Volume 280, number 2, 363–366 FEBS 09495 March 1991
© 1991 Federation of European Biochemical Societies 0014-5793/91/$3.50
ADONIS 001457939100205T

Characterization of human deoxycytidine kinase

Correlation with cDNA sequences

Staffan Eriksson[1], Ella Cederlund[2], Tomas Bergman[2], Hans Jörnvall[2] and Christina Bohman[1]

[1]*Medical Nobel Institute, Department of Biochemistry I* and [2]*Department of Chemistry I, Karolinska Institutet, Stockholm, Sweden*

Received 14 December 1990

Existing data on the structure of human deoxycytidine kinase (dCK) diverge. A monomeric 60 kDa form has been isolated and the cloning of a cDNA coding for 626 amino acids corresponding to a 71 kDa protein has been reported. However, pure dCK isolated from leukemic spleen is a dimer of 30 kDa subunits. Amino acid sequences of peptides from digests of this protein are now presented. None of the peptide structures obtained correspond to the cDNA for the 71 kDa protein, but to a cDNA for a 30.5 kDa dCK recently cloned. Furthermore, homology of the peptide sequences of dCK to parts of thymidine kinases and protein-tyrosine kinases are detected.

Deoxycytidine kinase; Peptide sequence; Amino acid composition; cDNA sequence

1. INTRODUCTION

Deoxycytidine kinase (dCK; EC 2.7.1.74) catalyzes the phosphorylation of deoxycytidine to deoxycytidine 5′-monophosphate using nucleoside triphosphates as phosphoryl donors. dCK can also phosphorylate purine deoxyribonucleosides and a large number of deoxynucleoside analogs. It has been clearly shown that the activity of dCK is a prerequisite for the antiviral and cytostatic effect of several therapeutically important nucleoside analogs.

dCK has been purified from many sources (reviewed in [1]) but it has been possible only recently to obtain apparently homogeneous preparations from leukemic spleen [2,3]. The active enzyme appeared to be a dimer of two 30 kDa subunits, as judged by SDS-polyacrylamide gel electrophoresis, gel exclusion chromatography and glycerol gradient centrifugation [3]. In accordance with these results, dCK from Molt-4 T-lymphoblasts was reported to be a dimer of two 30 kDa subunits [4]. However, other reports suggest [5,6] that native dCK from human leukemic T-lymphoblasts is a monomer of 60 kDa when prepared in the presence of protease inhibitors. In the absence of protease inhibitors, two fragments of 30 and 33 kDa, were observed, suggesting that proteolytic cleavage of a 60 kDa dCK during purification results in two shorter fragments. Furthermore, the cloning and sequence determination of a cDNA suggested to code for human dCK with 626 amino acids corresponding to a 71 kDa polypeptide has been reported [7].

To resolve these discrepancies and to obtain information for reliable cloning of the gene for human dCK, chemical characterization of the 30 kDa human spleen enzyme is essential. Here, we present data on the total composition of the protein and on peptide sequences of four segments. A correlation of these results with those published before [7] and with a very recent cDNA sequence for human dCK [8] resolves the question and unambiguously identifies the correct dCK structure.

Correspondence address: S. Eriksson, Medical Nobel Institute, Department of Biochemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden

2. MATERIALS AND METHODS

Human leukemic dCK was prepared as described [3]. The final preparation had a specific activity of 200 nmol dCMP formed/min/mg and was more than 95% pure as judged by SDS-polyacrylamide gel electrophoresis.

For analysis, protein (3–300 μg) was precipitated with 20% TCA (1 h on ice), washed twice with −20°C acetone, lyophilized, dissolved in 500 μl 8 M urea, 0.4 M Tris HCl, pH 8.5, 2 mM EDTA. Reduction was performed with 5 mM DTT (2 h at 37°C), after which the protein was carboxymethylated by addition of ^{14}C-labeled iodoacetate (15 mM) and incubated for 90 min at 37°C anaerobically in the dark. The reaction was stopped by addition of excess mercaptoethanol and the sample desalted on Sephadex G50 eluted with 30% acetic acid. After lyophilization the protein was dissolved in 1.0 ml 70% formic acid containing 0.2 g CNBr for methionine cleavage at 25°C for 24 h or dissolved in 8 M deionized urea, in 0.2 M ammonium bicarbonate, pH 8.0, and subsequently diluted to a final concentration of 1 M urea, 0.2 M ammonium bicarbonate for Lys-specific cleavage with Lys C-endoprotease (Boehringer), 4 h at 37°C with a protein:protease ratio of 10:1.

CNBr fragments were separated by HPLC on a μBondapak C_{18} column developed with a linear gradient of 0–80% acetonitrile in 0.1% aqueous trifluoroacetic acid for 80 min at a flow rate of 1 ml/min, while the Lys-cleaved peptides were separated on a Vydac C_{18} column

Table I

Total composition of the human deoxycytidine kinase analyzed

Present direct protein analysis		cDNA analysis	
Residue	mol/mol	[8]	[7]
Cys	4.4	6	6
Asx	20.4	24	71
Thr	16.1	15	25
Ser	24.0	20	34
Glx	42.7	40	80
Pro	9.9	10	39
Gly	18.0	11	32
Ala	16.1	11	48
Val	9.1[b]	12	54
Met	4.8	6	8
Ile	7.3[b]	13	29
Leu	21.8	27	47
Tyr	11.1	12	27
Phe	11.7	12	30
Trp	4.7[a]	7	6
Lys	17.1	17	60
His	4.8	5	7
Arg	13.2	12	23
Total	257	260	626

Values shown are molar ratios as determined by PTC analysis after acid hydrolysis. For comparison the values corresponding to those derived from the intial [7] and the recent [8] cDNA reports are also given, showing the disagreement of the former but agreement of the latter with the present direct protein analysis.

[a] Value obtained after hydrolysis with methane sulfonic acid

[b] Low yields upon acid hydrolysis was due to the presence of Val-Val and Ile-Ile sequences [8]

developed with a linear gradient of 0–60% acetonitrile in 0.1% trifluoroacetic acid for 75 min at 1 ml/min.

2.1. *Structural analysis*

Sequencer degradations were performed with Applied Biosystems 470A and 477A instruments. Phenylthiohydantoin derivatives were analyzed by HPLC using a sodium acetate/acetonitrile gradient on a Hewlett-Packard 1090 instrument [9] or the on-line Applied Biosystems 120A system. Total compositions were determined after PTC-modification of the amino acids as described [10].

3. RESULTS

3.1. *Protein characterization*

The amino acid composition of human dCK after affinity chromatography is shown in Table I. N-terminal sequence analysis of 400 pmol showed no significant amino acid residue for the first 4 cycles, suggesting that the N-terminus of dCK is blocked.

Reduced and carboxymethylated dCK (approximately 5 nmol) was cleaved with CNBr. Resulting peptides were separated by reverse-phase HPLC (Fig. 1A) and two peaks were analyzed for amino acid sequence. One (C1) was sufficiently pure for interpretation for 18 cycles (Fig. 1A), while the other (main peptide peak) contained at least three different peptides and could not be reliably interpreted. Another batch of dCK was subjected to digestion with the Lys-specific protease and

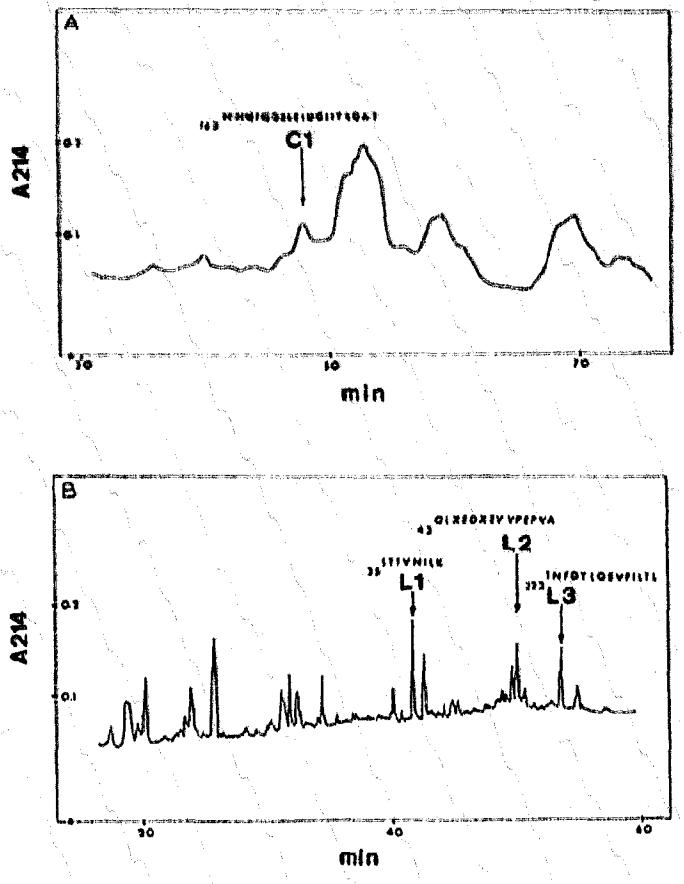

Fig. 1. HPLC chromatograms and amino acid sequences determined for peptides resulting from (A) CNBr cleavage and (B) Lys-specific cleavage of purified human dCK. Numbering of amino acid residues is that deduced from the cDNA sequence of Chottiner et al. [8] using the initiator methionine as position one. X denotes an incompletely identified amino acid residue.

fractionated by HPLC (Fig. 1B). The results of amino acid sequence analysis of peptides (L1, L2, L3) from this digest are shown in Fig. 1B.

3.2. *Correlation with cDNA sequences*

Surprisingly, none of the sequences (C1, L1, L2 or L3; Fig. 1A and B) showed more than 20% identity to any peptide sequence deduced from a previous cDNA clone [7], nor could the amino acid composition now determined be correlated with that of the entire 71 kDa peptide, cf. Table I [7] or its C-terminal or N-terminal halves.

After the completion of this analysis a second cDNA corresponding to dCK was cloned from a human T-cell library [8]. This clone could be used to transfect dCK negative mouse L-cells such that the transfectants expressed an active dCK. The deduced amino acid sequence contained 260 amino acid residues, corresponding to a protein of 30.5 kDa.

We could identify all our peptide sequences in the

deduced amino acid sequence of this new report [8] and the numbering of the amino acids in Fig. 1 refers to the positions in that cDNA-derived sequence with the initiator methionine as position one (Fig. 1).

In searches for sequence homology between the dCK peptides and other deoxynucleoside kinases characterized [12–14], we found a segment in peptide L1 that showed homology to a suggested triphosphate binding domain in the N-terminal part of human and mouse thymidine kinases I (Fig. 2A). Similarly, in peptide L3 a structure is observed that is homologous to the very C-terminal part of several human protein-tyrosine kinase oncogenes [13,15] (Fig. 2B).

4. DISCUSSION

A chemical characterization of leukemic 30 kDa, dCK was performed. We found the N-terminus to be blocked and therefore produced internal fragments by cleavage with CNBr and Lys-specific protease. After HPLC purification, several peptide sequences were determined. During the completion of this work a cDNA sequence of human dCK was published [7]. Surprisingly, no significant homology was found between that sequence and any of the peptide sequences obtained from the pure enzyme. Furthermore, the molecular weight of the cloned cDNA corresponded to a peptide of 71 kDa and we could not correlate this or the amino acid composition deduced with that of the protein we analyzed. Neither could the leukemic dCK be a proteolytic fragment of the larger protein as judged from the total composition.

Hence, we conclude that the previous cDNA report [7] concerns a protein different from dCK as recently also confirmed by Huang et al. [11], showing the 71 kDa protein to be a human stress protein homologous to murine ERp72.

However, very recently Chottiner et al. [8] have cloned and sequenced a cDNA which can be used to transfect mouse dCK-negative cells and confer an active enzyme to the transfectants. This cDNA codes for a protein corresponding to a 30.5 kDa polypeptide [8]. The deduced protein sequence does contain the four peptide sequences now described from pure leukemic dCK, and the amino acid composition deduced from the cloned cDNA agrees with that experimentally determined (Table I). Our data demonstrate that the cDNA cloned by Chottiner et al. indeed codes for the leukemic 30 kDa dCK now analyzed, and already previously isolated [3]. The fact that the cDNA-deduced molecular weight agrees with the one determined for the purified enzyme argues against major post-translational modifications of human dCK.

It is reasonable to assume that there will be regions of sequence homology between different nucleoside kinases. One such region, called the phosphate binding loop, has been identified in all known thymidine kinases and in a large number of the ATP binding proteins, e.g. protein kinases [12,13]. Peptide L1 appears to be homologous to part of this highly conserved region as shown in Fig. 2A where these sequences of human and mouse thymidine kinases are compared to that of L1 [12–15]. The sequence KST is found in all thymidine kinases and probably forms the first three amino acid residues in an α-helix that terminates the phosphate loop [13]. This hypothesis is based on models which have been constructed using the known structure of adenylate kinase to predict the structure of this region of the nucleoside kinases [14]. When the sequences of L2 and C1 were compared to other kinase sequences no significant homologies were apparent. However, the sequence LDG was found in human and mouse TK:s, starting at position 130, as well as in C1 but no other conserved amino acids were found in this peptide.

Another homology was detected in the L3 peptide which was found in the C-terminal part of the cDNA-deduced sequence [8] and showed homology to the very C-terminal part of the src-subfamily of protein-tyrosine kinases [13,16]. This proto-oncogene family, in addition to the cellular homologies of the oncogene products from Rous avian sarcoma (src), Gardner-Rasheed feline sarcoma (fgr) and the lymphoid cell protein-tyrosine kinase (LCK-JURKAT) also consists of several protein kinases with extensive similarities in primary structure. This part of the protein-tyrosine kinases corresponds to the subdomain XI in the nomenclature of Hanks et al. [13]. Arg at position 506 (Fig. 2B) is a residue that is conserved not only in this family but among all 65 protein kinases listed by Hanks et al. This region is located near the carboxyl-terminus of the catalytic domain but its function in protein kinases is thus far not known. These results indicate that dCK is part of a nucleoside kinase family that also may show structural similarity to the protein kinase family.

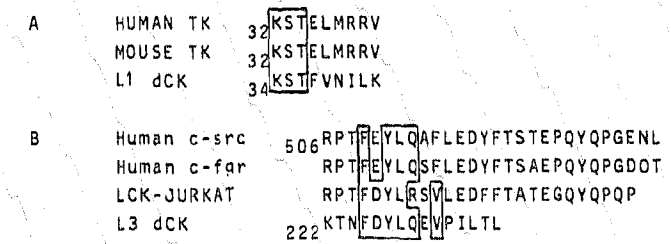

Fig. 2. Possible sequence homology (A) of peptide L1 with human and mouse thymidine kinases (TK:s) [12,14] and (B) peptide L3 with the C-terminal part of members of the src-subfamily of protein tyrosine-kinases [13,16]. The abbreviations used are explained in the text. Lys before L1 and L3 is shown because a Lys-specific protease was used to generate these peptides. Boxes indicate residue identities between characterized proteins and L1 or L3.

Acknowledgements: This work was supported by grants from the Swedish Medical Research Council, the Swedish Cancer Society, the National Board for Technical Development, Magnus Bergvall Foundation, Bengt Lundqvist, Anders Otto Swärd Foundations and the Medical Faculty of the Karolinska Institute. We thank B. Mitchell for sharing information regarding the human dCK cDNA sequence and J. Holcenberg regarding the homology of the 71 kDa protein to the murine ERp72 protein prior to publication of these results. We thank Birgitta Wallström, Anita Sjöberg Herrström and Monia Särd for excellent help.

REFERENCES

[1] Bohman, C. and Eriksson, S. (1990) Biochem. (Life Sci. Adv.) 9, 11–35.
[2] Bohman, C. and Eriksson, S. (1986) Adv. Exp. Med. Biol. 195B, 311–314.
[3] Bohman, C. and Eriksson, S. (1988) Biochemistry 27, 4258–4265.
[4] Datta, N.S., Shewach, D.S., Harley, M.C., Mitchell, B.S. and Fox, I.H. (1989) Biochemistry 28, 114–123.
[5] Kim, M.Y., Ikeda, I. and Ives, D.H. (1988) Biochem. Biophys. Res. Commun. 156, 92–98.
[6] Kim, M.Y. and Ives, D.H. (1989) Biochemistry 28, 9043–9047.
[7] Huang, S.H., Tomic, J.M., Wu, H., Jong, A. and Holcenberg, J. (1989) J. Biol. Chem. 264, 14762–14768.
[8] Chottiner, E.G., Shewach, D.S., Datta, N.S., Ashcraft, E., Gribbin, D., Ginsburg, D., Fox, I.H. and Mitchell, B.S. (1991) Proc. Natl. Acad. Sci. USA (in press).
[9] Kaiser, R., Holmquist, B., Hempel, J., Vallee, B.L. and Jörnvall, H. (1988) Biochemistry 27, 1132–1140.
[10] Bergman, T., Carlquist, M. and Jörnvall, H. (1986) In: Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J. and Erdmann, V.A., eds) pp. 45–55, Springer-Verlag, Berlin.
[11] Huang, S.-H. Gomer, C., Sun, G.-X., Wong, S., Wu, H., Lin, Y.-X. and Holcenberg, J. (1991) J. Biol. Chem., in press.
[12] Folkers, G., Trumpp, K.E., Roth, H.J. and Schwoebel, W. (1987) Med. Sci. Res. 15, 633–634.
[13] Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) Science 241, 42–52.
[14] Folkers, G. and Trumpp, S. (1987) Med. Sci. Res. 15, 1495–1496.
[15] Trevillyan, J.M., Lin, Y., Chen, S.J., Phillips, C.A., Canna, C. and Linna, T.J. (1986) Biochim. Biophys. Acta 888, 286–295.