Biochem. J. (1985) **227**, 1003–1007
Printed in Great Britain

Terminal deoxynucleotidyltransferase

Alignment of α- and β-subunits of the core enzyme along the primary translation product

Carol M. BEACH, S. K. CHAN, Thomas, C. VANAMAN and Mary Sue COLEMAN*
**Department of Biochemistry, University of Kentucky, Lexington, KY 40536-0084, U.S.A.*

(*Received 7 January 1985/20 February 1985; accepted 26 February 1985*)

Terminal deoxynucleotidyltransferase exists in multiple M_r forms, all apparently generated from a single polypeptide of 62kDa. On isolation and purification, the smallest catalytically active protein of this enzyme consists of two subunits, α (12kDa) and β (30kDa). Recently a complementary-DNA nucleotide sequence has been reported for a portion of the enzyme from human lymphoblast. We have pinpointed the locations of the α- and β-subunits within the elucidated nucleotide sequence. From these data, the portions of the nucleotide sequence coding for the catalytically important area of the transferase can be estimated. Here the amino acid sequence of a number of tryptic peptides from calf α- and β-subunits is presented. Because of the striking homology between the amino acid sequence of the calf enzyme and that predicted for human lymphoblast enzyme, it is possible for us to conclude that the α-subunit was generated from the *C*-terminus of the precursor protein and the β-subunit was non-overlapping and proximal.

Terminal deoxynucleotidyltransferase is a non-template-directed DNA polymerase expressed exclusively in pre-lymphocytes (in thymus and marrow) and in their neoplastic counterparts (lymphoblastic leukaemic cells) (reviewed in Coleman & Deibel, 1983). [The enzyme name has previously been abbreviated to 'TdT' (Bollum, 1978), but this is avoided here because of possible confusion with the dinucleotide similarly symbolized.] This enzyme has been detected in mammalian and avian species, but not in lower eukaryotes or prokaryotes (Goldschneider *et al.*, 1977). The biological function of the enzyme has not been elucidated. However, since this enzyme synthesizes DNA without template direction and is expressed maximally in pre-B- and -T-lymphocytes during immunological programming, it has been postulated that it may play a role in the development of the immune system (Baltimore, 1974). The enzyme appers to be synthesized as a catalytically active single polypeptide of approx. 62kDa (Deibel & Coleman, 1979; Bollum & Brown, 1979). However, lower-M_r forms of the enzyme can be detected in crude extracts of thymus tissues. Specific cleavage of the transferase molecule occurs during purification to produce fully catalytically active forms of 45kDa or a two-subunit enzyme consisting of 12kDa (α) and 30kDa (β) species. Since neither the α- nor the β-subunit alone exhibits catalytic activity (Bollum & Chang, 1981; Chang *et al.*, 1982; Modak & Gillerman-Cox, 1982), it must be concluded that the smallest enzymically active form of the transferase contains one α- and one β-subunit.

Recently, a cDNA nucleotide sequence has been reported for a portion of the human lymphoblast enzyme (Peterson *et al.*, 1984). The purpose of the present paper is to elucidate the structural relationship between the α- and β-subunits and the parent protein and the locations of the subunits in the parent protein. From these data, it is possible to pin-point the portions of the nucleotide sequences coding for the catalytically important area of the transferase.

Materials and methods

Chemicals were of the highest purity commercially available or were purified before use. Reagents for amino acid analysis were obtained from Beckman (Palo Alto, CA, U.S.A.) and reagents for protein-sequence analysis were from Applied Biosystems (Foster City, CA, U.S.A.).

Abbreviations used: PMSF, phenylmethanesulphonyl fluoride; cDNA, complementary DNA.

* To whom correspondence and reprint requests should be addressed.

Amino acid phenylthiohydantoin derivatives used for calibration purposes and trifluoroacetic acid were purchased from Pierce Chemical Company (Rockford, IL, U.S.A.). Carboxypeptidase A (PMSF-treated) was from Worthington Diagnostic Systems (Freehold, NJ, U.S.A.).

Enzyme, subunit and peptide purification

Calf thymus terminal deoxynucleotidyltransferase was purified as previously described (Deibel & Coleman, 1980). The protein obtained was shown to be homogeneous by catalytic specific activity (> 100000 units/mg) and by electrophoresis under native and denaturing conditions (Laemmli, 1970).

Subunits were isolated by reverse-phase chromatography on a Hewlett–Packard 1084B liquid chromatograph using a 4.6mm × 7.5cm Beckman Ultrapore RPSC (C_3) column. A linear gradient of 0.1% (v/v) trifluoroacetic acid in water versus 0.1% trifluoroacetic acid in 70% acetonitrile was run at room temperature with a flow rate of 1 ml/min, and the column effluent was monitored at 282nm with a reference wavelength of 500nm. A 6nmol portion of the transferase α-subunit was dissolved in 500μl of 0.05M-NH_4HCO_3 and was digested with four portions of diphenylcarbamyl chloride-treated trypsin (trypsin/protein, 1:20, w/w). Digestion by each portion of trypsin was allowed to continue for 4h at 37°C. A 10nmol portion of the transferase β-subunit was treated in a similar fashion by using two portions of trypsin (trypsin/protein, 2:25, w/w). The solutions were freeze-dried, dissolved in 0.1% trifluoroacetic acid and peptides were separated by reverse-phase h.p.l.c. on a column (4.6mm × 25cm) of Supelco LC-318. The solvents, flow rate and temperature used were the same as those used for subunit separation.

Amino acid sequence determination

Automated Edman degradation was performed by using an Applied Biosystems gas-phase sequencer operated with non-vacuum-drying methodology as currently supplied. Amino acid phenylthiohydantoin derivatives in sequencer cycles were identified and quantified by using a du Pont Zorbax ODS (4.6mm × 25cm) column and a Waters h.p.l.c. system consisting of two 6000A pumps, a λ max Model 490 detector, M730 data module, model 720 system controller and WISP 710B automatic injector. Identification of amino acid phenylthiohydantoin derivatives was similar to the method of Hunkapiller & Hood (1978).

Carboxypeptidase digestion of performic acid-oxidized subunits

Subunits of the transferase were individually oxidized with performic acid (Hirs, 1967). The dried oxidized protein was dissolved in 100μl of 0.2M-*N*-ethylmorpholine/acetate, pH8.5, and digested with carboxypeptidase A (Ambler, 1972). Free amino acids in the digestion mixture were analysed with a Beckman model 6300 amino acid analyser using sodium citrate buffers A (67mM, pH3.25), B (0.13M, pH3.95) and D (0.3M, pH6.40), with the standard 1h analysis program. The resulting values were corrected for amino acid values obtained from a control digestion.

Results and discussion

The α- and β-subunits were separated by reverse-phase chromatography as described in the Materials and methods section. The *N* termini of both subunits were not amenable to Edman degradation and were probably blocked. The inaccessibility of both *N*-termini precluded sequencing intact subunits. Therefore the amino acid sequences of tryptic peptides from both subunits were determined. The primary structure obtained for 16 peptides (eight from α-subunit and eight from β-subunit) ranging in size from a tripeptide to a peptide of 24 amino acid residues is shown in Table 1.

All eight peptides sequenced for α-subunit could be aligned with sequences previously deduced from a partial human cDNA. In all, 43 of the 45 amino acid residues contained in these eight peptides showed complete homology with the amino acid sequence predicted from the partial human cDNA. The peptides were aligned in a cluster just preceding the first predicted stop codon (Fig. 1). Of the peptides generated from the β-subunit, eight were sequenced (64 residues in total, see Table 1). Three of these (β-subunit, nos. 6–8) could be aligned with the partial human cDNA sequence (Fig. 1). The other five peptides sequenced were not found within the portion of the enzyme encoded by the partial human cDNA. Thus either these five peptides were located in the portion of β-subunit for which there is not as yet a nucleotide sequence available or they were derived from regions of the bovine protein that were not homologous with the human enzyme.

The first termination codon detected in the partial human cDNA is preceded by an alanine codon, and then, in reverse order, asparagine, arginine, glutamic, tryptophan and proline codons. α-Subunit tryptic peptide 5 included sequence up to the arginine residue adjacent to the first predicted stop codon (see Table 1). When 1 nmol of intact α-subunit was digested with carboxypeptidase A, 0.63nmol of alanine and 0.58nmol of asparagine were released, indicating that the *C*-terminal amino acid residue was either alanine or asparagine. As these were the last two amino acids predicted from the partial human cDNA before

Table 1. *Amino acid sequences corresponding to tryptic peptides of terminal deoxynucleotidyltransferase α- and β-subunits*
The bovine tryptic peptides were sequenced as detailed in the Materials and methods section. For each peptide, line (*a*) represents the predicted human sequence from a cDNA; (*b*) represents the bovine sequence obtained in this study; and (*c*) represents the amount (in pmol) of amino acid phenylthiohydantoin derivative obtained. Abbreviation: Xaa, amino acid identity not determined for the bovine peptide.

(i) α-Subunit

Peptide		Sequence
α1	(*a*)	Thr-Trp-Lys
	(*b*)	Thr-Trp-Lys
	(*c*)	240 220 630
α2	(*a*)	Tyr-Ala-Thr-His-Glu-Arg
	(*b*)	Tyr-Ala-Thr-His-Glu-Arg
	(*c*)	1750 450 130 100 50 10
α3	(*a*)	Ala-Leu-Tyr-Asp-Lys
	(*b*)	Ala-Leu-Tyr-Asp-Lys
	(*c*)	750 780 630 290
α4	(*a*)	Ala-Glu-Ser-Glu-Glu-Glu-Ile-Phe-Ala
	(*b*)	Ala-Glu-Ser-Glu-Glu-Glu-Ile-Phe-Ala
	(*c*)	280 280 20 190 210 250 110 180 150
α5	(*a*)	Leu-Gly-Leu-Asp-Tyr-Ile-Glu-Pro-Trp-Glu-Arg
	(*b*)	Leu-Gly-Leu-Asp-Tyr-Ile-Glu-Pro-Trp-Glu-Arg
	(*c*)	220 130 150 100 70 50 70 40 6 30
α6	(*a*)	Ile-Phe-Leu-Lys
	(*b*)	Val-Phe-Leu-Lys
	(*c*)	780 230 170 60
α7	(*a*)	Met-Ile-Leu-Asp-Asn-His
	(*b*)	Met-Met-Leu-Asp-Asn-His
	(*c*)	370 320 320 480 200 10
α8	(*a*)	Ala-Ile-Arg
	(*b*)	Ala-Ile-Arg
	(*c*)	440 460 420

(ii) β-Subunit

Peptide		Sequence
β1	(*a*)	Human predicted sequence undetermined
	(*b*)	Glu-Ala-Val-Trp-Met-Lys
	(*c*)	650 530 480 140 210 160
β2	(*a*)	Human predicted sequence undetermined
	(*b*)	Leu-Phe-Thr-Ser-Val-Phe
	(*c*)	480 460 60 20 120 60
β3	(*a*)	Human predicted sequence undetermined
	(*b*)	Tyr-Gln-Ser-Phe-Lys
	(*c*)	540 230 40 460 60
β4	(*a*)	Human predicted sequence undetermined
	(*b*)	Met-Gly-Phe-Arg
	(*c*)	1080 610 1070 110
β5	(*a*)	Human predicted sequence undetermined
	(*b*)	Phe-Thr-Lys
	(*c*)	810 200 980
β6	(*a*)	Ala-Phe-Leu-Pro-Asp-Ala-Phe-Val-Thr-Met
	(*b*)	Ala-Phe-Leu-Pro-Asp-Ala-Phe-Val-Xaa-Met
	(*c*)	300 200 150 100 60 60 30 30 10
β7	(*a*)	Met-Gly-His-Asp-Val-Asp-Phe-Leu-Ile-Thr-Ser-Pro-
	(*b*)	Ile-Gly-His-Asp-Val-Asp-Phe-Leu-Ile-Thr-Ser-Pro-
	(*c*)	250 200 110 190 180 130 220 120 150 40 10 90
		Gly-Ser-Thr-Glu-Asp-Glu-Glu-Gln-Leu-Leu-Gln-Lys
		Gly-Ser-Ala-Glu-Asp-Glu-Glu-Gln-Leu-Leu-Pro-Lys
		90 10 90 80 30 50 70 20 40 20 10 5
β8	(*a*)	Val-Met-Asn-Leu-Trp-Glu-Lys
	(*b*)	Val-Ile-Asn-Leu-Trp-Glu-Lys
	(*c*)	640 410 510 470 510 350 280

the first predicted stop codon, we conclude that the α-subunit occupied the *C*-terminal portion of the protein predicted from the partial human cDNA, and that it extended to the first predicted stop codon.

Our peptide alignment has shown that the β-subunit preceded the α-subunit in the partial human cDNA sequence and that the α-subunit extended to the first predicted stop codon. If the 62kDa parent form of the enzyme were the gene product of the entire cDNA, as predicted, then the α-subunit would be located at the *C*-terminus of the parent protein and the β-subunit would be located just before the α-subunit.

In order to determine a probable location for the *N*-terminus of the α-subunit, we calculated that the α-subunit contained approx. 85 amino acids. This calculation was based on a molecular mass of 12kDa as determined by dodecyl sulphate/polyacrylamide gel electrophoresis (Deibel & Coleman, 1980) and on amino acid analysis of the α-subunit (Chang & Bollum, 1971; C. M. Beach, unpublished work). A potential *N*-terminus for the α-subunit is a glutamine residue preceded by a tryptophan residue in the deduced partial sequence of the human enzyme (arrow at position 551, Fig. 1). Since it is known that *N*-terminal glutamine may cyclize to pyrrolidinecarboxylic acid, particularly at acidic pH (Blomback, 1967), the assignment of glutamine as the *N*-terminus of the α-subunit would account for the block to Edman degradation.

To identify the *C*-terminus of the β-subunit, preliminary carboxypeptidase analysis of the intact β-subunit was performed. The data obtained suggested that the β-subunit may possess a *C*-terminal arginine that could place it within nine amino acid residues of the *N* terminus of the α-subunit. However, the data were not sufficient to be unequivocal. Regardless, the 55 amino acid residues between the last β-subunit peptide sequenced and the first α-peptide sequenced were insufficient to account for all of the protein excised from the 62kDa parent protein. A portion of the *N*-terminus of the parent protein must have been removed to give rise to the α- and β-subunits.

The protein-sequence data we have generated for the α-subunit indicated almost total sequence homology (95%) beween calf and human enzymes. Whereas the peptides sequenced for β-subunit also indicated a high degree of homology (90%), a large portion of the β-subunit remains to be sequenced and may be more variable than the α-subunit in different species. These data suggest that the amino acid sequences surrounding the active site of the enzyme, located on either the α- or β-subunit or on a combination of both, are highly conserved. High-molecular-mass forms of the protein contain portions that are not absolutely required for enzyme activity, and these portions may diverge in amino acid sequence between species, as evidenced by the existence of non-cross-reacting monoclonal antibodies that have been produced against both human and calf enzyme (Fuller *et al.*, 1983; Bollum *et al.*, 1984).

Having located the α- and β-subunits on the gene

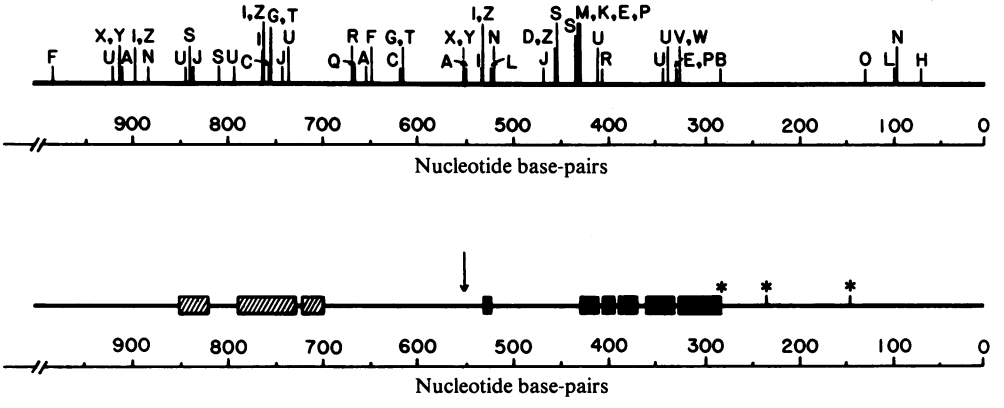

Fig. 1. *Alignment of the α- and β-subunits of calf terminal deoxynucleotidyltransferase along the primary sequence of the human enzyme predicted from a cDNA nucleotide sequence*

The top line shows the locations of restriction-endonuclease sites determined from the reported nucleotide sequence of the fragment. The identities of the restriction enzymes are as follows: A, *Alu*I; B, *Avr*II; C, *Bln*I; D, *Cau*II; E, *Cfo*I, F, *Dde*I; G, *Dpn*I, H, *Eco*RI; I, *Eco*RII; J, *Fok*I; K, *Gdi*II; L, *Hae*I; M, *Hae*II; N, *Hae*III; O, *Hqi*CI; P, *Hha*I; Q, *Hin*dII; R, *Hin*fI; S, *Hpa*II; T, *Mbo*I; U, *Mbo*II; V, *Mlu*I, W, *Mst*I; X, *Nsp*BII; Y, *Pvu*II; Z, *Scr*FI. The lower shows the locations of the sequenced α- and β-peptides along the translated cDNA fragment. The arrow found near the centre of the lower line indicates the probable start of the α-subunit coding region. ■, α-subunit peptides; ▨ β-subunit peptides; *, stop codon.

and knowing that the enzyme catalytic site is contained within these subunits will aid future studies to definitively identify the active site by mutagenesis *in vitro* or by studies in which enzyme peptides are modified by active-site-directed agents. Thus terminal deoxynucleotidyltransferase may be the first eukaryotic DNA polymerase readily amenable to detailed studies of the catalytic domain.

This work was supported by the National Cancer Institute (grants CA19492 and CA26391). C. M. B. is the recipient of a fellowship from the Physicians' Service Plan. We are grateful for the technical expertise of Mr. Steven Gathy, Ms. Yong Hui Ahn and the secretarial assistance of Ms. Sara Thompson.

References

Ambler, R. P. (1972) *Methods Enzymol.* **25**, 143–154
Baltimore, D. (1974) *Nature (London)* **248**, 409–411
Blomback, B. (1967) *Methods Enzymol.* **11**, 398–413
Bollum, F. J. (1978) *Adv. Enzymol.* **47**, 347–374
Bollum, F. J. & Brown, M. (1979) *Nature (London)* **278**, 191–192
Bollum, F. J. & Chang, L. M. S. (1981) *J. Biol. Chem.* **256**, 8767–8770
Bollum, F. J., Augl, C. & Chang, L. M. S. (1984) *J. Biol. Chem.* **259**, 5848–5850
Chang, L. M. S. & Bollum, F. J. (1971) *J. Biol. Chem.* **246**, 909–916
Chang, L. M. S., Plevani, P. & Bollum, F. J. (1982) *J. Biol. Chem.* **257**, 5700–5706
Coleman, M. S. & Deibel, M. R., Jr. (1983) in *Enzymes of Nucleic Acid Synthesis and Modification*, vol. 1 (*DNA Enzymes*), pp. 93–118, CRC Press, Boca Raton, FL
Deibel, M. R., Jr. & Coleman, M. S. (1979) *J. Biol. Chem.* **254**, 8634–8640
Deibel, M. R., Jr. & Coleman, M. S. (1980) *Arch. Biochem. Biophys.* **202**, 414–419
Fuller, S. A., Philips, A. & Coleman, M. S. (1983) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **42**, 1973
Goldschneider, I., Gregoire, K. E., Barton, R W. & Bollum, F. J. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 734–738
Hirs, C. H. W. (1967) *Methods Enzymol* **11**, 197–199
Hunkapiller, M. W. & Hood, L. E. (1978) *Biochemistry* **17**, 2124–2133
Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
Modak, M. J. & Gillerman-Cox, E. (1982) *J Biol. Chem.* **257**, 15105–15109
Peterson, R. C., Cheung, L. C., Mattaliano, R. T., Chang, L. M. S. & Bollum, F. J. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4363–4367