# Complementation of Fragments of Triosephosphate Isomerase Defined by Exon Boundaries†

Bonnie L. Bertolaet*,‡ and Jeremy R. Knowles

*Departments of Chemistry and Biochemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138*

*Received August 4, 1994; Revised Manuscript Received February 9, 1995*®

ABSTRACT: Chicken triosephosphate isomerase (TIM) has been fragmented by inserting single "splits" at three separate exon/exon boundaries, and the complexes have been assayed for catalytic activity. *In vivo* studies show that the expression of *both* portions of each of the three different split genes complements the TIM deficiency of *Escherichia coli* strain DF502 when grown on selective media. The expression of only one fragment of each split gene does not complement the TIM-minus genotype. To assess the catalytic activity that derives from the fragmented protein complex, the individual peptide products of one of the three split genes were expressed and purified. The purified complex showed isomerase activity, albeit of low specific catalytic activity. A catalytically active multichain complex composed of separate peptide products of a gene singly split at exon/exon junctions has thus been created.

Triosephosphate isomerase (TIM) is an ancient, ubiquitous, and relatively highly conserved protein that is believed to have evolved before the archaebacteria−prokaryotic−eukaryotic division (Marchionni & Gilbert, 1986). Much is known about its structure and catalytic mechanism (Knowles, 1991a,b), and it is well suited for the study of protein evolution, stability, and structure. The work described here seeks to examine the plasticity of an enzyme's structure in the context of the evolution of functional proteins. To this end, the gene that encodes chicken TIM has been "split" at several of the known exon/exon junctions, and the combination of the resulting peptide fragments has been assessed for isomerase catalytic activity.

Ever since the discovery in 1977 that eukaryotic genes are not continuous but are interrupted by noncoding sequences (Berget et al., 1977; Chow et al., 1977), the origin and function of these sequences, known as introns, have been subjects of much speculation. It is not yet understood why eukaryotic genes are interrupted by introns, nor is it known what determines the position of these intervening sequences. Although there are many theories, each with its own variations and subtleties, two views dominate current thinking. One view holds that introns are ancient, that they predate the divergence of eukaryotes and prokaryotes, and that ancestral genes contained these interrupting sequences. The contrary view attests that introns have "invaded" eukaryotic genes that were once continuous. Each of these positions begs a range of questions relating to the nature and purpose of introns (as distinct from the timing of their appearance) as features of eukaryotic genes. Recently, in an effort to account for the origin of introns, we put forward the "exon microgene" theory (Seidel et al., 1992), in which we sought to interpret "the faint outline of ancient patterns" that could be perceived in modern genes. We suggested that exon/intron and intron/exon boundaries were originally determined early in evolution by terminating amber codons (TAG) of exons that each encoded independently translated peptides which assembled to form catalytically active complexes.

Our hypothesis aimed to explain the origin of the consensus RNA splice site sequences in protein-encoding genes at exon/intron (..AG/..) and intron/exon (..ag/..) boundaries, which are quite distinct from those found in intron-containing rRNA and tRNA genes (Shapiro & Senapathy, 1987). We proposed that the AG of the consensus sequence (for protein-encoding genes) could derive from TAG of original amber termination codons. We also suggested that the independently translated peptides need not have any intrinsic structural stability but that they could combine together to form catalytic multichain assemblies (Figure 1). That such complementation is possible is well documented, and there are many examples where protein fragments can combine to generate catalytic activity (Anfinsen et al., 1971; Richards & Wyckoff, 1971; Li & Bewley, 1976; Holmgren & Slaby, 1979; Galakatos & Walsh, 1987; Saint Girons et al., 1987; Burbaum & Schimmel, 1991; Toyama et al., 1991; Eder & Kirschner, 1992). Intron-mediated exon shuffling and migration, by the early counterparts of molecular mechanisms that survive today, would have brought together those exons that contribute to a particular catalytic activity (Gilbert, 1978, 1985), allowing the linked inheritance of the microgenes that encoded the components of a particular functional assembly. This step is consistent with the notion that genes were constructed by exon shuffling mechanisms facilitated by some recognition element (*i.e.*, their common amber 3′-termini). Excision of the introns by means presumably analogous to modern splicing mechanisms would lead to a spliced message that still encoded individually translated peptides. In some cases, the excised introns can be thought of as "failed" exons in the sense that they encoded peptides that did not add to the overall stability or function of any catalytic complex. Finally, read-through of the amber stop codons of those exons that were in frame would yield crude but continuous single-chain protomers (Figure 1), which were the substrates for catalytic refinement and improvement over evolutionary time.

† Supported by National Institutes of Health Grant GM37007.

* To whom correspondence should be addressed.

‡ Present address: Cancer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0684.

® Abstract published in *Advance ACS Abstracts,* April 1, 1995.

0006-2960/95/0434-5736$09.00/0 

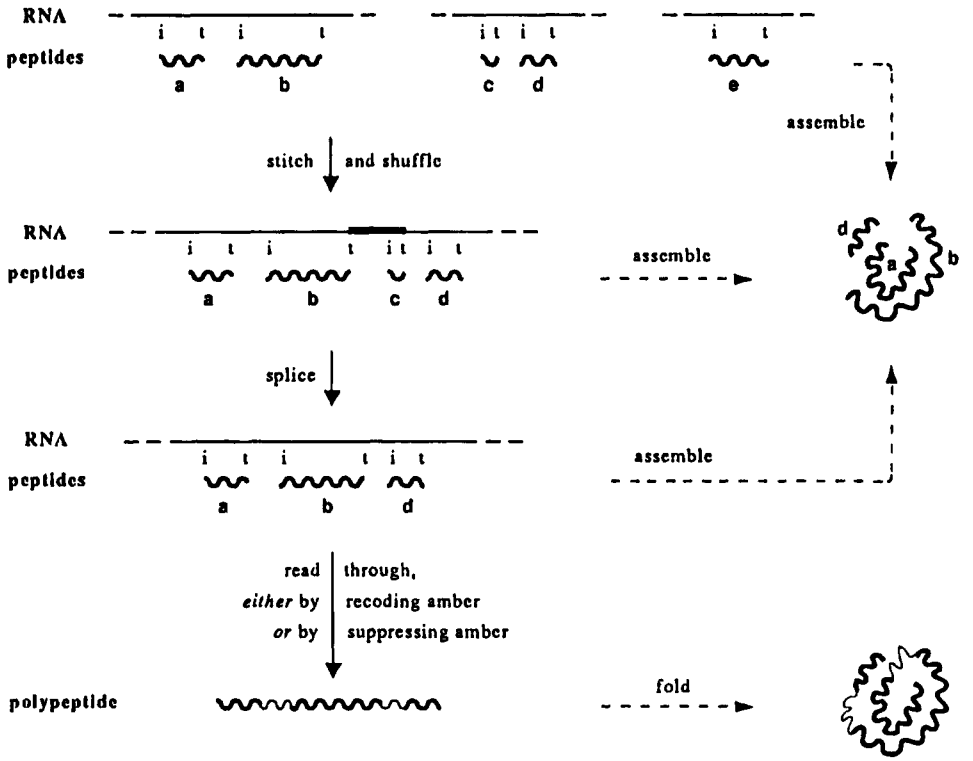


FIGURE 1: Putative development of functional proteins. From a variety of segments of RNA, initiation of protein synthesis (at i) and termination of protein synthesis (at t) produce a library of peptides (**a**, **b**, **c**, etc.) from which catalytic activity is generated by spontaneous peptide assembly. Use of the termination amber codons (t) as a recognition element for stitching useful RNA segments together and for shuffling these microgenes [by mechanisms analogous to those of present-day RNA splicing, which is illustrated by the excision of the intron (heavy bar)] allows the rough juxtaposition and (therefore) the linked inheritance of the microgenes that encode the components of a particular functional assembly. Read-through of the terminating amber codons by one of the two mechanisms indicated (recoding amber or suppressing amber) then produces a continuous polypeptide that has greater thermal stability than the segmented assembly of smaller peptides. This figure was adapted from Seidel et al. (1992b).

If the juxtaposition of a pair of microgenes were to result in neighboring exons being out of frame, then read-through would be selected against, and the intervening sequence would have to be excised later to produce a functional, single-chain protein. The spliceosome would have evolved after the constraint to recognize AG had been established, originally with AG in frame, but later less constrained, due to junctional sliding, allowing for the distribution of intron phase classes observed today (Federov et al., 1992). The migration of the splice sites evidently did not result in the complete randomization of intron phases but retained the greatest frequency for phase 0 introns (*i.e.*, those falling between codons), although to varying extents depending on the species (Federov et al., 1992).

Since we do not present data on TIM genes that have been split at positions *other* than exon/exon junctions, this study does not strictly test the exon microgene theory. Sadly, there is no way to design a satisfying negative control for such a study, short of splitting the protein at every single position, which is a technically daunting problem. Yet even such a notional control is not perfect. An ancestral TIM gene could at some time have contained introns that have not survived to the present time in any organism. Positions of splits that yielded a catalytically active complex and that did not correspond to any known intron locations could conceivably correspond to ancient introns that have been lost. These considerations illustrate the difficulty of attempting to recreate primitive enzymes from modern ones. The work presented here simply examines the functional activity of complementary peptides defined by exon/exon boundaries. With the caveats presented above, the exon microgene theory provides an intriguing backdrop for our results.

Chicken muscle TIM is the paradigm for the 8-fold $\alpha\beta$-barrel structure and is the parent of a family of many other proteins that share this structural motif (Farber & Petsko, 1990). The enzyme is a homodimer with a subunit molecular weight of 26 500, and the chicken gene encoding this protomer has seven exons interrupted by six introns. It has been noted by Gilbert that if the location of introns in a gene were random (for example, as a consequence of the insertion of intron segments into an originally continuous gene), a statistical distribution would be expected in the size of exons (Straus & Gilbert, 1985). This is not the case, however, either generally or for the gene encoding chicken TIM. The exons for TIM encode peptides of roughly two sizes: exons 1, 2, 4, and 7 encode peptides of 38, 41, 45, and 47 amino acids, respectively, and exons 3, 5, and 6 encode peptides of 28, 29, and 29 amino acids, respectively. This and other observations led Gilbert to hypothesize that the TIM gene was originally constructed by the multiplication of "genetic units" encoding $\alpha\beta$ motifs (Straus & Gilbert, 1985).

All but two of the splice junctions in the genomic sequence for chicken TIM conform to the eukaryotic consensus sequence of ..AG/.. for exon/intron junctions and ..ag/.. for intron/exon junctions (Shapiro & Senapathy, 1987). The possibility that the consensus RNA splice site sequences at these junctions in the TIM gene could at one time have been part of a translational stop sign (UAG) is consistent with the structure of the chicken TIM monomer. Figure 2 illustrates the surface location of exon/exon junctions in the

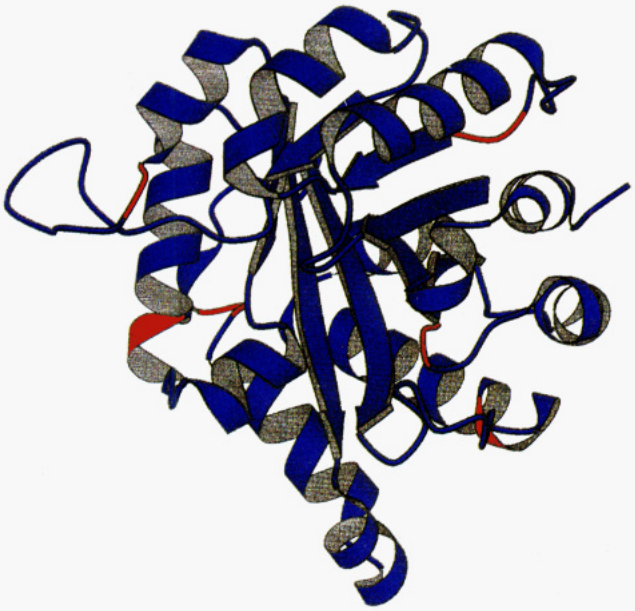

FIGURE 2: Side view of a monomer of chicken triosephosphate isomerase. Surface locations of exon/exon junctions are highlighted in red.

TIM monomer, which is a general phenomenon (Craik et al., 1982). The exon microgene theory is consistent with this fact in suggesting that surface locations could have been driven by the need to solvate and stabilize the charged N- and C-termini of the originally independently translated peptides. The surface location of these termini could also have been advantageous when read-through occurred (creating the extra peptide sequences between t and i, illustrated in Figure 1), allowing the peptide fragments to be joined by a peptide loop that would be most readily accommodated at the surface of the folded polypeptide.

TIM is highly conserved in terms of its sequence, structure, and kinetic characteristics. Most of the conserved residues lie in or near the active site of the enzyme, which lies toward the carboxyl ends of the $\beta$ strands, most of which are encoded by separate exons. The codons for active site residues are therefore distributed among the exons of the TIM gene (Straus & Gilbert, 1985; Knowles, 1991a). Exon 1 encodes lysine-12, a conserved residue essential for catalytic activity, as well as residues that contact the adjacent subunit (Phillips et al., 1977; Alber et al., 1981; Lolis & Petsko, 1990; Davenport et al., 1991). Exon 2 encodes the loop that extends into the active site of the adjacent subunit. Exon 3 encodes histidine-95, which acts as a general acid in the enolization step (Belasco & Knowles, 1980; Komives et al., 1991; Nickbarg et al., 1988; Lodi & Knowles, 1991). Histidine-95 is located at the positive end of a conserved $\alpha$ helix that is also encoded by exon 3 (Lodi & Knowles, 1993). Exon 5 encodes glutamate-165, the catalytically critical enzymic base (Straus et al., 1985; Raines et al., 1986; Lolis & Petsko, 1990; Davenport et al., 1991), as well as the active-site "lid" (168−177) that envelops the substrate once it is bound (Banner et al., 1975; Alber et al., 1981; Lolis & Petsko, 1990). Finally, exon 7 encodes the $\alpha$ helix containing glycine-232 and -233 that provides backbone hydrogen bonds to the substrate's phospho group (Alber et al., 1987). The distribution of essential residues along the length of the protein and the consequential spread of their codons among many exons make TIM an attractive candidate for this study. Because the nature and location of the exon/intron junctions rather than the domains of the protein are the subject of this work, the precision of the cleavage sites is of great importance. Proteolysis could not reliably produce the appropriate peptides, so the TIM gene was modified so as to encode polypeptide fragments corresponding precisely to those defined by the exon/exon junctions.

## EXPERIMENTAL PROCEDURES

*Materials.* *Escherichia coli* strain DF502, a streptomycin-resistant strain of *E. coli* that lacks the chromosomal triosephosphate isomerase (constructed in a manner analogous to DF500 by Dr. D. Fraenkel: Babul, 1978; Fraenkel, 1986), was a generous gift from Dr. D. Fraenkel. Plasmid pBSTIM has been described previously (Hermes et al., 1990). All chemicals were from Sigma Chemical Co. (St. Louis, MO), restriction enzymes were from New England Biolabs (Beverly, MA), and other enzymes and protease inhibitors were from Boehringer Mannheim Biochemicals (Indianapolis, IN). Bromohydroxyacetone phosphate (BHAP) was the kind gift of Dr. E. Komives (de la Mare et al., 1972). Oligonucleotides were synthesized on a MilliGen/Biosearch 7500 DNA synthesizer (Millipore Corp., Bedford, MA).

*Mutagenesis.* Site-directed mutagenesis was performed by the method of Eckstein (Taylor et al., 1985a,b; Nakamaye & Eckstein, 1986; Sayers et al., 1988) using materials from Amersham Corp. (Arlington Heights, IL). Insertional Eckstein mutagenesis was performed using the method of Burbaum and Schimmel (1991). DNA sequencing was performed using the dideoxy chain termination method (Sanger et al., 1977) using radiolabeled nucleotides from Amersham Corp.

*Construction of Vectors pBSBLB-1, -2, and -3.* Vectors pBSBLB-1, -2, and -3 encode split genes 1, 2, and 3, respectively, and were constructed by the method of Burbaum and Schimmel (1991). Insertion of an 18-nucleotide sequence in the case of split genes 1 and 2 and a 17-nucleotide sequence in the case of split gene 3 (see Figure 3) consisting of a stop codon (TAA), a ribosomal binding site (AGGA), and an ATG start codon was accomplished by use of primers 70 oligonucleotides in length on the single-stranded form of phagemid pBSTIM as template.

*Construction of Vectors pBSBLB-1a, -2a, and -3a.* Vectors pBSBLB-1a, -2a, and -3a encode gene fragment *a* of split genes 1, 2, and 3, respectively, and were constructed by site-directed mutagenesis. Elimination of the initiation codon of gene fragment *b* and the introduction of a second stop codon [...TAACTAGGACC(C)ATCATG... changed to ...TAACTAGGACC(C)ATCATA**TAA**... (extra C nucleotide in pBS-1 and -2)] were accomplished by use of primers 21 oligonucleotides in length on the single-stranded form of pBSBLB-1, -2, and -3 as templates.

*Construction of Vectors pBLB-1b, -2b, and -3b.* Vectors pBSBLB-1b, -2b, and -3b encode gene fragment *b* of split genes 1, 2, and 3, respectively, and were constructed by amplifying the appropriate gene fragment from a pBSTIM template by the polymerase chain reaction (PCR) with modified 5′ and 3′ primers to introduce restriction sites for subcloning into pBSTIM.

*Polymerase Chain Reaction.* The reactions were performed using *Taq* DNA polymerase (Promega, Madison, WI) as recommended by the manufacturer, using priming oligo-

nucleotides at 0.5 μM each and pBSTIM as template DNA. PCR reactions were carried out in a programmable thermal cycler (PTC-100, MJ Research, Cambridge, MA).

*Selection for Transformants with Isomerase Activity.* Each of the vectors pBSBLB-1, -2, -3, -1a, -2a, -3a, -1b, -2b, and -3b was used individually to transform *E. coli* strain DF502. Selection for fragmented proteins with isomerase activity was performed on M63 minimal plates (Miller, 1972) with lactate (0.2% w/v) as the sole carbon source, as described by Hermes et al. (1987, 1989). Plasmid DNA was isolated from colonies grown on selective plates and sequenced.

*Enzyme Assay.* Triosephosphate isomerase activity was determined by using glyceraldehyde 3-phosphate as substrate in a coupled enzyme assay, based on the method of Putman et al. (1972), at 30 °C in triethanolamine hydrochloride buffer (100 mM, pH 7.6) containing EDTA (10 mM).

*Construction of pGEX-1a and -1b Vectors.* Vectors pGEX-1a and -1b encode fusion proteins of gene fragments *a* and *b*, respectively, of split gene 1 with glutathione *S*-transferase (GST). These vectors were constructed by amplifying the appropriate gene fragment from a pBSTIM template by PCR with a modified 5′ primer to introduce a *Bam*HI site and factor Xa protease recognition site and a modified 3′ primer to introduce an *Eco*RI site for subcloning into pGEX1 (Pharmacia Biotech Inc., Piscataway, NJ).

*Purification of Fusion Proteins* **1a** *and* **1b** *from E. coli BL21(DE3)(pGEX1a) and BL21(DE3)(pGEX1b), Respectively. E. coli* strain BL21(DE3) (Novagen, Inc., Madison, WI) cells harboring pGEX1a or pGEX1b were grown on ZYG media [bacto tryptone (10 g), bacto yeast extract (5 g), NaCl (5 g), and glucose (4 g) in a total volume of 1 L] containing ampicillin (200 μg/mL). Starter cultures (100 mL) were used to inoculate culture medium, were grown with shaking for 1.5 h at 37 °C, and then induced with isopropyl β-D-thiogalactopyranoside (IPTG; final concentration 0.5 mM) for 4 h. The cells were harvested by centrifugation (5000*g*) at 4 °C. The cell paste was resuspended in Tris-buffered saline (TBS) buffer [10 mL; Tris-HCl (25 mM, pH 8.0) containing NaCl (140 mM) and KCl (2 mM)] containing EDTA (1 mM), phenylmethanesulfonyl fluoride (PMSF; 0.4 mg/mL), Pefabloc (0.5 mg/mL), *N*-tosyl-L-lysine chloromethyl ketone (TCLK; 0.4 mg/mL), aprotinin (2 μg/mL), and pepstatin (1.4 μg/mL), then treated with lysozyme (10 mg) for 1 h, and lysed by passing the cell suspension through a French pressure cell (Aminco, Urbana, IL) at 20 000 psi. Triton X-100 (final concentration 1% v/v) was added to the resulting lysate which was then cleared by centrifugation for 10 min at 10000*g*. Insoluble fusion protein present in the resulting cell pellet was solubilized in 8 mL of triethanolamine (25 mM, pH 8.0) containing *N*-lauroylsarcosine (1.5%) and EDTA (1 mM) with stirring at 4 °C. The resolubilized pellet was cleared by centrifugation for 10 min at 10000*g*. Triton X-100 [final concentration 2% (v/v)] and $CaCl_2$ (final concentration 1 mM) were added to the supernatant. The original cell-free extract and supernatant from the resolubilized pellet were mixed and then loaded onto a glutathione−agarose column (bed volume 2 mL; Pharmacia) preequilibrated with TBS buffer. The column was washed with TBS buffer and then eluted with Tris-HCl buffer (100 mM, pH 8.0) containing reduced glutathione (20 mM) and NaCl (150 mM). The eluant was equilibrated with Tris-HCl buffer (100 mM, pH 8.0) using a PD-10 gel filtration column (Pharmacia) and then concentrated to 2 mL using Amicon 3 ultrafiltration units (Amicon Inc., Beverly, MA). The concentrated sample was then loaded onto an FPLC Mono Q HR 5/5 column (Pharmacia) preequilibrated with buffer A (Tris-HCl, 100 mM, pH 8.0). The column was eluted isocratically with 9.5% buffer B (Tris-HCl, 100 mM, pH 8.0, 1 M KCl) and then with a linear gradient from 9.5% buffer B to 35% buffer B, followed by a second linear gradient from 35% buffer B to 43.7% buffer B. The column was then washed with buffer B followed by buffer A (10 min, 0.5 mL/min). The fractions containing the highest ratio of fusion protein to free GST, as determined by SDS−PAGE, were pooled and treated with BHAP as described by Plaut and Knowles (1972).

*Cleavage of Fusion Proteins with Factor Xa.* After BHAP treatment, the individual preparations of fusion proteins **1a** and **1b** were pooled and equilibrated with Tris-HCl buffer (50 mM, pH 7.5) containing NaCl (150 mM) and concentrated by ultrafiltration. For some experiments, fusion proteins **1a** and **1b** were combined and simultaneously cleaved with factor Xa in a ratio of 100:1 (w/w) in cleavage buffer [Tris-HCl (50 mM, pH 7.5) containing NaCl (150 mM) and $CaCl_2$ (2 mM)]. The reaction mixture was divided into separate portions and incubated at 25 °C. Portions were denatured by treatment with guanidinium chloride (6 M, pH 8.0) containing dithiothreitol (DTT, 5 mM) every 4 h. Aliquots were removed from each sample for SDS−PAGE analysis before addition of the denaturant. Control cleavage reactions with either fusion protein **1a** or fusion protein **1b** alone were performed in the same manner. The denatured products from the simultaneous cleavage of fusion proteins **1a** and **1b** were concentrated by ultrafiltration and dialyzed against TBS buffer (250 mL) using Spectra/Por cellulose ester membranes (molecular mass cutoff 0.5 kDa; Spectrum Medical Industries, Inc., Houston, TX) at 4 °C with buffer changes every 2 h for 8 h. The refolded samples were then concentrated by ultrafiltration and assayed for catalytic activity.

*Protein Determination.* The amount of peptides **1a** and **1b** released during the simultaneous cleavage of fusion proteins **1a** and **1b** by factor Xa was determined by densitometry scanning of a gradient polyacrylamide gel (4−20%) containing the products of the cleavage reaction using a Fuji PDI Bio-Imaging Analyzer, Model DNA 35 (Fuji Photo Film Co., Ltd., Japan).

## RESULTS AND DISCUSSION

*Splitting the TIM Gene.* The continuous coding sequence of the chicken TIM gene was split at three exon/exon boundaries so as to encode a set of TIM proteins each fragmented into two polypeptides. Single splits were introduced into the cDNA clone for TIM at the three exon junctions shown in Figure 3. These splits were made by inserting a sequence that includes a stop codon (TAA) at the end of one exon, a ribosomal binding site (AGGA), and a start codon (ATG) at the beginning of the next exon. The sequence contains 18 nucleotides in the case of split genes 1 and 2 and 17 nucleotides in the case of split gene 3. The intercistronic region of split gene 3 is not a multiple of three and thus eliminates concerns of read-through translation and production of a continuous polypeptide containing an extra peptide loop. [Although each of these constructs actually creates a two-cistron message, we shall refer to this poly-

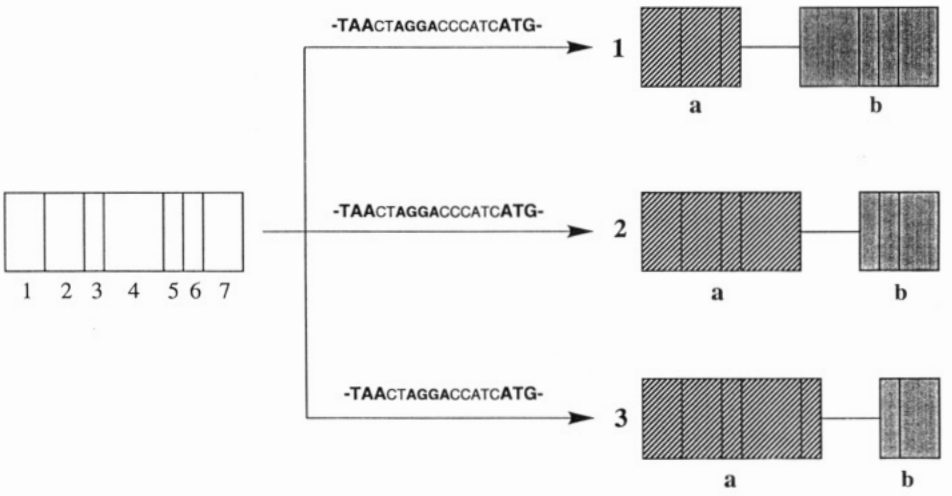


FIGURE 3: Single "splits" (**1**, **2**, **3**) introduced into the gene for chicken triosephosphate isomerase at three different exon/exon junctions by inserting an 18- or a 17-nucleotide sequence that contains a stop codon (TAA) at the end of the one exon, a ribosomal binding site (AGGA), and a start codon (ATG) at the beginning of the next exon.

cistronic message as a "split gene".] Such constructions produce pairs of separate but translationally coupled peptides and have been used in the expression of heteromultimeric proteins (Omer et al., 1993; Hoffman et al., 1990) as well as domains and fragments of single protomeric proteins (Burbaum & Schimmel, 1991; Toyama et al., 1991; Eder & Kirschner, 1992).

The sizes of the peptide products corresponding to the split TIM genes are 11.8 and 15.5 kDa for split 1, 16.6 and 10.7 kDa for split 2, and 19.8 and 7.5 kDa for split 3. The fragmentation of the protein encoded by splits 1, 2, and 3 is illustrated in Figure 4. While the introduction of single splits into the TIM gene may seem conservative, the fact that TIM is catalytically active only as a dimer requires that four peptide fragments assemble to form a catalytically active "dimeric" enzyme in each case. The splits can be seen relative to the dimerization interface of the intact protein, which is formed by a loop (shown in the upper-left corner of the structures in Figure 4) of one monomer that "interdigitates" with the residues along one wall of the active site cleft of the other monomer.

*The in Vivo Selection System for TIM.* To test the ability of the peptide products from the split genes to assemble and generate catalytic activity, *in vivo* complementation was employed. An *in vivo* selection system provides the best way to explore whether the translation products of exons (or sets of exons) can assemble into a catalytically active complex inside the cell. Unlike an *in vitro* assay of purified polypeptide fragments, which would necessarily be performed in an artificial environment devoid of competing packing interactions and degradative proteolytic processes, the *in vivo* complementation assay provides a more rigorous test for catalytic activity.

The *in vivo* selection system for TIM is based on *E. coli* strain DF502, from which the endogenous TIM gene has been excised (Babul, 1978; Fraenkel, 1986). Although strain DF502 lacks TIM, this strain can grow on glucose as the sole carbon source [from which both glyceraldehyde 3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP) are produced] or on lactate *plus* glycerol (which are precursors of GAP and DHAP, respectively; Blacklow, 1990). In contrast, DF502 cells cannot survive on lactate alone or glycerol alone as the sole carbon source. Indeed, the only way DF502 cells can live on either lactate or glycerol is if they carry a plasmid that expresses a functional TIM gene. Plasmid-mediated expression of wild-type TIM, when transcription is driven from the *trc* promoter, complements the TIM deficiency of DF502 and permits growth of these cells on *either* lactate or glycerol alone (Hermes et al., 1990). Transformants expressing as little as 1−10 picounits/cell of TIM activity can grow on lactate but not on glycerol (Cooper, 1984; Blacklow, 1990).

*Assessment of Catalytic TIM Activity from Split TIM Genes.* DF502 cells were transformed with pBSBLB-1 (which harbors split gene 1) and plated on minimal medium supplemented with lactate alone. After 5 days, transformants carrying pBSBLB-1 produced colonies (Table 1A). Analogous experiments were conducted with pBSBLB-2 and pBSBLB-3, yielding similar results (Table 1A). It thus appears that expression of any of the three split genes behind an effectively constitutive promoter complements the TIM deficiency of DF502 when grown on a selective medium. When the plasmid DNA was isolated from these surviving colonies and sequenced, it was shown that the fragmented TIM gene indeed retained the split sequence.

The expression of only one of the individual gene fragments of split 1, *i.e.*, either *1a* or *1b*, does not complement the TIM deficiency of DF502 when grown on lactate plates. No colonies grew after 10 days when DF502 was transformed with either the first or the second gene fragment (from pBSBLB-1a or pBSBLB-1b, respectively) and grown on lactate plates (Table 1B). Analogous experiments were conducted for split genes 2 and 3, yielding similar results (Table 1B). The catalytic activity responsible for the complementation of DF502 on selective media therefore results from the presence of *both* peptide fragments in each case: the ability to complement the TIM deficiency of DF502 is not conferred by either fragment alone.

When DF502 is transformed with plasmids containing the genes of intact isomerases carrying point mutations, the assayable catalytic activity roughly correlates both with the amount of time needed for colonies to grow and with the size of the resulting colonies (Hermes et al., 1990). Colonies producing more active isomerases grow more quickly and are larger than those producing more sluggish enzymes. However, since western blotting experiments showed that

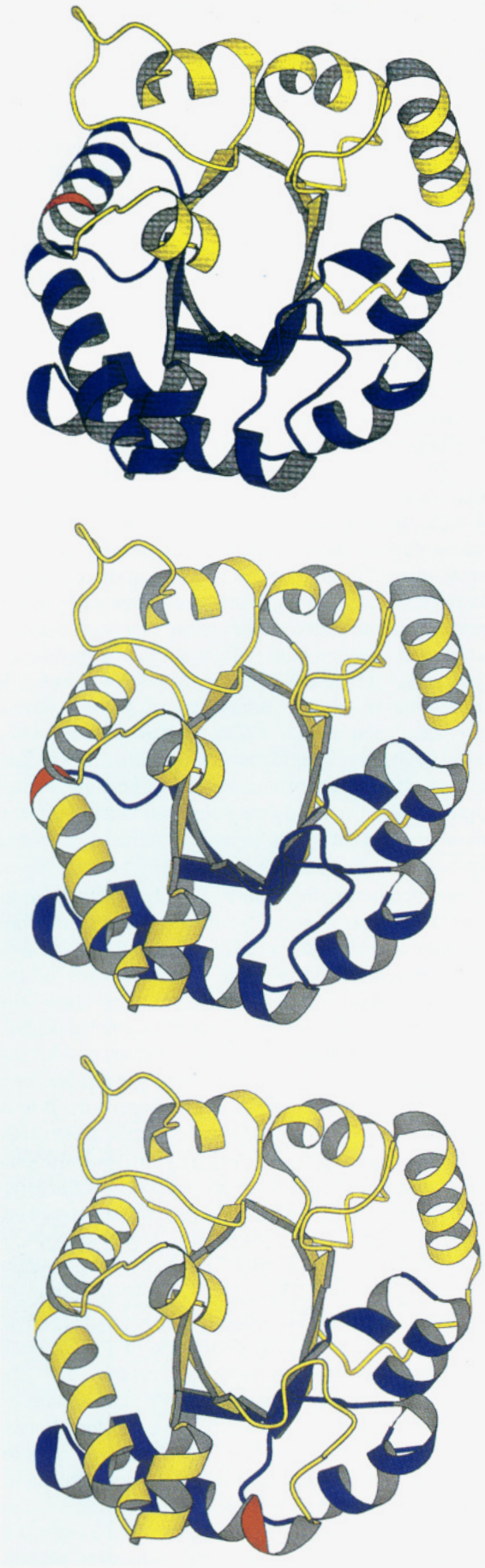

FIGURE 4: View down the $\beta$-barrel core of the monomer of chicken triosephosphate isomerase. Exon/exon junctions are highlighted in red, peptide fragment **a** is shown in light yellow, and peptide fragment **b** is shown in blue. Panels: (A, top) protein fragmentation encoded by split gene 1; (B, middle) protein fragmentation encoded by split gene 2; (C, bottom) protein fragmentation encoded by split gene 3.

Table 1: Growth Patterns of DF502 Transformants

| | glycerol + lactate | lactate alone |
|---|---|---|
| (A) Split-Gene Plasmids | | |
| DF502 (untransformed) | + | − |
| DF502 pBSTIM | + | + |
| DF502 pBSBLB-1 | + | + |
| DF502 pBSBLB-2 | + | + |
| DF502 pBSBLB-3 | + | + |
| (B) Single-Cistron Plasmids | | |
| DF502 pBSBLB-1a | + | − |
| DF502 pBSBLB-1b | + | − |
| DF502 pBSBLB-2a | + | − |
| DF502 pBSBLB-2b | + | − |
| DF502 pBSBLB-3a | + | − |
| DF502 pBSBLB-3b | + | − |

the two peptide fragments were expressed from each split gene quite disproportionately (at least a 10-fold difference; data not shown), this correlation could not be used to estimate the specific catalytic activity of the split TIM enzymes. Thus, we can conclude that either (i) the split isomerases have a specific activity somewhat lower than that of the H95N mutant (which has a specific activity $10^4$ lower than that of the wild-type enzyme and takes about 4 days to grow on lactate plates) or (ii) the complex of the two fragments has a much higher specific activity (cells containing the wild-type enzyme take 1 day to grow on lactate plates) but the low expression of the second fragment severely limits assembly of the active complex. While we cannot quantitate the specific catalytic activity of the complexes produced, the *in vivo* complementation results show that sufficient catalytic activity derives from these "rudimentary" TIMs to complement TIM-deficient cells.

*Determination of the Activity of the Peptide Fragment Complex.* When whole cell extracts from cells transformed with any of the split gene plasmids (pBSBLB-1, -2, or -3) are subjected to SDS−PAGE, it is evident that the "upstream" fragment is expressed at a much higher level than the "downstream" fragment. Efforts to modify the intercistronic region did not alter the uneven expression of the two peptide products of split gene 1 (Bertolaet, 1993). To assess the specific catalytic activity of the protein complex encoded by split gene 1, each of the peptide fragments was separately purified. To achieve this, the peptide fragments were expressed individually as fusion proteins with glutathione *S*-transferase (GST) (Simons & Vander Jagt, 1977; Smith & Johnson, 1988). Using the protease factor Xa, which cleaves after the four amino acid recognition site -IEGR-, the peptides can be released from the GST fusion proteins without extraneous amino acids at their N-termini. The gene fragments encoding the peptide products of split gene 1 were therefore individually subcloned into the vector pGEX1. The codons that encode the factor Xa cleavage site were introduced at the 5′ end of the gene fragment by amplification using PCR with a modified primer such that the coding sequence for the exon peptide product would begin immediately after the protease cleavage site and would remain in frame.

The fusion proteins were initially purified from BL21-(DE3) transformants harboring either pGEXBLB-1a or

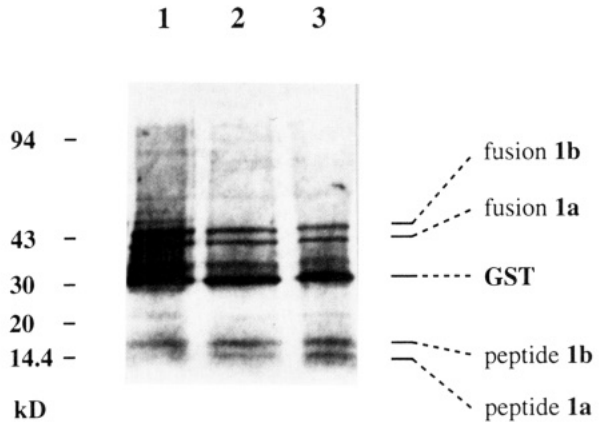


FIGURE 5: SDS−PAGE showing the time course of the simultaneous cleavage of fusion proteins **1a** and **1b** by factor Xa. The release of peptide fragments **1a** (11.8 kDa) and **1b** (15.5 kDa) from the specific cleavage of fusion proteins **1a** (39.3 kDa) and **1b** (43.0 kDa), respectively, is indicated in lanes 2 and 3. The molecular mass of GST is 27.5 kDa. Lanes: 1, cleavage reaction of fusion proteins **1a** and **1b** with factor Xa at time 0; 2, cleavage reaction of fusion proteins **1a** and **1b** with factor Xa after 8 h; 3, cleavage reaction of fusion proteins **1a** and **1b** with factor Xa after 16 h.

pGEXBLB-1b, containing peptide fragments **1a** and **1b**, respectively, by affinity chromatography on glutathione−cross-linked agarose beads. Endogenous *E. coli* BL21(DE3) TIM activity was eliminated from the fusion proteins by anion-exchange chromatography and by treatment with the isomerase inactivator bromohydroxyacetone phosphate (BHAP; de la Mare et al., 1972). Because BHAP rapidly inactivates TIM *and* has a relatively short lifetime in aqueous solution, any endogenous *E. coli* TIM activity is eliminated, and the BHAP is degraded, well before any mixing of the two fusion proteins or their cleavage products.

After treatment with BHAP but before mixing, the individual fusion proteins were tested for catalytic activity to ensure the complete absence of isomerase activity. Fusion proteins **1a** and **1b**, expressed from vectors pGEXBLB-1a and pGEXBLB-1b, respectively, were then mixed and treated with factor Xa protease. The fusion proteins after mixing but prior to cleavage had no isomerase activity. The cleavage reaction was monitored by SDS−PAGE, and samples were tested for catalytic activity throughout the 16-h time course of the cleavage reaction. This analysis showed that the fusion proteins were being cleaved, but no catalytic activity was detected. The two released fragments were evidently not spontaneously folding together to form an active complex. Since prior denaturation is sometimes required before the proper three-dimensional structure and catalytic activity can be observed (Saint Girons et al., 1987; Burbaum & Schimmel, 1991; Pookanjanatavip et al., 1992), the products of the cleavage reaction were denatured in guanidine hydrochloride and then dialyzed to allow refolding and complexation. From these experiments, the mixture of cleaved peptide fragments showed easily detectable levels of TIM activity.

The amount of released peptide relative to the total amount of protein in the fusion protein mixture prior to cleavage was estimated by densitometric scanning of SDS−polyacrylamide gels (Figure 5). Given that the yield of folded peptides after denaturation and dialysis could not be readily determined, the observed specific catalytic activity of the peptide complex is only a lower limit. This lower limit was found to be 0.1 unit/mg, roughly 75 000-fold less than that of the wild-type enzyme (Blacklow, 1990).

Control experiments were conducted in which the fusion proteins **1a** and **1b** were cleaved individually under the same conditions. Upon dialysis of these denatured samples, no isomerase activity was detected, showing that the TIM catalytic activity detected in the mixed samples derives from the presence of *both* peptides **1a** and **1b** and confirming that the BHAP inactivation of the endogenous BL21(DE3) isomerase activity during the fusion protein purification is irreversible under the conditions used.

## CONCLUSION

The relevance of the work presented in this paper to the plasticity of protein structure with regard to catalytic function is clear. The location of the splits at exon boundaries allows for speculation about the evolution of proteins, in the context of the exon microgene theory. The possibility of exon product complementation has been investigated by introducing single splits into a eukaryotic protein at three different exon/exon junctions. While there are many examples of fragment complementation studies in the literature, none has been concerned with proteins split at sites corresponding to exon junctions. The focus of earlier studies was either on protein folding or on the robustness of protein tertiary structure (Anfinsen et al., 1971; Richards et al., 1971; Holmgren & Slaby, 1979; Galakatos & Walsh, 1987; Saint Girons et al., 1987). The studies described here, in contrast, have used peptide fragment complementation in an effort to recreate a possible early step in the evolution of functional proteins.

One interpretation of the observed catalytic activity of the fragmented isomerases is that we have generated a "primitive" enzyme, the catalytic activity of which derives from the complementation of the two copies of each of two peptides. Although the fragmented enzymes show only modest activity, the fact that the matrix of assembled peptides is active at all is significant. We know that the architecture of the active site and the position of active site residues need to be precise if the activity of the wild-type enzyme is to be approached (Knowles, 1991a). For example, moving the active site base by roughly 1 Å results in a $10^3$-fold loss of activity (Raines et al., 1986); reducing the size of the active site lid produces a $10^5$-fold loss in $k_{cat}$ (Pompliano et al., 1990); and replacing the active site histidine with an asparagine results in a $10^4$-fold loss in catalytic potency (Blacklow & Knowles, 1990).

The peptide fragment complexes have enough catalytic activity to overcome the TIM deficiency of *E. coli* strain DF502 under selective conditions and catalyze the TIM reaction at least $10^4$-fold more effectively than acetate ion (Richard, 1984). A rudimentary enzyme has evidently been created, the activity of which lies about halfway between the rate of the nonenzymic isomerization and that of the diffusion-limited enzyme that exists today. While we must be careful not to overinterpret these results in the context of any particular theory, our findings are consistent with the view that the first catalytically active entities could have derived from the assembly of peptide fragments translated from what we now call exons.

## ACKNOWLEDGMENT

We thank Shari Wang and Andrea Wong for their technical assistance and Elizabeth Komives for her help in rendering the ribbon structures.

## REFERENCES

Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D., Rivers, P. S., & Wilson, I. A. (1981) *Philos. Trans. R. Soc. London B293*, 159−171.
Alber, T. C., Davenport, R. C., Jr., Giammona, D. A., Lolis, E., Petsko, G. A., & Ringe, D. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 603−613.
Anfinsen, C. B., Cuatrecasas, P., & Taniuchi, H. (1971) *Enzymes, 3rd Ed. 4*, 177−204.
Babul, J. (1978) *J. Biol. Chem. 253*, 4350−4355.
Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., & Waley, S. G. (1975) *Nature 255*, 609−614.
Belasco, J. G., & Knowles, J. R. (1980) *Biochemistry 19*, 472−477.
Berget, S. M., Moore, C., & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3171−3175.
Bertolaet, B. L. (1993) Ph.D. Thesis, Harvard University, Cambridge, MA.
Blacklow, S. C. (1990) Ph.D. Thesis, Harvard University, Cambridge, MA.
Blacklow, S. C., & Knowles, J. R. (1990) *Biochemistry 29*, 4099−4108.
Burbaum, J. J., & Schimmel, P. (1991) *Biochemistry 30*, 319−324.
Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. T. (1977) *Cell 12*, 1−8.
Cooper, R. A. (1984) *Annu. Rev. Microbiol. 38*, 49−68.
Craik, C. S., Sprang, S., Fletterick, R., & Rutter, W. J. (1982) *Nature 299*, 180−182.
Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A., & Ringe, D. (1991) *Biochemistry 30*, 5821−5826.
de la Mare, S., Coulson, A. F. W., Knowles, J. R., Priddle, J. D., & Offord, R. E. (1972) *Biochem. J. 129*, 321−331.
Eder, J., & Kirschner, K. (1992) *Biochemistry 31*, 3617−3625.
Farber, G. K., & Petsko, G. A. (1990) *Trends Biochem. Sci. 15*, 228−234.
Federov, A., Suboch, G., Bujakov, M., & Federova, L. (1992) *Nucleic Acids Res. 20*, 2553−2557.
Fraenkel, D. (1986) *Annu. Rev. Blochem. 55*, 317−337.
Galakatos, N. G., & Walsh, C. T. (1987) *Biochemistry 26*, 8475−8480.
Gilbert, W. (1978) *Nature 271*, 501.
Gilbert, W. (1985) *Science 228*, 823−824.
Hermes, J. D., Blacklow, S. C., & Knowles, J. R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 597−602.
Hermes, J. D., Parekh, S. M., Blacklow, S. C., Koster, H., & Knowles, J. R. (1989) *Gene 84*, 143−151.
Hermes, J. D., Blacklow, S. C., & Knowles, J. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 696−700.
Hoffman, S. J., Looker, D. L., Roehrich, J. M., Cozart, P. E., Durfee, S. L., Tedesco, J. L., & Stetler, G. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8521−8525.
Holmgren, A., & Slaby, I. (1979) *Biochemistry 18*, 5591−5599.
Knowles, J. R. (1991a) *Nature 350*, 121−124.
Knowles, J. R. (1991b) *Philos. Trans. R. Soc. London B332*, 115−121.
Komives, E. A., Chang, L. C., Lolis, E., Tilton, R. F., Petsko, G. A., & Knowles, J. R. (1991) *Biochemistry 30*, 3011−3019.
Li, C. H., & Bewley, T. A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1476−1479.
Lodi, P. J., & Knowles, J. R. (1991) *Biochemistry 30*, 6948−6956.
Lodi, P. J., & Knowles, J. R. (1993) *Biochemistry 32*, 4338−4343.
Lolis, E., & Petsko, G. A. (1990) *Biochemistry 29*, 6619−6625.
Marchionni, M., & Gilbert, W. (1986) *Cell 46*, 133−141.
Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Nakamaye, K., & Eckstein, F. (1986) *Nucleic Acids Res. 14*, 9679−9698.
Nickbarg, E. B., Davenport, R. C., Petsko, G. A., & Knowles, J. R. (1988) *Biochemistry 27*, 5948−5960.
Omer, C. A., Kral, A. M., Diehl, R. E., Prendergast, G. C., Powers, S., Allen, C. M., Gibbs, J. B., & Kohl, N. E. (1993) *Biochemistry 32*, 5167−5176.
Phillips, D. C., Rivers, P. S., Sternberg, M. J. E., Thornton, J. M., & Wilson, I. A. (1977) *Biochem. Soc. Trans. 5*, 642−647.
Plaut, B., & Knowles, J. R. (1972) *Biochem. J. 129*, 311−320.
Pompliano, D. L., Peyman, A., & Knowles, J. R. (1990) *Biochemistry 29*, 3186−3194.
Pookanjanatavip, M., Yuthavong, Y., Greene, P. J., & Santi, D. V. (1992) *Biochemistry 31*, 10303−10309.
Putman, S. J., Coulson, A. F. W., Farley, I. R. T., Riddleston, B., & Knowles, J. R. (1972) *Biochem. J. 129*, 301−310.
Raines, R. T., Sutton, E. L., Straus, D. R., Gilbert, W., & Knowles, J. R. (1986) *Biochemistry 25*, 7142−7154.
Richard, J. P. (1984) *J. Am. Chem. Soc. 106*, 4926−4936.
Richards, F. M., & Wyckoff, H. W. (1971) *Enzymes, 3rd Ed. 4*, 647−806.
Saint Girons, I., Gilles, A.-M., Margarita, D., Michelson, S., Monnot, M., Fermandjian, S., Danchin, A., & Barzu, O. (1987) *J. Biol. Chem. 262*, 622−629.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Sayers, J. R., Schmidt, W., & Eckstein, F. (1988) *Nucleic Acids Res. 16*, 791−802.
Seidel, H. M., Pompliano, D. L., & Knowles, J. R. (1992) *Science 257*, 1489−1490.
Shapiro, M. B., & Senapathy, P. (1987) *Nucleic Acids Res. 15*, 7155−7174.
Simons, P. C., & Vander Jagt, D. L. (1977) *Anal. Biochem. 82*, 334−341.
Smith, D. B., & Johnson, K. S. (1988) *Gene 67*, 31−40.
Straus, D., & Gilbert, W. (1985) *Mol. Cell. Biol. 5*, 3497−3506.
Straus, D., Raines, R., Kawashima, E., Knowles, J. R., & Gilbert, W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2272−2276.
Taylor, J. W., Schmidt, W., Cosstick, R., Okruszek, A., & Eckstein, F. (1985a) *Nucleic Acids Res. 13*, 8749−8764.
Taylor, J. W., Ott, J., & Eckstein, F. (1985b) *Nucleic Acids Res. 13*, 8765−8785.
Toyama, H., Tanizawa, K., Yoshimura, T., Asano, S., Lim, Y.-H., Esaki, N., & Soda, K. (1991) *J. Biol. Chem. 266*, 13634−13639.

BI9417933