DNA Sequence of the Yeast Transketolase Gene[†,‡]

Thomas S. Fletcher,[§,∥] Ingrid L. Kwee,[*,⊥,#] Tsutomu Nakada,[⊥,#] Corey Largman,[§,∥] and Brian M. Martin[○]

Biochemistry and Neurochemistry Research Laboratories, Veterans Affairs Medical Center, Martinez, California 94553, Departments of Internal Medicine and Neurology, University of California, Davis, California 95616, and Clinical Neuroscience Branch, National Institute of Mental Health, Bethesda, Maryland 20892

Received June 27, 1991; Revised Manuscript Received November 22, 1991

ABSTRACT: Transketolase (EC 2.2.1.1) is the enzyme that, together with aldolase, forms a reversible link between the glycolytic and pentose phosphate pathways. We have cloned and sequenced the transketolase gene from yeast (*Saccharomyces cerevisiae*). This is the first transketolase gene of the pentose phosphate shunt to be sequenced from any source. The molecular mass of the proposed translated protein is 73 976 daltons, in good agreement with the observed molecular mass of about 75 000 daltons. The 5′-nontranslated region of the gene is similar to other yeast genes. There is no evidence of 5′-splice junctions or branch points in the sequence. The 3′-nontranslated region contains the polyadenylation signal (AATAAA), 80 base pairs downstream from the termination codon. A high degree of homology is found between yeast transketolase and dihydroxyacetone synthase (formaldehyde transketolase) from the yeast *Hansenula polymorpha*. The overall sequence identity between these two proteins is 37%, with four regions of much greater similarity. The regions from amino acid residues 98–131, 157–182, 410–433, and 474–489 have sequence identities of 74%, 66%, 83%, and 82%, respectively. One of these regions (157–182) includes a possible thiamin pyrophosphate (TPP) binding domain, and another (410–433) may contain the catalytic domain.

Transketolase (EC 2.2.1.1) is the enzyme of the pentose phosphate pathway which transfers a ketol group to an aldehyde acceptor molecule. Its substrate specificity is very broad, being able to utilize three-carbon to seven-carbon sugars. Together with aldolase, transketolase forms a reversible link between the glycolytic and pentose phosphate pathways, thereby enabling the cell to shuttle ribose 5-phosphate and glycolytic intermediates between the two pathways. In plants, transketolase also plays a vital role in the Calvin cycle, catalyzing the reaction of fructose 6-phosphate with glyceraldehyde 3-phosphate to yield xylulose 5-phosphate and erythrose 4-phosphate, and the reaction of sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate to yield ribose 5-phosphate and xylulose 5-phosphate. Purified transketolases from a variety of sources have similar properties; bakers's yeast transketolase (Kochetov, 1982) has a molecular mass of 159 kDa and consists of two identical subunits (Cavalieri et al., 1975); human transketolase from red blood cells (Himmo et al., 1988; Takeuchi et al., 1986) and leukocytes (Mocali & Paoletti, 1989) also have a similar subunit molecular mass of about 70 kDa (Himmo et al., 1988; Mocali & Paoletti, 1989; Takeuchi et al., 1986), as does transketolase prepared from rabbit liver (Masri et al., 1988). The enzyme requires thiamin pyrophosphate (TPP) as a cofactor, as well as Mg(II) ions for optimal activity. Transketolase is active as a homodimer, and it is known that the two TPP-binding sites exhibit negative cooperativity (Egan & Sable, 1981).

This enzyme, which has a broad substrate range, TPP as a cofactor, and negative cooperativity with regard to cofactor binding, should be an excellent model system to study cofactor and substrate interactions with an enzyme as well as the nature of the interaction between subunits. The application of site-directed mutagenesis to the study of transketolase should allow us to resolve many questions regarding its biochemical properties. As a first step, we have cloned and sequenced the transketolase gene from yeast (*Saccharomyces cerevisiae*). This is the first transketolase gene of the pentose phosphate shunt to be sequenced from any source.

Experimental Procedures

Yeast transketolase was purchased from Sigma Chemical (Sigma T-6133, St. Louis, MO). Transketolase was subjected to automated Edman degradation to determine the N-terminal sequence of the protein. To determine the sequence of other regions of the protein, transketolase was partially digested with TPCK-treated trypsin. The digest was separated by reversed-phase (C-18) HPLC, and well-resolved peaks were sequenced by automated Edman degradation using a Beckman 890M amino acid sequencer. PTH amino acids were identified by HPLC.

The oligonucleotide probes synthesized were partially degenerate (Suggs et al., 1981), and deoxyinosine was used at most ambiguous codon positions to maximize the chances of the oligonucleotide hybridizing to the yeast transketolase gene (Ohtsuka et al., 1985). A yeast (*Saccharomyces cerevisiae* X2180, ATCC 26109) genomic library in λgt11 (Clontech YL1001) was plated onto NZCYM plates and screened with two probes end labeled with [γ-^{32}P]ATP. Positive plaques were purified by repeated screening. Phage DNA was prepared, and the insert was subcloned into M13mp18 and M13mp19. Recombinant M13 phage were isolated and sequenced using chain-terminating inhibitors (Sanger et al., 1977). Both an M13 universal primer and specific primers

[†] This work was supported by the Department of Veterans Affairs.
[‡] The nucleotide sequence in this paper has been submitted to GenBank/EMBL Data Bank under Accession Number M63302.
[*] To whom correspondence should be addressed at the Neurochemistry Research Laboratory, Veterans Affairs Medical Center.
[§] Biochemistry Research Laboratory, Veterans Affairs Medical Center.
[∥] Department of Internal Medicine, University of California.
[⊥] Neurochemistry Research Laboratory, Veterans Affairs Medical Center.
[#] Department of Neurology, University of California.
[○] Clinical Neuroscience Branch, National Institute of Mental Health.

0006-2960/92/0431-1892$03.00/0

```
Peptide AA sequence (residues 2-9)      GlnPheThrAspIleAspLys

Actual gene sequence                5' CAATTCACTGACATTGATAA 3'
                                    3' GTTAAGTGACTGTAACTATT 5'

Synthesized probe (TKN-1)           3' GTTAAITGACTITAICTITT 5'
                                       C     C
                                             G
                                             T

Peptide AA sequence (residues 597-605)  AspPheThrPheAspLysGlnPro

Actual gene sequence                5' GATTTCTTCACTTTTGACAAACAACC 3'
                                    3' CTAAAGAAGTGAAAACTGTTTGTTGG 5'

Synthesized probe (TK24-1)          3' CTIAAIAAITGAAAICTITTTGTTGG 5'
                                             C        C  C
                                             G
                                             T
```

FIGURE 1: Design of oligonucleotide probes for library screening. Oligonucleotide probes were synthesized on the basis of amino acid sequences derived from yeast transketolase. The complementary sequence of the cloned transketolase gene is also shown for comparison.

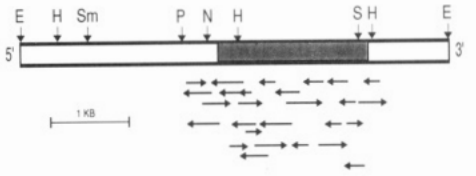

FIGURE 2: Sequencing strategy and partial restriction map of the yeast transketolase gene. The original 5-kb clone isolated by oligonucleotide screening was subcloned into M13mp18 and M13mp19. It was sequenced using both universal and specific sequencing primers. The portion of the DNA sequence that codes for the transketolase protein is shaded. The horizontal arrows indicate the direction and position of the sequence determinations. The vertical arrows indicate some restriction sites (E = *Eco*RI, H = *Hin*dIII, N = *Nhe*I, P = *Pvu*I, S = *Sac*I, and Sm = *Sma*I).

were used for sequencing. Restriction enzymes were purchased from New England Biolabs (Beverly, MA) and BRL (Gaithersburg, MD). DNA sequences were analyzed with computer programs supplied by Dr. H. Martinez (Biomathematics Computation Laboratory, Department of Biochemistry, UCSF, San Francisco, CA).

Results and Discussion

Amino Acid Sequence. An automated Edman degradation performed on an intact transketolase molecule yielded a sequence (TK-NT) which was used to establish the amino terminus of the active native transketolase protein. Yeast transketolase was also partially digested with trypsin. Ten peptides were purified to homogeneity and sequenced. Antisense oligonucleotide probes designed from the N-terminal sequence and from one tryptic peptide (TK-24) were synthesized. The amino acid sequence of the peptides and the derived oligonucleotide probes are presented in Figure 1. The corresponding nucleotide sequence of the yeast transketolase gene is also presented to illustrate the similarity between the probes and the actual gene.

DNA Sequence. When a yeast (*Saccharomyces cerevisiae*) genomic library in λgt11 was screened with the two end-labeled probes, the TK24-1 probe yielded two positive plaques. One of these contained a 5-kb insert when digested with *Eco*RI. Restriction fragments of the 5-kb insert were subcloned into M13 and sequenced. A 1500 bp *Hin*dIII fragment located in the middle of the 5-kb insert was found to code for 9 of the 10 tryptic peptides. We subcloned additional fragments which overlapped the ends of the 1500 bp sequence and found that these contained the 3′-untranslated region and the 5′ portion of the gene, including the N-terminal peptide and a large portion of the transketolase promoter. The sequencing strategy is shown in Figure 2.

Figure 3 shows the entire sequence of the yeast transketolase gene. All 10 of the transketolase peptide sequences were found in a single open reading frame of amino acid sequence deduced from the DNA sequence. Three serine residues encoded by the DNA in the region of peptide TK-12 were not detected in the peptide sequence due to the inability of the amino acid sequencer to detect low levels of this amino acid.

Peptide "TK-NT", which is the N-terminal peptide of transketolase, is, as expected, at the most N-terminal portion of the translated DNA sequence. It is preceded by a methionine which we have assigned as amino acid 1 and assumed to be the translation start site. There are no other methionine codons between this methionine codon and an upstream, in-frame stop codon at 175 nt. The molecular mass of the translated protein from the methionine translation start codon (271 nt) to the stop codon at 2305 nt is estimated to be 73 976 daltons. This is in good agreement with the observed protein molecular mass of about 75 000 daltons (Kochetov, 1982). The codon usage of this gene is characteristic of highly expressed yeast genes which exhibit a codon bias which is different from genes expressed at low levels (Sharp et al., 1988). Out of 20 amino acids, 18 of them exhibit the same codon bias in transketolase as is seen in highly expressed yeast genes.

The 5′-nontranslated region of the gene is similar to other yeast genes. It contains two possible TATA boxes at −220 and −161 nt, a possible CAAT box at −209 nt, a CT-rich region at −108 to −82 nt (Dobson et al., 1982), and a region adjacent to a potential initiation codon (−8 to −1 nt) that is highly homologous to the sequence of the yeast glyceraldehyde-3-phosphate dehydrogenase genes (Holland & Holland, 1980). There is no evidence of 5′-splice junctions or branch points in the sequence. The 3′-nontranslated region contains the polyadenylation signal (AATAAA) at 2386 nt (Proudfoot & Brownlee, 1976), 80 base pairs downstream from the termination codon (Figure 3).

A search of the GenBank DNA sequence database found no sequences resembling yeast transketolase, but a comparison of the amino acid sequence with that of a related enzyme, dihydroxyacetone synthase (formaldehyde transketolase), from the yeast *Hansenula polymorpha* (Janowicz et al., 1985) revealed an overall amino acid sequence identity of 37%. The sequence similarities are concentrated in four regions of yeast transketolase: amino acid residues 98–131, 157–182, 410–433, and 474–489, which have sequence homologies of 74%, 66%, 83%, and 82%, respectively (Figure 4). One of these regions (157–182) is homologous to a consensus sequence [GDG-(X_8)E(X_4)A(X_{13})N] previously identified by Hawkins (Hawkins et al., 1989) in a survey 16 TPP-binding enzymes to contain the thiamin pyrophosphate (TPP) cofactor-binding domain.

While this is the only domain of these two transketolases that can be identified by amino acid sequence homology, if the regions of homology represent regions of conserved function, it should be possible to identify the functional domains by correlating the amino acid composition of these domains with studies in which chemical modifications of specific amino acids have changed the catalytic properties of yeast transketolase. There are reports that the modification of arginine residues (Kremer et al., 1980; Usmanov & Kochetov, 1983), tyrosine residues (Kuimov et al., 1988), and aspartic/glutamic acid residues (Kuimov et al., 1985; Meshalkina et al., 1984) can inactivate yeast transketolase.

Modification of arginine residues with either 2,3-butanedione or phenylglyoxal results in a loss of catalytic activity (Kremer et al., 1980; Usmanov & Kochetov, 1983). The

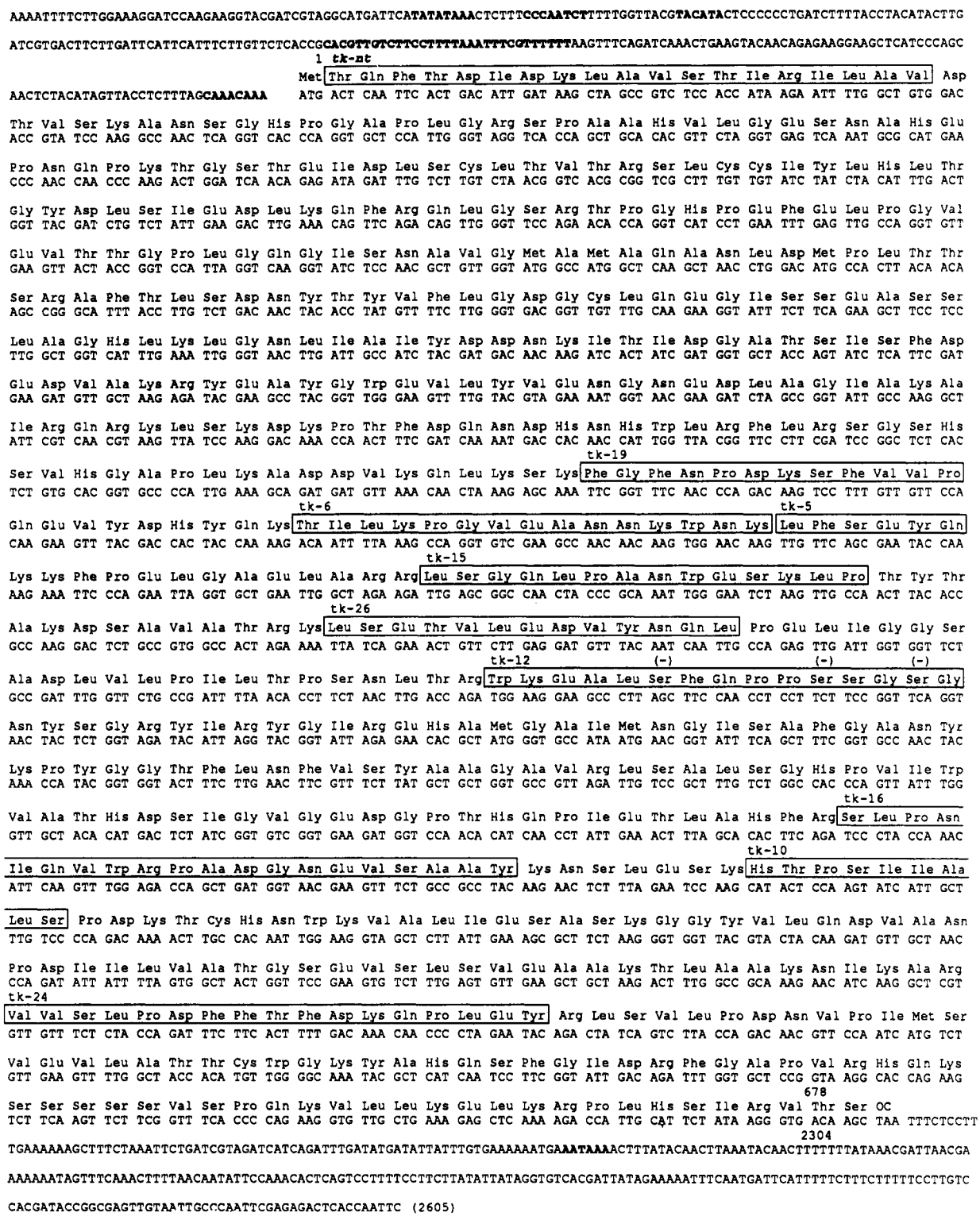

FIGURE 3: Nucleotide sequence of the yeast transketolase (EC 2.2.1.1) gene and the deduced amino acid sequence. The amino acid sequence is numbered sequentially from a proposed methionine start site located immediatedly before the amino acid determined by amino acid sequencing to be the N-terminal amino acid of the purified transketolase protein. The amino acid sequences of the HPLC-purified peptides are boxed, and differences between the deduced and actual amino acid sequence are indicated by showing the actual amino acid sequence of the peptide above the deduced amino acid sequence. A "(-)" indicates that no amino acid was detected by the amino acid sequencer during the cycle corresponding to that residue. The numbering of the DNA sequence and the deduced amino acid sequence is indicated to the left of the sequences. The proposed polyadenylation site is indicated by boldface at 2386 nt. The portions of the 5′-nontranslated region of interest are boldfaced. They include two possible TATA boxes at −220 and −161 nt, a possible CAAT box at −209 nt, a CT-rich region at −108 to −82 nt, and a region adjacent to a potential initiation codon (−8 to −1) that is homologous to the sequence of the yeast glyceraldehyde-3-phosphate dehydrogenase genes.

addition of TPP (a cofactor) or the substrate-binding inhibitors phenyl phosphate or phenyl sulfate does not affect the kinetics of inactivation (Kremer et al., 1980), implying that these arginine residues are located in the catalytic site of transketolase and not in the TPP-binding or substrate-binding domains. Examination of the conserved regions of the two

FIGURE 4: Amino acid sequence comparison of two transketolases. The translated amino acid sequences of yeast transketolase (top) and formaldehyde transketolase from *Hansenula polymorpha* (bottom) are compared using the Needleman–Wunsch method of alignment (Martinez, 1988). The regions of amino acid sequences that we have identified as highly conserved are boxed.

transketolases reveals that only the region from 410 to 433 contains conserved arginine residues (Arg-423 and Arg-430). Additional support for the assignment of this region as part of the catalytic domain is that the tyrosine-specific reagent *N*-acetylimidazole destroys the catalytic activity of transketolase (Kuimov et al., 1988) and of the four conserved regions only residues 410–433 contain conserved tyrosines (Tyr-413, Tyr-417, and Tyr-420) in these two transketolases.

The reaction of transketolase with either CDI [1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide] or Woodward's reagent K results in complete inactivation (Meshalkina et al., 1984). The kinetics of inactivation suggest that modification of a single carboxyl group per active site is sufficient to inactivate the protein. There are a total of nine aspartic and glutamic residues present in the four conserved regions of transketolase, making it impossible to identify any one of them as containing the critical carboxyl group. However, the region from 410 to 433 does contain a single glutamic acid (Glu-424) which is consistent with this region being the active site of the enzyme.

In addition to the proposed TPP-binding domain (157–181) and the proposed catalytic site (410–433), the two transketolases contain two other conserved regions (98–131 and 474–489). We propose that these sequences are possibly involved in either substrate binding or dimerization. While the chemical modification of transketolase has yielded information to make tentative assignments of function, by combining an understanding of the crystal structure of this enzyme (Schneider et al., 1989) with site-directed mutagenesis and the introduction of modified transketolase sequences into yeast cells, the enzymatic properties of transketolase can be dissected and it can be determined if the conserved regions of the protein are critical for its biochemical properties.

ACKNOWLEDGMENTS

We thank Teresa Redilla-Flores for her excellent technical assistance and all the members of the Core Biochemistry Laboratory at the Veterans Affairs Medical Center (Martinez, CA) for their helpful discussions.

REFERENCES

Cavalieri, S. W., Neet, K. E., & Sable, H. Z. (1975) *Arch. Biochem. Biophys. 171*, 527–532.

Dobson, M. J., Tuite, M. F., Roberts, N. A., Kingsman, A. J., & Kingsman, S. M. (1982) *Nucleic Acids Res. 10*, 2625–2637.

Egan, R. M., & Sable, H. Z. (1981) *J. Biol. Chem. 256*, 4877–4883.

Hawkins, C. F., Borges, A., & Perham, R. N. (1989) *FEBS Lett. 255*, 77–82.

Himmo, S. D., Thomson, M., & Gubler, C. J. (1988) *Prep. Biochem. 18*, 261–276.

Holland, J. P., & Holland, M. J. (1980) *J. Biol. Chem. 255*, 2596–2605.

Janowicz, Z. A., Eckart, M. R., Drewke, C., Roggenkamp, R. O., Hollenberg, C. P., Maat, J., Ledeboer, A., Visser, C., & Verrips, C. T. (1985) *Nucleic Acids Res. 13*, 3043–3062.
Kochetov, G. A. (1982) *Methods Enzymol. 90*, 209–223.
Kremer, A. B., Egan, R. M., & Sable, H. Z. (1980) *J. Biol. Chem. 255*, 2405–2510.
Kuimov, A. N., Meshalkina, L. E., & Kochetov, G. A. (1985) *Biochem. Int. 11*, 913–920.
Kuimov, A. N., Kovina, M. V., & Kochetov, G. A. (1988) *Biochem. Int. 17*, 517–521.
Martinez, H. M. (1988) *Nucleic Acids Res. 16*, 1683–1691.
Masri, S. W., Ali, M., & Gubler, C. J. (1988) *Comp. Biochem. Physiol. B: Comp. Biochem. 90B*, 167–172.
Meshalkina, L. E., Kuimov, A. N., Kabakov, A. N., Tsorina, O. N., & Kochetov, G. A. (1984) *Biochem. Int. 9*, 9–16.
Mocali, A., & Paoletti, F. (1989) *Eur. J. Biochem. 180*, 213–219.
Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., & Matsubara, K. (1985) *J. Biol. Chem. 260*, 2605–2608.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature 263*, 211–214.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5476.
Schneider, G., Sundstrom, M., & Lindqvist, Y. (1989) *J. Biol. Chem. 264*, 21619–21620.
Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H., & Wright, F. (1988) *Nucleic Acids Res. 16*, 8207–8211.
Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., & Itakura, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6613–6617.
Takeuchi, T., Nishino, K., & Itokawa, Y. (1986) *Biochim. Biophys. Acta 872*, 24–32.
Usmanov, R. A., & Kochetov, G. A. (1983) *Biochem. Int. 6*, 673–683.

CORRECTION

Primary Structure of the Thermostable Formyltetrahydrofolate Synthetase from *Clostridium thermoaceticum*, by Charles R. Lovell,* Alan Przybyla, and Lars G. Ljungdahl, Volume 29, Number 24, June 19, 1990, pages 5687–5694.

Page 5689. In Table II, the numbers of Lys residues predicted from the nucleotide sequences of the *C. thermoaceticum* and *C. acidiurici* formyltetrahydrofolate synthetase genes were inadvertently reversed. They should read 36 for *C. thermoaceticum* and 47 for *C. acidiurici*. The sums of the basic residues (Arg + Lys) should be 62 and 64, respectively.

Page 5692. In column 2, line 15 should read 6 more basic residues (including Tyr) overall were found for the *C. thermoaceticum* enzyme, not 20.