Volume 285, number 1, 85–88 FEBS 09876 July 1991
© 1991 Federation of European Biochemical Societies 00145793/91/$3.50
ADONIS 001457939100582R

Divergent evolution of chloroplast-type ferredoxins

Shigeaki Harayama, Alessandra Polissi and Monique Rekik

Department of Medical Biochemistry, University Medical Center, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland

Received 30 March 1991; revised version received 24 April 1991

The TOL plasmid pWW0 of *Pseudomonas putida* encodes a set of enzymes required for the oxidation of toluene to Krebs cycle intermediates. The structural genes for these enzymes are encoded in two operons which comprise the *xylCMABN* and *xylXYZLTEGFJQKIH* genes, respectively. The function of the *xylT* gene has not yet been identified. The nucleotide sequence of *xylT* was determined in this study and putative gene product was shown to contain a sequence characteristic for chloroplast-type ferredoxins. The *nahT* gene, the homologue of *xylT*, present on NAH plasmid NAH7 encoding naphtalene-degrading enzymes, was also sequenced. The sequence conservation between *xylT* and *nahT* strongly suggests that both gene products have some physiological function. Chloroplast-type ferredoxins have been discovered in photosynthetic organisms (plants, algae, cyanobacteria and *Rhodobacter*) and *Halobacterium* species. Furthermore, chloroplast-type ferredoxin-like sequences have been found in the electron-transfer components of some oxygenases. The sequences of XylT and NahT were compared with those of the previously identified chloroplast-type ferredoxins, in order to examine their evolutionary relationships.

TOL plasmid; NAH plasmid; Ferredoxin; Chloroplast-type; Protein evolution

1. INTRODUCTION

TOL plasmid pWW0 and NAH plasmid NAH7 have originally been discovered in *P. putida* strains mt-2 and PpG7, and encode catabolic functions for the mineralization of toluene and naphthalene, respectively [1–3]. In these catabolic pathways, the initial substrates, namely toluene and naphthalene, are transformed into catechol, which is further oxidized to Krebs cycle intermediates. In our previous study, the gene order of the *meta* operon of TOL plasmid pWW0, which encodes the enzymes necessary for the mineralization of benzoate via catechol has been determined to be: *xylXYZLTEGFJQKIH* [4]. The *xylXYZL* genes are required for the transformation of benzoate to catechol while the *xylEGFJKIH* genes are required for the oxidation of catechol to Krebs cycle intermediates (Fig. 1). The function of the *xylT* product is not yet known. The catabolic genes on the NAH7 plasmid have also been chaacterized. The *sal* operon on the plasmid encodes *nahGHINLJK*; the *nahG* gene is the structural gene for salicylate hydroxylase while *nahH, nahI, nahN, nahL, nahJ* and *NahK* are isofunctional to *xylE, xylG, xylF, xylJ, nahH* and *xylI*, respectively. Extensive sequence homology exists between *xylE* and *nahH* and between *xylG* and *nahI* [5,6] Salicylate hydroxylase, encoded by *nahG*, catalyzes a reaction completely different from that catalyzed by the *xylL* product (Fig. 1), and no DNA sequence similarity exists between *xylL* and *nahG* [6a] (Harayama, S. and Rekik, M., unpublished result). We have, however, found that homologous recombination could occur between the *xylT* region on pWW0 and the upstream region of *nahH* on NAH7 (Harayama and Rekik, unpublished result). This observation suggests that the NAH7 plasmid encodes a gene analogous to *xylT* in a region between *nahG* and *nahH* (Fig. 2). The evolutionary conservation of the *xylT*-like sequence in the NAH7 plasmid indicates that *xylT* and its homologue on the NAH7 plasmid named *nahT*, play some physiological role in the metabolism of catechol. In this study, we sequenced the DNA regions of *xylT* and *nahT*.

Correspondence address: S. Harayama, Department of Medical Biochemistry, University Medical Center, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland. Fax: (41) (22) 473300.

2. MATERIALS AND METHODS

2.1. *DNA sequencing and analysis*

*Sau*3A-generated fragments of pGSH2939 carrying the *nahT* sequence were cloned into M13 bacteriophage derivatives, tg130, tg131 and mp18 as described previously [5]. DNA sequencing using M13 derivatives was carried out by the established method [7]. Klenow fragment was used for the sequencing of the M13 derivatives. Two deletion derivatives of pGSH2939, Δ(*Xba*I-*Nco*I) and Δ(*Nco*I-*Eco*RI), were sequenced using T7 polymerase according to instructions by Pharmacia. The *xylT* region of the pWW0 DNA was subcloned from pGSH3042 [5] into pGEM-7Zf(−) [8] from which nested deletions were obtained using exonuclease III according to Henikoff [9]. These plasmids were sequenced by T7 polymerase. The alignment of *xylT* and *nahH* was carried out manually whereas the multiple alignment of chloroplast-type ferredoxins was done with the CLUSTAL program [10] in the PC/GENE software package (developed by A. Bairoch and available from IntelliGenetics).

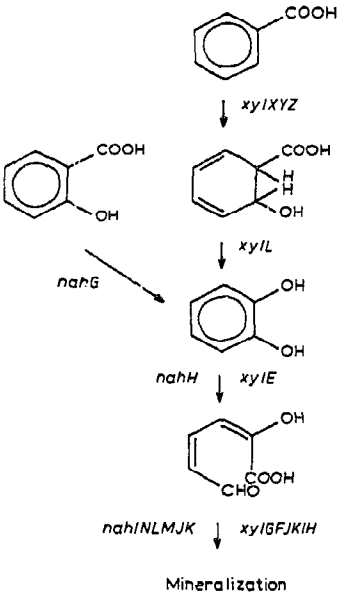

Fig. 1. Parts of the catabolic pathways encoded by TOL plasmid pWW0 and NAH plasmid NAH7. Enzymatic steps involved in the transformation of benzoate into 2-hydroxymuconic semialdehyde in the TOL catabolic pathway, and those involved in the transformation of salicylate into 2-hydroxymuconic semialdehyde in the NAH catabolic pathway are presented.

3. RESULTS AND DISCUSSION

The strategy for DNA sequencing in the region between *xylL*, the structure gene for 1,2-dihydroxycyclohexa-2,4-diene carboxylate dehydrogenase, and *xylE*, the structural gene for catechol 2,3-dioxygenase is shown in Fig. 3. As shown in Fig. 4, this region contains an open-reading frame which allows the synthesis of a polypeptide of 12 034 daltons. A putative initiation codon for *xylT* is preceded by a Shine-Dalgarno-like sequence, GGA. The initiation codon of *xylE* overlaps with the TGA stop codon of *xylT*. The DNA region upstream of *nahH* was determined by the strategy shown in Fig. 3 and its nucleotide sequence and a possible gene product from this region of DNA is presented

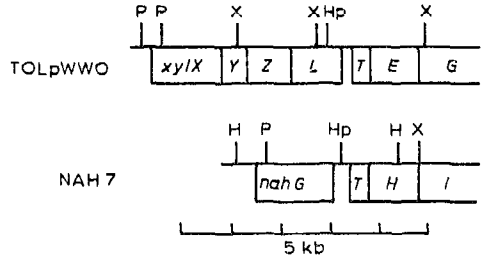

Fig. 2. Order of the catabolic genes on TOL plasmid pWW0 and NAH plasmid NAH7. The functions of these genes are indicated in Fig. 1, except *xylT* and *nahT* whose function is not yet known. Restriction sites are: P, *Pst*I; X, *Xho*I; Hp, *Hpa*I; and H, *Hin*dIII. Boxes represent gene locations.

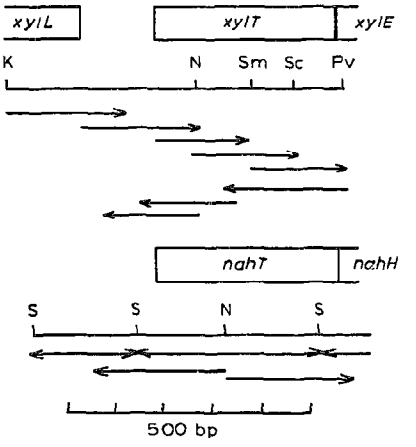

Fig. 3. Strategies for the sequencing of *xylT* and *nahT*. Restriction sites are: K, *Kpn*I; N, *Nco*I; Sm, *Sma*I; Sc, *Sac*I; Pv, *Pvu*II; and S, *Sau*3A.

```
nahT    GATCTAAGCATGGGCGGCCTACCGACCCTAGCGTCAGTGGGCCGCAGCCACAACGGCAAC
xylT    TCGACAGCAGCCTTATGAAACGCTACGGAAGCATTGACGAGCAGGTCGAGGCAATTCTGT

nahT    CTAATATGGTTTCCCTTGCATGCTGCGCCGCAGAGCGTCGCTGGAGATAACAATGCACAA
xylT    TCCTTGCTTCTGACGCCGCCTCCTACATCACCGGTATAACTCTTCCGGTGGCAGGGGGAG

nahT    CAATAACAAGATGCTTTGCTCGTTCCGTTCCACCTTTACCGTCAGTGCTCTGGCATTGGC
xylT    ACCTCGGCTGCCAGAGCTGTTCCGTCATGTTTAGTGTATCTGGGTGACCCTTATGTCTGG

nahT    CCTCTGCATTTCCCTGGCAAGAGCGGTCAAGGCGGCCGAACCGATCAAGATTGGTCTTTT
xylT    TTAAGAAGAGAAATCGACATGCGAAGAAGCAACGTACATAAGACCCCTGAGGCTCATTTT

nahT    GATTTCCGACAGCGGCATCTTCGCGGCCCCGAGTGAAGCTACTCCGTTTCTGCCACGGTT
xylT    CGGGGTTATGGCGGCATCACCCAGAGCTGTTGGGGGATACTTCCGTCATGTTTAGTGTAT

nahT    GAGCGGAGGGAGATATGTCAGAGGTCTTTGAAATCACTGTGCAGCCTGGTGGAGAGCGCT
xylT    CTGGGATGAATATGAACAGTGCCGGCTACGAGGTGTTCGAAGTGCTAAGCGGCCAGTCAT
NahT                  M  S  E  V  F  E  I  T  V  Q  P  G  G  E  R
XylT               M  N  S  A  G  Y  E  V  F  E  V  L  S  G  Q  S

nahT    TTGTCTGTCAGCCTCAGCAATCAGCGTTGCATGCCATGGAGACGCAGGGCAAGCGCTGCT
xylT    TCCGCTGTGCGGAGGGCCAGTCGGTACTGCGCGCCATGGAAGCCCAGGGCAAGCGCTGCA
NahT    F  V  C  Q  P  Q  Q  S  A  L  H  A  M  E  T  Q  G  K  R  C
XYLT    F  R  C  A  E  G  Q  S  V  L  R  A  M  E  A  Q  G  K  R  C

nahT    TACCTGTGGGCTGTCGCGGCGGCGGTTGTGGCTTGTGCAAGGTGAGGGTGCTTGCCGGTG
xylT    TACCGGTGGGCTGTCGCGGTGGCGGTTGCGGCCTTTGTAGAGTGCGGGTGCTCAGCGGAG
NahT    L  P  V  G  C  R  G  G  G  C  G  L  C  K  V  R  V  L  A  G
XylT    I  P  V  G  C  R  G  G  G  C  G  L  C  R  V  R  V  L  S  G

nahT    ACTACGAGAGCGGGCGCGTGAGCTGCAAGCACCTACCGGTAGAAGCACGCGAACAAGGCT
xylT    CCTACCGGAGCGGACGCATGAGCCGCGGTCACGTGCCGGCCAAGGCCGCCGCCGAAGCGC
NahT    D  Y  E  S  G  R  V  S  C  K  H  L  P  V  E  A  R  E  Q  G
XylT    A  Y  R  S  G  R  M  S  R  G  H  V  P  A  K  A  A  A  E  A

nahT    ATGCCTTGGCCTGCCGACTGTTTGCCCGCAGCGATCTTTGTATCGAGCGTTACTCAAAGC
xylT    TGGCCCTGGCCTGTCAAGTGTTTCCGCAAACCGACTTGACCATCGAGTACTTTCGCCACG
NahT    Y  A  L  A  C  R  L  F  A  R  S  D  L  C  I  E  R  Y  S  K
XylT    L  A  L  A  C  Q  V  F  P  Q  T  D  L  T  I  E  Y  F  R  H

nahT    CGTGCAGTGAAAGTACGGTCGACCAACAACAAAGAGAATAGGTGATTTCATG
xylT    TTGGCGGAAACAAACCTGACAACATGAACTATGAAGAGGTGACGTCATGA
NahT    P  C  S  E  S  T  V  D  Q  Q  Q  R  E  -
XylT    V  G  G  N  K  P  D  N  M  N  Y  E  E  V  T  S  -
```

Fig. 4. The nucleotide sequences of *xylT* and *nahT* and the amino acid sequences of the putative *xylT* and *nahT* products. Shine-Dalgarno-like sequences for the *xylT* and *nahT* genes are underlined. The ATG sequences present at the end of *xylT* and *nahT* are the initiation codons for *xylE* and *nahH*, respectively.

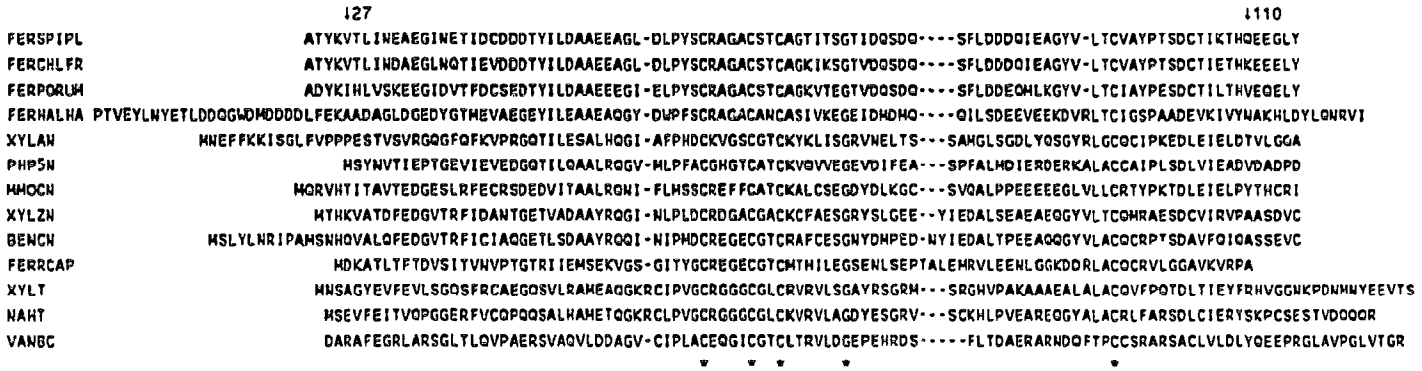

Fig. 5. Alignment of the XylT and NahT sequences with other chloroplast-type ferredoxins. (–) = no corresponding amino acid, (*) = identical amino acid. Sources of the amino acid data are: ferredoxin from *Spirulina platensis* (FERSPIPL) [19], ferredoxin from *Chlorogloepsis fritschii* (FERCHLFR) [20], ferredoxin from *Porphyra umbilicalis* (FERPORUM) [21], ferredoxin from *Halobacterium halobium* (FERHALHA) [22], the amino-terminal sequence of XylA, an electron transfer component of xylene mono-oxygenase [16], the amino-terminal sequence of Polypeptide 5 of phenol hydroxylase (PHP5N) [15], the amino-terminal sequence of the electron transfer component of methane mono-oxygenase [14], the amino-terminal sequence of XylZ, an electron transfer component of toluate 1,2-dioxygenase (XYLZN, Harayama et al., unpublished), the amino-terminal sequence of BenC, an electron transfer component of benzoate 1,2-dioxygenase (BENCN, Neidle et al., submitted), ferredoxin from *Rhodobacter capsulatus* (FERRCAP) [12] and carboxyl-terminal sequence of VanB, an electron transfer component of vanillate decarboxylase (VANBC) [13]. The fraction of homologous sequences in the aligned sequences between coordinates 27 and 110 was counted for all pairs of the sequences, and summarized in Table I.

in Fig. 4. The gene encoded upstream of *nahH* was named *nahT*. The *nahT* region has also been sequenced by You et al. [6a]. The nucleotide sequence similarity between *xylT* and *nahT*, and the amino acid sequence similarity between XylT and NahT are evident in Fig. 4. Inspection of the amino acid sequences of XylT and NahT revealed the presence of the sequences characteristic to chloroplast-type ferredoxins, namely Cys-XXXX-Cys-XX-Cys [10]. This class of ferredoxins contains one [2Fe-2S] cluster, and the four non-variant cysteine residues bind two iron atoms. These ferredoxins have originally been found in photosynthetic organisms, namely plants, algae and cyanobacteria. The chloroplast-type ferredoxins are also found in non-photosynthetic archaebacteria, such as the *Halobacterium* species [11]. More recently, a chloroplast-type ferredoxin was found in *Rhodobacter capsulatus*, a photosynthetic bacterium [12]. Furthermore, the sequences similar to those of the chloroplast-type ferredoxins have been found in the electron transfer components of some oxygenases, namely in the carboxyl-terminal sequence of the VanB component of vanillate decarboxylase [13], and in the amino terminal of the MmoC component of methane mono-oxygenase [14], of Polypeptide 5 of phenol hydroxylase [15], of the XylA component of xylene mono-oxygenase [16], of the BenC component of benzoate dioxygenase [17] (Neidle et al., submitted), and of the XylZ component of toluate dioxygenase [18] (Harayama et al., manuscript in preparation). Therefore, it would seem that these oxygenase subunits have emerged by fusion of ancestral chloroplast-type ferredoxins with other proteins. The alignment of XylT and NahT with other chloroplast-type ferredoxins is presented in Fig. 5, and the fraction of homologous amino acids is presented in Table I.

Table I

Percent homologies between chloroplast-type ferredoxins

	FERSPIPL	FERCHLFR	FERPORUM	FERHALHA	PHP5N	MMOCN	XYLAN	XYLZN	BENCN	FERRCAP	XYLT	NAHT	VANBC
FERSPIPL	100												
FERCHLFR	89	100											
FERPORUM	70	67	100										
FERHALHA	37	40	40	100									
XYLAN	28	31	27	28	100								
PHP5N	18	22	22	21	32	100							
MMOCN	22	22	26	25	31	22	100						
XYLZN	24	23	25	23	27	32	33	100					
BENCN	20	20	20	21	26	29	35	63	100				
FERRCAP	20	21	12	24	20	26	18	25	26	100			
XYLT	19	19	22	24	28	24	23	26	27	20	100		
NAHT	19	17	20	20	32	25	29	28	24	19	55	100	
VANBC	18	20	18	22	22	17	11	17	15	20	25	23	100

The percent homologies between all pairs of aligned chloroplast-type ferredoxins are shown. The comparison was made between the residues 27 and 110 shown in Fig. 4. Gaps are considered to be mismatched.

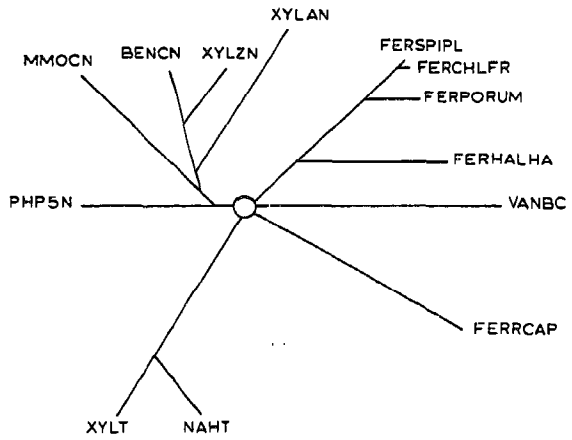

Fig. 6. Proposed phylogenetic relationships derived from the comparison of partial ferredoxin sequences. See Fig. 5 for the sequence names.

Possible evolutionary relationships between these ferredoxins were examined based on the results in Table I. As shown in Fig. 6, the chloroplast-type ferredoxins could be classified into five distinct subfamilies. The first subgroup comprises ferredoxins from chloroplasts, cyanobacteria and the *Halobacterium* species. These ferredoxins are characterized by their low mid-point redox potential (−300 mV or less). The second subgroup consists of the ferredoxin sequences found in BenC, MmoC, Polypeptide 5, XylZ and XylA. All the proteins in this subfamily are composite polypeptides which consist of ferredoxin-like structure in the amino-terminal region and NADH ferredoxin oxidoreductase-like structure in the carboxyl-terminal region. Since these ferredoxins are involved in the electron transfer from NADH to oxygen, their mid-point redox potential may be significantly higher than those of the first subfamily. The ferredoxins from *xylT* and *nahT* are similar to each other, however these proteins are distantly related to any of the other ferredoxins. Chloroplast-type ferredoxin from *R. capsulata* and carboxyl-terminal sequence of VanB are also strongly diverged from other ferredoxin sequences. XylA, XylT and XylZ are encoded on TOL plasmid pWW0. However, the sequence similarities between these proteins are not stronger than those between other ferredoxin pairs. This observation suggests that the direct progenitors of the ferredoxin genes on TOL plasmid pWW0 are not common, but these genes were derived from different ancestral ferredoxin genes.

Since the *xylT* and *NahT* genes were conserved on two different catabolic plasmids, their products may be involved in a metabolic step common to toluene- and naphthalene-degradative pathways, namely one of the steps for the transformation of catechol to Krebs cycle intermediates. We are in the process of isolating mutants of pWW0 defective in *xylT* in order to assess the role of XylT in the metabolism of catechol.

Acknowledgements: We thank Drs I.-S. You, D. Ghosal and I.C. Gunsalus for communicating us the nucleotide sequence of the *nahT* region prior to the publication, J.P. Shaw for careful reading of the manuscript, and F. Rey for typing the manuscript. This work was supported by the Swiss National Science Foundation.

REFERENCES

[1] Worsey, M.J. and Williams, P.A. (1975) J. Bacteriol. 124, 7–13.
[2] Yen, K.-M. and Gunsalus, I.C. (1982) Proc. Natl. Acad. Sci. USA 79, 874–878.
[3] Dunn, N.W. and Gunsalus, I.C. (1973) J. Bacteriol. 114, 974–979.
[4] Harayama, S. and Rekik, M. (1990) Mol. Gen. Genet. 221, 113–120.
[5] Harayama, S., Rekik, M., Wasserfallen, A. and Bairoch, A. (1987) Mol. Gen. Genet. 210, 241–247.
[6] Ghosal, D., You, I.-S. and Gunsalus, I.C. (1987) Gene 55, 19–28.
[6a] You, I.S., Ghosal, D. and Gunsalus, I.C. (1991) Biochemistry 30, 1635–1641.
[7] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[8] Promega protocols and applications guide, Promega Co., Madison, WI, USA.
[9] Henikoff, S. (1984) Gene 28, 351–359.
[10] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.
[11] Bruschi, M. and Guerlesquin, F. (1988) FEMS Microbiol. Rev. 54, 155–176.
[12] Saeki, K., Miyatake, Y., Young, D.A., Marrs, B.L. and Matsubara, H. (1990) Nucleic Acids Res. 18, 1060.
[13] Brunel, F. and Davison, J. (1988) J. Bacteriol. 170, 4924–4930.
[14] Stainthorpe, A.C., Lees, V., Salmond, G.P.C., Dalton, H. and Murell, J.C. (1990) Gene 9, 27–34.
[15] Nordlund, I., Powlowski, J. and Shingler, V. (1990) J. Bacteriol. 172, 6826–6833.
[16] Suzuki, M., Hayakawa, T., Shaw, J.P., Rekik, M. and Harayama, S. (1991) J. Bacteriol. 173, 1690–1695.
[17] Neidle, E.L., Shapiro, M. and Ornston, L.N. (1987) J. Bacteriol. 169, 5496–5503.
[18] Harayama, S., Rekik, M. and Timmis, K.N. (1986) Mol. Gen. Genet. 202, 226–234.
[19] Tanaka, M., Haniu, M., Yasunobu, K.T., Rao, K.K. and Hall, D.O. (1976) Biochem. Biophys. Res. Commun. 69, 759–765.
[20] Takahashi, Y., Hase, T., Matsubara, H., Hutber, G.N. and Rogers, L.J. (1982) J. Biochem. 92, 1363–1368.
[21] Takruri, I., Haslett, B.G., Boulter, D., Andrew, P.W. and Rogers, L.J. (1978) Biochem. J. 173, 459–466.
[22] Hase, T., Wakabayashi, S., Matsubara, H., Mebarech, M. and Werber, M.M. (1980) Biochim. Biophys. Acta 623, 139–145.