cDNA and Amino Acid Sequence of Human Adenosine Deaminase[a]

PETER E. DADDONA, STUART H. ORKIN,[b]
DONNA S. SHEWACH, AND WILLIAM N. KELLEY

Departments of Internal Medicine and Biological Chemistry
University of Michigan Medical Center
Ann Arbor, Michigan 48109

[b]*Division of Hematology-Oncology*
Children's Hospital and the Dana Farber Cancer Institute
and
Department of Pediatrics
Harvard Medical School
Boston, Massachusetts 02115

Human adenosine deaminase (ADA) is a purine catabolic enzyme that catalyzes the irreversible hydrolytic deamination of adenosine and deoxyadenosine. The enzyme is broadly distributed in human tissues with highest activity in lymphoid cells.[1-3] Adenosine deaminase purified from human erythrocytes is a soluble monomeric protein with an estimated molecular weight of about 38,000 daltons that exhibits post-translational modification.[4,5] In all tissues the enzyme is encoded for by a single genetic locus on the long arm of chromosome 20.[6,7]

An inherited deficiency of adenosine deaminase in its most extreme form is associated with an autosomal recessive form of severe combined immunodeficiency disease (SCID).[8] Patients have severe T cell and variable B cell dysfunction. In the absence of ADA activity, immature T cells selectively accumulate and are sensitive to dATP and deoxyadenosine toxicity. Patients have recurrent severe infections that usually become fatal within the first few years of life.[9] These patients have undetectable enzyme activity in their erythrocytes and residual lymphocytes, but may have variable amounts of ADA enzymatic activity and immunoreactive protein in other tissues.[10] As evidence for genetic heterogeneity in human ADA deficiency, a smaller group of partially ADA-deficient subjects have been identified through a genetic screening program or by serendipity. Subjects have undetectable ADA activity in senescent erythrocytes but 1-20% of normal lymphocyte enzyme activity.[11-13] The partial ADA lymphocyte activity apparently protects these individuals against immunodeficiency and they remain healthy. Transformed B lymphoblast cell lines derived from these partially ADA-deficient individuals show characteristically low enzyme activity and structural mutations in the enzyme associated with electrophoretically altered[13] or heat-labile ADA protein.[12] Recent studies by Hutton and associates have described an elevated level of ADA mRNA in a B lymphoblast cell line derived from a severely and a partially ADA-deficient subject.[14,15] The mutation in both cases may be associated with an unstable ADA protein.

We have previously reported the isolation of a partial human T cell ADA cDNA gene sequence (0.8 kilobase insert).[16] Subsequently, ADA cDNA sequences have been isolated from rat, mouse, and human cDNA libraries by others.[14,17–19] In this report we describe the complete ADA primary amino acid sequence, an essentially complete T cell ADA cDNA sequence, and ADA mRNA levels in B lymphoblasts from a group of severely ADA-deficient subjects.

RESULTS AND DISCUSSION

Adenosine Deaminase Peptide Sequencing

Adenosine deaminase was purified from human erythrocyte protein using a previously described method.[5] The enzyme was reduced and cystine residues were alkylated with vinylpyridine to prevent protein cross-linking. An aliquot of enzyme was
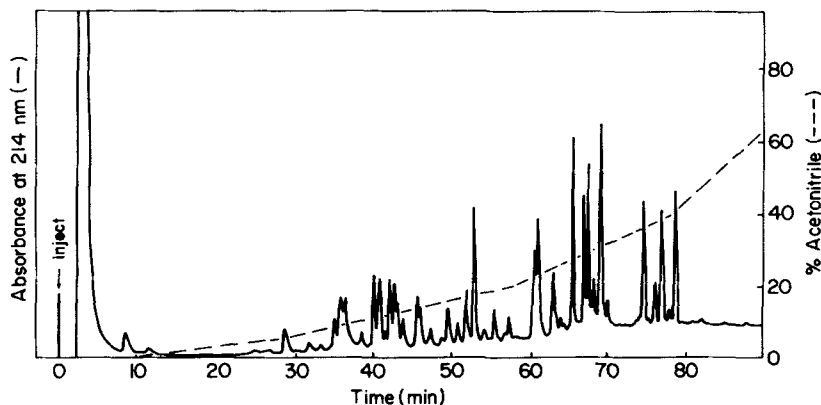
FIGURE 1. Tryptic ADA peptide separation. Pyridylethylated ADA protein was digested with TPCK-trypsin. The peptide mixture was fractionated by high-pressure liquid chromatography on a Syn Chrom RP-P (300Å, 25 cm) column using an aqueous acetonitrile gradient containing 0.1% trifluoroacetic acid.

digested with TPCK-trypsin and another aliquot with cyanogen bromide.[20] FIGURE 1 shows a representative separation of tryptic ADA peptides using reverse phase high-pressure liquid chromatography. Based on the amino acid composition of ADA, tryptic cleavage should generate a minimum of 40 peptides. We resolved 45 peptides in this analysis. Individual peptides were collected, assessed for purity, and sequenced by a manual Edman degradation technique. In a similar manner nine ADA peptides were generated by cyanogen bromide cleavage (CNBr peptides), purified, and sequenced.

Adenosine Deaminase cDNA Cloning

The cloning of human ADA cDNA sequences was undertaken using partial ADA peptide sequence information to design a specific ADA oligonucleotide probe. From

partial ADA peptide sequencing, the peptide designated CNBr8 (Gly-Phe-Thr-Glu-Glu-Glu-Phe-Lys-Arg-Leu-Asn-Ile-Asn-Ala-Ala) was determined to be suitable for construction of an oligonucleotide mixture for direct screening of recombinant cDNA clones. Human T cell human peripheral blood acute lymphoblastic leukemia mRNA was used to construct a cDNA library inserted into the PstI site of pBR322 plasmid.[16] By screening over 100,000 cDNA clones using the ^{32}P-17mer oligonucleotide as probe, a hybridization frequency of 1/1000 was observed. The clone bearing the largest cDNA insert, designated D1, was selected and sequenced. The cDNA insert was 0.8 kilobases (kb) in length. The region corresponding to the oligonucleotide probe was localized by DNA sequencing and the deduced amino acid sequence of the CNBr8 peptide as well as 16 other ADA peptides could be identified in this clone.[16] The D1 clone contained an in-phase termination signal (TGA) and encoded for the ADA C-terminal peptide (CNBr4). Given the expected size of the ADA protein (about 38,000 daltons), the D1 cloned insert represented only about two-thirds of the structure coding region of the polypeptide. Failure to isolate a full-length cDNA was probably related to difficulties in first strand DNA synthesis in the cDNA library preparation. Thus, in order to isolate the remaining ADA cDNA sequences, a new T cell cDNA library was constructed. FIGURE 2 shows a schematic of the proposed ADA cDNA
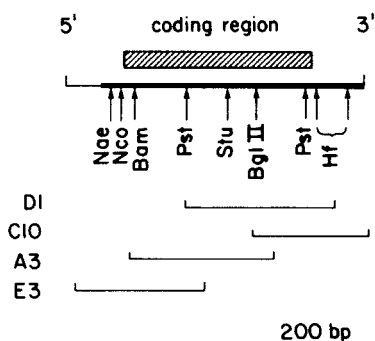

FIGURE 2. Linear map of ADA cDNA clones. The coding region from the initiator ATG to the translation terminator is hatched. D1, C10, A3, and E3 refer to various isolated cDNA clones. Restriction sites from which Maxam-Gilbert sequencing was performed are indicated. The bold portion of the restriction map includes those sequences belonging to ADA mRNA. The 5'-most 250 nucleotides of clone E3 are derived from another unknown mRNA species via a cloning artifact. (From Daddona et al.[20] Reprinted by permission of J. Biol. Chem.)

sequences for reference. The HPB-ALL cDNA library was screened using a 3' PstI restriction fragment of D1 as a probe. Clone C10 was isolated; it had a poly(A) addition signal and a short poly(A) tail designating the complete 3' end of the cDNA. Using a 5' BglII restriction fragment of clone D1 as probe, clones A3 and E3 were isolated from the library. Sequencing of DNA was performed on all clones as previously described.[20]

ADA Sequences

FIGURE 3 shows the cDNA sequence corresponding to ADA mRNA and the complete primary protein sequence deduced from sequences of cloned ADA cDNAs combined with direct ADA peptide sequencing analysis. The cDNA sequences span 1478 nucleotides, contain a consensus poly(A) addition signal and a short poly(A) tail denoting the 3' end of the cDNA. At the 5' end the consensus sequence

FIGURE 3. Adenosine deaminase cDNA and primary amino acid sequence. The DNA sequence of the strand corresponding to the mRNA is displayed above the protein sequence. Experimentally determined amino acid sequences of purified erythrocyte ADA peptides produced from either tryptic (—) or CNBr (---) cleavage are indicated. The poly(A) addition signal is indicated by a box. (From Daddona et al.[20] Reprinted by permission of J. Biol. Chem.)

CC(A or G)CCAUG(G) for translation start sites surrounds the initiator methionine codon. At the 3' end TGA is the termination signal for the polypeptide. The ADA protein is encoded for by 1088 nucleotides.

The protein sequence is 362 amino acids excluding the initiator methionine and gives a calculated molecular weight of 40,638, a value which is in good agreement with previously published estimates.[4,5,21] The N-terminal amino acid deduced from the cDNA sequence is alanine. In the purified erythrocyte ADA protein the N-terminal is blocked to Edman degradation. Many soluble erythrocyte proteins are blocked at the N-terminal amino acid and in a majority of the cases the N-terminal has been shown to be acetylated.[22] The C-terminal amino acid of ADA is a leucine. Direct amino acid sequencing determined 73% of the protein sequence and was in good agreement with the deduced sequence in all cases. The amino acid composition determined for the intact purified erythrocyte protein and that predicted from the DNA sequencing also agree well.[20]

After the completion of this study, Hutton and associates reported the cDNA and deduced amino acid sequence for human T cell ADA.[23] Our data agrees favorably with their findings and our direct analysis confirms the ADA sequences. During their cloning of T cell ADA cDNA, they encountered in a small fraction of their ADA cDNA a short nonstructure coding sequence within the structure coding portion. They propose that this sequence may represent a nonprocessed intron in the ADA mRNA.

ADA mRNA Analysis

In an attempt to gain further insight into the mechanism(s) of ADA deficiency, we have examined specific ADA mRNA levels in B lymphoblast cell lines derived from patients with severe ADA deficiency ($< 1\%$ enzyme activity) and SCID. Total mRNA was isolated from lymphoblasts and fractionated on a 1% agarose-formaldehyde gel as previously described.[20] After Northern blotting, ADA mRNA was hybridized to the nick-translated insert fragment of pHADA-1 (D1, 0.65 kb) and visualized by autoradiography.

FIGURE 4 shows a Northern blot analysis of mRNA derived from two normal B lymphoblast cell lines, GM 333 and GM 558, six lymphoblast cell lines from unrelated patients with severe ADA deficiency and SCID, GM 4258, GM 2825, GM 1715, GM 2756, GM 2606, and GM 2471, and an ADA-deficient null lymphoblast cell line, DHL-9, derived from a null cell leukemia patient.[20] As shown, the major ADA mRNA in normal cells is about 1.6 to 1.8 kb in size and agrees with the size of full-length ADA cDNA. A minor and variable RNA hybridizing species is also present at 5.8 kb and has been observed in mouse T cell, MOLT-4, HeLa and B lymphoblast mRNA by others.[4,17,18,24] The 5.8 kb hybridizing RNA species is too small to be a primary precursor RNA of the mature ADA mRNA since a minimum estimate for the human ADA gene is 18-23 kb.[16] It is possible that this RNA species could be a minor processed transcript of ADA; however the amount of this species is variable relative to the 1.6-1.8 kb ADA mRNA under stringent hybridization and washing conditions, suggesting minimal shared sequence homology between the two species.

As shown in FIGURE 4, mRNA from an ADA-deficient null cell line (DHL-9) shows no detectable 1.6-1.8 kb transcript but the presence of 5.8 kb RNA. The six patients with ADA deficiency and SCID all have detectable ADA mRNA of apparently normal size, suggesting the absence of major mRNA defects such as a large deletion
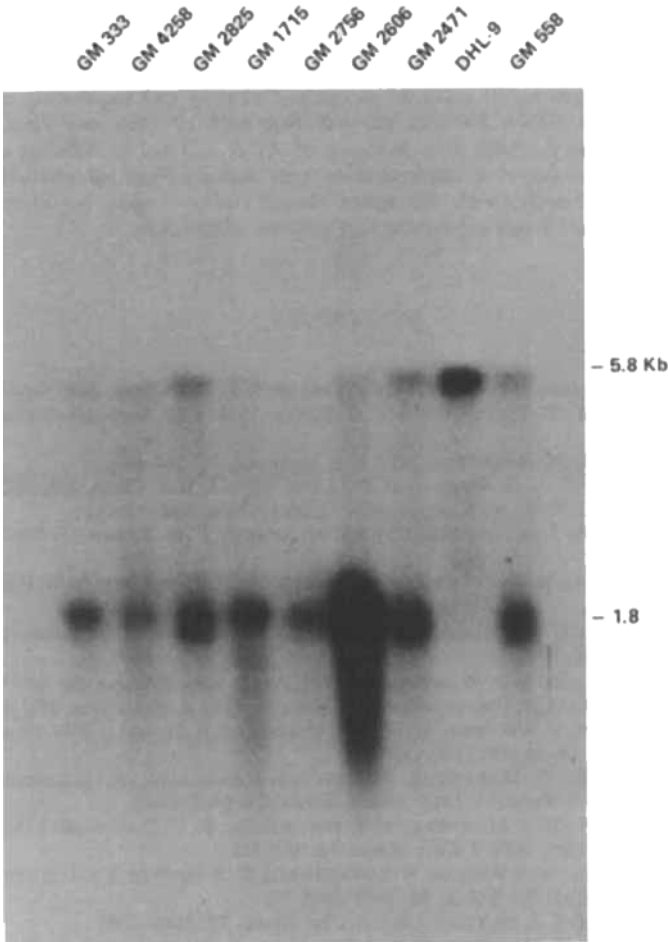
FIGURE 4. Northern blot analysis of mRNA from human lymphoblast cell lines. GM 333 and GM 558 represent mRNA from normal B lymphoblast cell lines. GM 4258, GM 2825, GM 1715, GM 2756, GM 2606, and GM 2471 are ADA-deficient B lymphoblasts derived from ADA-deficient SCID patients. DHL-9 is an ADA-deficient null lymphoblast cell line from a null cell leukemia patient and represents a negative control. (From Daddona *et al.*[20] Reprinted by permission of J. Biol. Chem.)

mutation or an mRNA processing defect. However, this analysis cannot detect small deletions or insertions of under 100 nucleotides or the presence of point mutations in the mRNA. Quantitation of hybridizable mRNA shows that all of these cell lines have a virtually normal ADA mRNA level. Cell line GM 2606 is an exception, having a fourfold elevation of ADA mRNA. A similar result to this has been reported by Hutton *et al.*[14] Southern blot analysis of GM 2606 DNA has ruled out ADA gene amplification as a mechanism for the elevated mRNA level.

Now that the ADA cDNA sequences are available, it will be possible to characterize fully both mRNA and DNA in ADA deficiency. Since all ADA-deficient cell lines derived from patients with immunodeficiency are ADA mRNA positive, this will allow mRNA analysis by S1 nuclease assays and cloning and sequencing of mutant ADA cDNA from cDNA libraries derived from each of these cell lines. Further knowledge of primary amino acid sequence of ADA will aid in defining structural mutations leading to functional defects in enzyme variants from subjects with partial ADA deficiency. Finally, with the entire cloned coding region for human ADA available, gene transfer and expression can now be undertaken.

REFERENCES

1. VAN DER WEYDEN, M. B. & W. N. KELLEY. 1976. J. Biol. Chem. **251:** 5448-5456.
2. HIRSCHHORN, R., F. MARTINIUK & F. S. ROSEN. 1978. Clin. Immunol. Immunopathol. **9:** 287-292.
3. ADAMS, A. & R. A. HARKNESS. 1976. Exp. Immunol. **26:** 647-649.
4. SCHRADER, W. P., A. R. STACY & B. POLLARA. 1976. J. Biol. Chem. **251:** 4026-4032.
5. DADDONA, P. E. & W. N. KELLEY. 1977. J. Biol. Chem. **252:** 110-115.
6. CREAGAN, R. P., J. A. TISCHFIELD, E. A. NICHOLS & F. H. RUDDLE. 1974. Lancet **2:** 1449.
7. TISCHFIELD, J. A., R. P. CREAGAN, E. A. NICHOLS & F. H. RUDDLE. 1974. Hum. Hered. **24:** 1-11.
8. GIBLETT, E. R., J. E. ANDERSON, F. COHEN, B. POLLARA & A. J. MEUWISSEN. 1972. Lancet **2:** 1067-1069.
9. MARTIN, D. W., JR. & E. W. GELFAND. 1981. Annu. Rev. Biochem. **50:** 845-847.
10. DADDONA, P. E., M. A. FROHMAN & W. N. KELLEY. 1980. J. Biol. Chem. **255:** 5681-5687.
11. HIRSCHHORN, R., V. ROEGNER, T. JENKINS, C. SEAMAN, S. PIOMELLI & W. BORKOWSKY. 1979. J. Clin. Invest. **64:** 1130-1139.
12. HIRSCHHORN, R., F. MARTINIUK, V. ROEGNER-MANISCALCO, A. ELLENBOGEN, J-L. PERIGNON & T. JENKINS. 1983. J. Clin. Invest. **71:** 1887-1892.
13. DADDONA, P. E., B. S. MITCHELL, H. J. MEUWISSEN, B. L. DAVIDSON, J. M. WILSON & C. A. KOLLER. 1983. J. Clin. Invest. **72:** 483-492.
14. WIGINTON, D. A., G. S. ADRIAN, R. L. FRIEDMAN, D. P. SUTTLE & J. J. HUTTON. 1983. Proc. Natl. Acad. Sci. U.S.A. **80:** 7481-7485.
15. ADRIAN, G. S. & J. J. HUTTON. 1983. J. Clin. Invest. **71:** 1649-1660.
16. ORKIN, S. H., P. E. DADDONA, D. S. SHEWACH, A. F. MARKHAM, G. A. BRUNS, S. C. GOFF & W. N. KELLEY. 1983. J. Biol. Chem. **258:** 12753-12756.
17. HUNT, S. W., III & P. A. HOFFEE. 1983. J. Biol. Chem. **258:** 13185-13192.
18. YEUNG, C-Y., E. G. FRAYNE, M. R. AL-UBAIDI, A. G. HOOK, D. E. INGOLIA, D. A. WRIGHT & R. E. KELLEMS. 1983. J. Biol. Chem. **258:** 15179-15185.
19. VALERIO, D., M. G. C. DUYVESTEYN, P. MEERA KHAN, A. GEURTS VAN KESSEL, A. DE WAARD & A. J. VAN DER EB. 1983. Gene **25:** 231-240.
20. DADDONA, P. E., D. S. SHEWACH, W. N. KELLEY, P. ARGOS, A. F. MARKHAM & S. H. ORKIN. 1984. J. Biol. Chem. **259:** 12101-12106.
21. WIGINTON, D. A., M. S. COLEMAN & J. J. HUTTON. 1981. Biochem. J. **195:** 389-397.
22. SCHROEDER, W. A. 1968. The Primary Structure of Proteins. Harper & Row. New York. pp. 78-79.
23. WIGINTON, D. A., G. S. ADRIAN & J. J. HUTTON. 1984. Nucleic Acids Res. **12:** 2439-2446.
24. VALERIO, D., M. G. C. DUYVESTEYN, H. VAN ORMONDT, P. MEERA KHAN & A. J. VAN DER EB. 1984. Nucleic Acids Res. **12:** 1015-1024.