encoded by a distinct gene in this higher plant species. Isolation of cDNA encoding this protein is needed to elucidate this possibility.

DHY-CYC homologies in plant, bacterial, yeast and mammalian proteins

Two databases (EMBL-24 and GenBank 65) were searched for the primary structures of characterized DHY-CYC proteins and C₁-THF synthases. Synthase sequences were obtained for human (32), rat (34), yeast mitochondria (26), and yeast cytosol (25), respectively. In addition, the databases had sequences for the DHY-CYC proteins of E. coli (8), mouse (13) and human cells (14). These sequences were aligned with the deduced sequence pea DHY-CYC protein as shown in Figure 2. For the trifunctional enzymes, only the DHY and CYC domains were considered. The alignment shown is a composite of the PC/GENE Clustal program with further refinement done by inspection. Figure 2 shows that there are several large blocks of similarity that were common to all eight sequences. In earlier studies, D'Ari and Rabinowitz (8) compared the amino acid sequence of E. coli DHY-CYC with the other six sequences noted above. These workers reported a ITPVPGGVGPMTVA sequence in all seven proteins with no substitutions or gaps. This sequence was not found in any other protein contained in the databases examined (8). The alignment shown in Figure 2 suggests that pea DHY-CYC contains this consensus sequence (residues 263-276) with the exception that the last amino acid A, is replaced by T.

The degree of identity of pea DHY-CYC, relative to each of the other seven sequences was examined by the Palign program of PC/GENE (Table 1). The mono-functional, NAD-dependent methylenetet-

Table 1. Percentage identity shown between the amino acid sequence of pea DHY-CYC and the corresponding enzymes or domains of other species.

| Source | % Identity |
|--|------------|
| Human C ₁ -THF synthase | 49.0 |
| Human bi-functional DHY-CYC | 46.3 |
| Rat C ₁ -THF synthase | 49.0 |
| Mouse bi-functional DHY-CYC | 48.3 |
| Yeast mitochondrial C ₁ -THF synthase | 46.3 |
| Yeast cytosolic C ₁ -THF synthase | 52.4 |
| Yeast NAD-dependent dehydrogenase | 25.2 |
| E. coli bi-functional DHY-CYC | 49.0 |

rahydrofolate dehydrogenase of yeast (34,35) was included in this comparison. The percentage identities ranged from 46% to 52%, except for the yeast monofunctional dehydrogenase protein. This is perhaps surprising because the mitochondrial DHY-CYC proteins of human and mouse are NAD-specific while the other proteins all have NADP-dependencies. The highest percentage identity (52%) was obtained when the pea protein was compared with the yeast cytosolic C₁-THF synthase. As expected, the lowest identity (25%) was observed when the plant protein was compared to the yeast monofunctional dehydrogenase. Since the DHY and CYC activities of the C₁-THF synthase are kinetically dependent (36), and probably share a common folatebinding site (37), it follows that the yeast monofunctional dehydrogenase will lack amino acid residues required for CYC activity.

Acknowledgments

This work was supported by grants to E.A.C and E.E.N. from the Natural Sciences and Engineering Research Council of Canada We thank Dr. David Macherel, Centre dEtude Nucleaires, Grenoble, France, for kindly providing the pea leaf λgt11 expression libraries used in this study. We also thank Gary Ritzel and Patricia Murray, Department of Biological Sciences, University of Alberta for excellent technical assistance.

References

- Appling DR. Compartmentation of folate-mediated one-carbon metabolism in eukaryotes. FASEB J 1991;
 2645-2651.
- Cossins EA, Chen L. Folates and one-carbon metabolism in plants and fungi. Phytochemistry 1997; 45: 437-452.
- MacKenzie RE. Biogenesis and interconversion of substituted tetrahydrofolates. In: Blakley RL, Benkovic, SJ, eds. Folates and Pterins. New York: Wiley-Interscience, 1984; 1: 255-306.
- 4. Whitehead TR, Rabinowitz JC. Nucleotide sequence of the Clostridium acidiurici_(Clostridium acidi-urici) gene for 10-formyltetrahydrofolate synthetase shows extensive amino acid homology with the trifunctional enzymes C₁-tetrahydrofolate synthase from Saccharomyces cervisiae. J Bacteriol 1988; 170: 3255-3261.
- Lovell CR, Przybyła A, Ljungdahl LG. Primary structure of the thermostable formyl-tetrahydrofolate synthetase from *Clostridium thermoaceticum*. Biochemistry 1990; 29: 5687-5694.
- Dev IK, Harvey RJ. A complex of N⁵, N¹⁰-methylenetetrahydrofolate dehydrogenase and N⁵, N¹⁰-methenyltetrahydrofolate cyclohydrolase in *Escherichia coli*: Purification, subunit structure, and allosteric inhibition

- by N^{10} -formyltetrahydrofolate. J Biol Chem 1978; 253: 4245-4253.
- Ljungdahl LG, OBrien WE, Moore MR, Liu MT. Methylenetetrahydrofolate dehydro-genase from Clostridium formicoaceticum and methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase (combined) from Clostri-dium themoaceticum. Methods Enzymol 1980; 66: 599-609.
- D'Ari L, Rabinowitz JC. Purification, characterization, cloning, and amino acid sequence of the bifunctional enzyme 5,10-methylenetetrahydrofolate dehydrogenase/ 5,10-methenyltetrahydrofolate cyclohydrolase from Escherichia coli. J Biol Chem 1991; 266: 23953-23958.
- Pawelek PD, MacKenzie RE. Methylenetetrahydrofolate dehydrogenase-cyclohydrolase from *Photobacterium phosphoreum* shares properties with a mammalian mitochondrial homologue. Biochim Biophys Acta 1996; 1296: 47-54.
- Wagner C. Symposium on the subcellular compartmentation of folate metabolism. J Nutri 1996; 126: 1228S-1234S.
- Mejia NR, MacKenzie RE. NAD-dependent methylenetetrahydrofolate dehydro-genase-methenyltetrahydrofolate cyclohydrolase from ascites tumor cells. J Biol Chem 1986; 261: 9509-9513.
- Mejia NR, MacKenzie RE. NAD-dependent methylenetetrahydrofolate dehydro-genase-methenyltetrahydrofolate cyclohydrolase in transformed cells is a mitochondrial enzyme. Biochem Biophys Res Commun 1988; 155: 1-6.
- Belanger C, MacKenzie RE. Isolation and characterization of cDNA clones encoding the murine NAD-dependent methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase. J Biol Chem 1989; 264: 4837-4843.
- Peri KG, Belanger C, MacKenzie RE. Nucleotide sequence of the human NAD-dependent methylenetetrahydrofolate dehydrogenase-cyclohydrolase. Nuc Acid Res 1989; 17: 8853.
- Chen L, Chan SY, Cossins EA. Distribution of folate derivatives and enzymes for synthesis of 10-formyltetrahydrofolate in cytosolic and mitochondrial fractions of pea leaves. Plant Physiol 1997; 115: 299-309.
- Kirk CD, Imeson HC, Zheng L, Cossins EA. The affinity of pea cotyledon 10-formyltetrahydrofolate synthetase for polyglutamate substrates. Phytochemistry 1994; 35: 291-296.
- Nour JM, Rabinowitz JC. Isolation, characterization, and structural organization of 10-formyltetrahydrofolate synthetase from spinach leaves. J Biol Chem 1991; 266: 18363-18369.
- Nour JM, Rabinowitz JC. Isolation and sequencing of the cDNA coding for spinach 10-formyltetrahydrofolate synthetase. J Biol Chem 1992; 267: 16292-16296.
- 19. Kirk CD, Chen L, Imeson HC, Cossins EA. A 5,10-methylenetetrahydrofolate dehydrogenase:5,10-methenyltetrahydrofolate cyclohydrolase protein from *Pisum sativum*. Phytochemistry 1995; 39: 1309-1317.
- Volgelstein B, Gillespie D. Preparative and analytical preparation of DNA from agarose. Proc Nat Acad Sci USA. 1979; 76: 615-619.

- Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: A laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1989.
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. Current Protocols in Molecular Biology. New York: Wiley-Interscience 1992.
- Imeson HC, Zheng L, Cossins EA. Folylpolyglutamate derivatives of *Pisum sativum* L.: Determination of polyglutamate chain lengths by high performance liquid chromatography following conversion to p-aminobenzoylpolyglutamates. Plant Cell Physiol 1990; 31: 223-231.
- Wilkinson M. Purification of RNA. In: Essential Molecular Biology: A practical approach. In: Brown TA, ed. New York: Oxford University Press, 1991; 1:74-86.
- Staben C, Rabinowitz JC. Nucleotide sequence of the Saccharomyces cerevisiae ADE3 gene encoding C₁-tetrahydrofolate synthase. J Biol Chem 1986; 261: 4629-4637.
- Shannon KW, Rabinowitz JC. Isolation and characterization of the *Saccharomyces cerevisiae MIS1* gene encoding mitochondrial C₁-tetrahydrofolate synthase. J Biol Chem 1988; 263: 7717-7725.
- 27. Joshi CP. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nuc Acid Res 1987; 15: 6643-6653.
- Campbell WH, Gowri G. Codon usage in higher plants, green algae, and cyanobacteria. Plant Physiol 1990; 92: 1-11
- 29. Appling DR, Rabinowitz JC. Evidence of overlapping active sites in a multifunctional enzyme: Immunochemical and chemical modification studies on C₁-tetrahydrofolate synthase from *Saccharomyces cerevisiae*. Biochemistry 1985; 24: 3540-3547.
- Barlowe CK, Appling DR. Molecular genetic analysis of Saccharomyces cerevisiae C₁-tetrahydrofolate synthase mutants reveals a noncatalytic function of the ADE3 gene product and an additional folate-dependent enzyme. Mol Cell Biol 1990; 10: 5679-5687.
- Shannon KW, Rabinowitz JC. Purification and characterization of a mitochondrial isozyme of C₁-tetrahydrofolate synthase from *Saccharomyces cerevisiae*. J Biol Chem 1986; 261: 12266-12271.
- 32. Hum DW, Bell AW, Rozen R, MacKenzie RE. Primary structure of a human trifunctional enzyme: Isolation of a cDNA encoding methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase. J Biol Chem 1988; 263: 15946-15950.
- 33. Thigpen AE, West MG, Appling DR. Rat C₁-tetrahydrofolate synthase: cDNA isolation, tissue-specific levels of the mRNA, and expression of the protein in yeast. J Biol Chem 1990; 265: 7907-7913.
- Barlowe CK, Appling DR. Isolation and characterization of a novel eukaryotic mono-functional NAD⁺dependent 5,10-methylenetetrahydrofolate dehydrogenase. Biochemistry 1990; 29: 7089-7094.
- 35. West MG, Barlowe CK, Appling DR. Cloning and characterization of the *Saccharomyces cerevisiae* gene encoding NAD-dependent 5,10-methylene-tetrahydrofolate dehydrogenase. J Biol Chem 1993; 268: 153-

160.

- 36. Drummond D, Smith S, MacKenzie RE. Methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase from porcine liver: Evidence to support a common dehydrogenase-cyclohydrolase site. Can J Biochem Cell Biol
- 1983; 61: 1166-1171.
- 37. Smith DDS, MacKenzie RE. Methylenetetrahydrofolate dehydrogenase-methenyl-tetrahydrofolate-cyclohydrolase-formyltetrahydrofolate synthetase: Affinity labeling of the dehydrogenase-cyclohydrolase active site. Biochem Biophys Res Commun 1985; 128: 148-154.