Volume 11 Number 18 1983 Nucleic Acids Research

Sequence for human argininosuccinate synthetase cDNA

Hans-Georg O.Bock*, Tsung-Sheng Su, William E.O'Brien, and Arthur L.Beaudet

Department of Pediatrics and Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA

Received 10 June 1983; Revised and Accepted 31 August 1983

ABSTRACT

The nucleotide sequence for human argininosuccinate synthetase cDNA was determined by analysis of six clones isolated from a single experiment. The sequence covered 1623 nucleotides including 76 bases of poly(A) and contained a 1236 nucleotide open reading frame encoding a protein of 46,434 daltons. In one cDNA isolate, a cloning artifact or perhaps RNA polymerase error involving addition of an A in a region of six A's within the coding sequence was documented. Single base variations in the 3' untranslated region were examined in detail since detection of DNA polymorphisms in the cDNAs could imply over-expression of both alleles at the active locus in canavanine-resistant cells, i.e. a trans-acting mechanism for enzyme overproduction. However, the sequence from five cDNAs suggested some single base artifacts, and DNA polymorphism remains uncertain. The occurrence of three tandem arginine codons in the 5' untranslated region of the cDNA suggested the possibility of an interaction of arginyl-tRNA with mRNA to regulate RNA processing or half-life as a mechanism for arginine-mediated repression.

INTRODUCTION

Argininosuccinate synthetase functions in the urea cycle to convert citrulline and aspartate to argininosuccinate. A deficiency of this enzymic activity in humans is recognized as the disorder citrullinemia, which is inherited in an autosomal recessive manner (1). In some cultured cell lines, argininosuccinate synthetase is subject to metabolite regulation by arginine (2). Variant cultured human cell lines resistant to the arginine analog canavanine have an argininosuccinate synthetase specific activity which is 180-fold higher than that of parental cells. This canavanine-resistant phenotype is stable and the enzymic activity is not subject to an arginine-mediated regulation (3,4). The steady state level of this mRNA correlates closely with the specific activity of the enzyme in the cell extract, indicating regulation at a pretranslational level (4,5). There is no evidence for gene amplification in canavanine-resistant cells (5). There are multiple processed argininosuccinate synthetase pseudogenes (6). However, there is abundant

© IRL Press Limited, Oxford, England.

evidence that none of these pseudogenes is expressed, and all of the available data are consistent with the presence of a single active gene (6-8).

We report here the DNA sequence of several cDNA clones for human argininosuccinate synthetase obtained from a single cloning experiment. The sequence provides data relevant to possible mechanisms of canavanine resistance and arginine-mediated regulation.

METHODS

The isolation of the cDNA clones labeled pAS1, pAS2, and pAS3 from a recombinant cDNA library constructed from pBR322 and poly(A)$^+$ RNA from canavanine-resistant RPMI-2650 cells was described previously (5). The cDNA clones pAS4, pAS9 and pAS12 were isolated by rescreening that library with nick-translated pAS1. All isolates were from a single cloning experiment using mRNA from one cell harvest.

Sequence Analysis

Restriction maps of all isolates were prepared. Restriction fragments were isolated, labeled at 5′ ends, and sequenced by the method of Maxam and Gilbert (9).

RESULTS

Sequencing Strategy and Nucleic Acid Sequence

A total sequence of 1623 nucleotides, including 76 bases of poly(A), was obtained from human argininosuccinate synthetase cDNAs of clones pAS1, pAS2, pAS3, pAS4, pAS9, and pAS12 (Figures 1 and 2). Seventy-five bases 5′ to the first ATG sequence were determined. Beginning with the most 5′ ATG, there was an open reading frame of 1236 nucleotides, which encoded a protein with a molecular weight of 46,434 daltons. The amino acid content and molecular weight predicted for argininosuccinate synthetase by this nucleotide sequence agreed very well with values determined for the bovine liver enzyme (10).

The 3′ untranslated sequence contained 233 nucleotides between the TAG terminator codon and the poly(A) sequence. Potential poly(A) recognition sequences were present at nucleotides 1526-1532 (ATAAAAA) and nucleotides 1537-1544 (AATTAAAA). Each is a variation of the canonical eukaryotic sequence, AATAAA, which typically is located 11-30 nucleotides 5′ to the poly(A) addition site (11,12). The proximity of the poly(A) sequence to the AATTAAAA sequence suggested that bases 1526-1532 (ATAAAAA) may represent the actual poly(A) recognition sequence.

The codon usage was nonrandom. Whereas the coding region (bases 76-

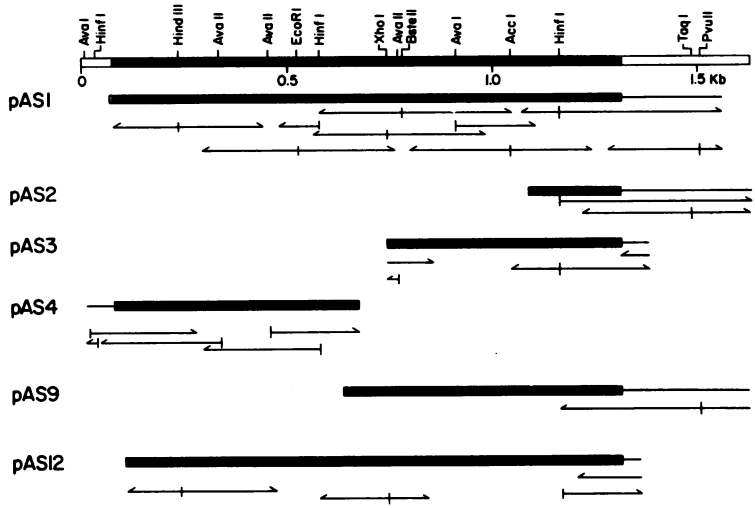

Figure 1. The sequencing strategy for human argininosuccinate synthetase cDNAs. The uppermost line provides a restriction map for the entire cDNA sequence. The sequence numbering proceeds in a 5′ to 3′ direction. Shown below and relative to this restriction map are the six cDNA inserts studied, each presented together with its own sequencing strategy. The solid, heavy lines represent protein coding regions, while the open, heavy lines represent untranslated regions. The cDNA fragments were labeled at the 5′ ends, and the direction and extent of each sequence is indicated directly below its respective cDNA map.

1311) had an overall G + C content of 56%, the G + C contents of positions 1, 2, and 3 of these codons were 57%, 37% and 75%, respectively. Relatively few codons contained the dinucleotide CG, similar to other eukaryotic DNA sequences.

Sequence Variations

Our efforts were placed originally on sequencing pAS1, the longest clone available. Upon completion of this sequencing, we were unable to identify a reading frame to encode a protein of the expected size. Analysis of additional cDNAs revealed that pAS1 contained a single base (A) insertion to give seven A's at a site where the sequences of both pAS3 and pAS12 contained only 6 tandem A's (Figures 2 and 3, nucleotides 757-762). Deletion of one of the seven A's at this position in pAS1 provided an open reading frame of the expected size. Additional evidence that only six A's occurred in this region was available from the sequence of the exon containing this region from a genomic clone and from the sequence of three processed pseudogenes (H.G.

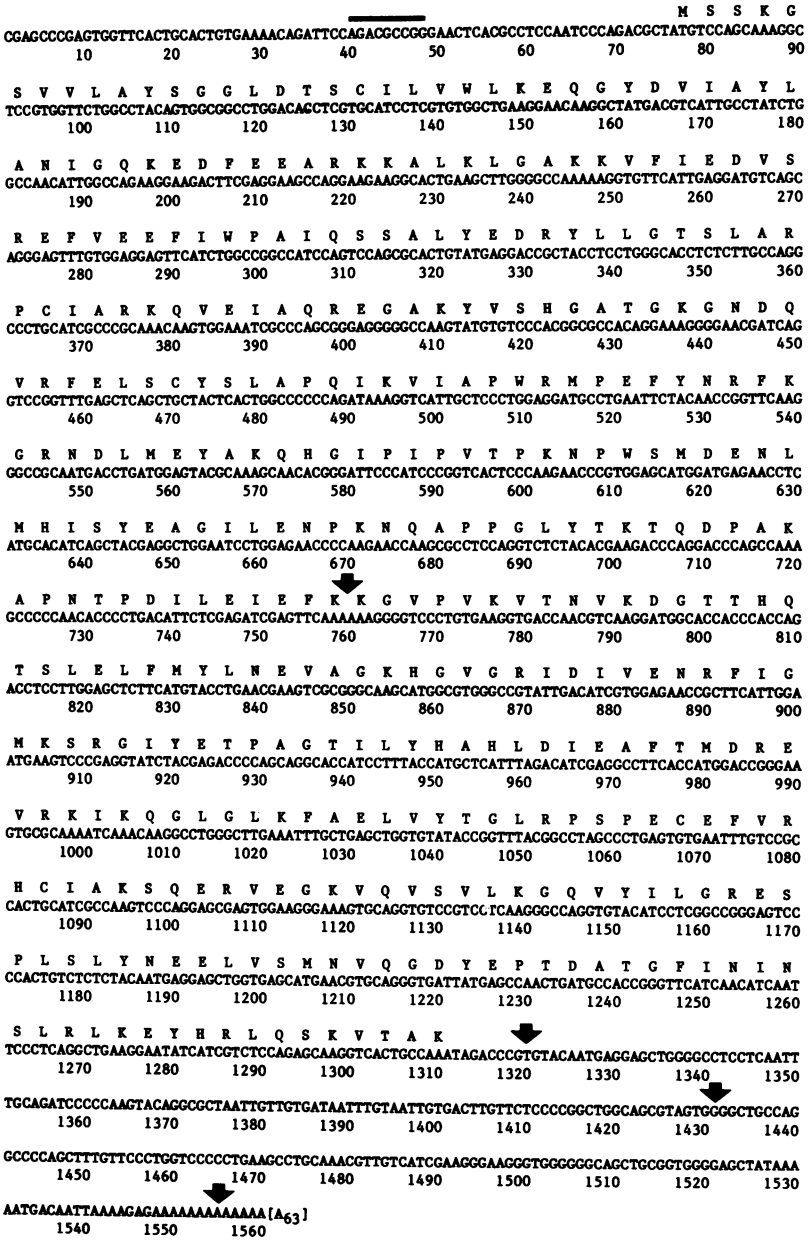

Figure 2. The nucleotide and amino acid sequences of human argininosuccinate synthetase cDNA. Three tandem arginine codons 5′ to the initiator ATG are overlined. Arrows indicate the locations of the observed nucleotide variations.

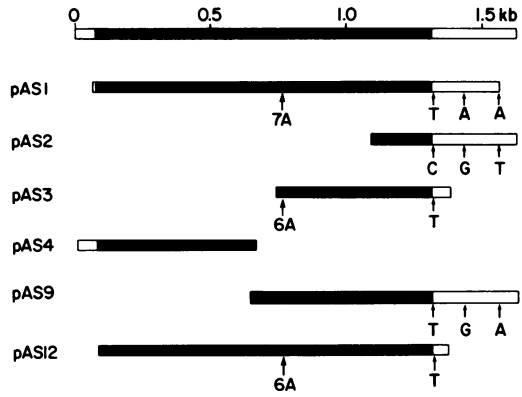

Figure 3. Summary of the Nucleotide Variations observed within Multiple Human Argininosuccinate Synthetase cDNA Clones (positions 757-762, 1320, 1431, and 1555).

Bock and S.O. Freytag, unpublished observations).

Nucleotide variations also were detected at positions 1320, 1431, and 1555 in the 3′ untranslated region (Figures 2 and 3). At position 1320, pAS2 had a C while pAS1, pAS3, pAS9, and pAS12 had T's. At position 1431, pAS1 had an A, while pAS2 and pAS9 had G's. At position 1555, pAS1 and pAS9 both had an A, while pAS2 had a T.

Arginine Codons in the 5′ Untranslated Region

Three tandem arginine codons occurred in the 5′ untranslated region. They were in the same reading frame as the coding sequence and were separated from the initiator ATG by 27 nucleotides. No ATG initiator codon was present upstream from the arginine codons although two GTG codons (an initiator codon in some instances) did occur upstream in this reading frame. No terminator codons were between the arginine codons and the initiator ATG. This sequence was from canavanine-resistant cells and could conceivably be different in wild type cells. The 5′ untranslated region was 59% G + C (44/75 bases), and this could favor the occurrence of the CGX series of arginine codons.

DISCUSSION

There are multiple argininosuccinate synthetase-like sequences in the human genome. Our approach to interpreting and integrating the data from six cDNA clones assumed the existence of a single active diploid locus for argininosuccinate synthetase, including canavanine-resistant cells. The

strongest argument for this interpretation was the analysis of cultured skin fibroblasts with mutations in the structural gene. These data indicated that all detectable mRNA was derived from a single locus in those cells (8). In addition, extensive analysis of genomic clones indicated a large expressed gene with at least eleven introns but no evidence for gene duplication (6).

The general features of the cDNA sequence were typical for a eukaryotic gene. The assigned coding region predicted an amino acid content in close agreement with that published for bovine liver enzyme (10). The content of basic amino acids (21 arginines and 33 lysines in 412 amino acids) was unusually high and was consistent with the pI of 9.0 reported for the human enzyme (13).

The occurrence of sequence variations both in the coding region and in the 3′ untranslated region led to the sequencing of additional cDNA clones from a single cloning experiment. Assuming that there is only one active diploid locus, we interpreted the extra A at position 757-762 in pAS1 to represent an RNA polymerase error or cloning artifact. The occurrence of six A's in two cDNAs, in three processed pseudogenes and in the genomic exon supported this interpretation. We could not eliminate the possibility that the tissue culture cell line was heterozygous for a frameshift mutation. Sequence variations in the 3′ untranslated region occurred at three sites, but the most 3′ variation could have involved slight differences in the site of polyadenylation and variations at bases 1320 and 1431 were less ambiguous. Initially this region was sequenced in multiple clones in an attempt to demonstrate genetic heterogeneity in the 3′ untranslated portion of the mRNA. Additional attempts to determine if DNA polymorphisms were present were unsuccessful. Although the sequence at position 1320 would involve the presence or absence of the sequence CGCG, a restriction enzyme recognition site for FnuDII or ThaI, efforts to analyze for such a variation were unsuccessful due in part to the presence of a very large number of pseudogenes which complicate Southern blotting analysis of genomic DNA (5). We have been unsuccessful, as have most other laboratories, in using S1 nuclease analysis to detect single base variations in the mRNA sequence (8). Assuming a single active locus, the isolation of two sequences in a small sample of clones would suggest that enzyme overproduction in canavanine-resistant cells involves increased expression of both alleles at the locus, i.e. a trans-acting mechanism. Three different combinations were identified at the 1320/1431 bases; T/A in pAS1, C/G in pAS2, and T/G in pAS9 (Figure 3). Again assuming a single active locus, these three combinations indicated at

least one artifactual result although polymorphism still could have occurred at either position. Single base artifacts at position 1320 in pAS2 and at position 1431 in pAS1 could explain the data. In concert with this possibility, the sequence in two processed pseudogenes agreed with that in pAS9 (S.O. Freytag and H.G. Bock, manuscript in preparation). Other approaches to distinguish whether the canavanine-resistant phenotype is trans-acting or cis-acting are being pursued.

The occurrence of three tandem arginine codons, nucleotides 40-48, in the 5′ untranslated region may be relevant to the repression of this enzymic activity by arginine. The steady state level of mRNA for argininosuccinate synthetase is regulated by the arginine concentration of tissue culture medium (2,5). The sequence determined may include virtually all of the 5′ untranslated region since two processed pseudogenes diverged from each other within four bases of the 5′ end of pAS4 (S.O. Freytag and H.G. Bock, manuscript in preparation). A limited survey of the 5′ untranslated region from over thirty other eukaryotic cDNAs indicated a need for cautious interpretation. Arginine codons occurred in clusters in a number of cDNAs, particularly if the G + C content was high, but no other instance of three tandem arginine codons was observed. The presence of tandem codons for a regulatory amino acid is similar to findings in bacterial amino acid operons which are regulated by attenuation (14-17). A similar mechanism may occur in yeast (18) and in SV40 (19). In our system, the absence of any evidence for translation of this region and the absence of a sequence resembling a termination site for transcription suggested that attenuation as described in bacteria would be unlikely. However, interaction of arginine codons with arginyl-tRNA could still serve as a component of a regulatory mechanism. A recent report (20) that nuclear RNA processing or stability was influenced by the occurrence of a nonsense codon in the β-globin gene, provides a precedent for interaction of mRNA with tRNA to affect RNA processing or stability in higher eukaryotic cells.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Research Fellowship Award GM07466-02 and grant GM 27593. We thank Drs. Svend O. Freytag and Jeffrey Rosen for constructive criticism of this manuscript and Lynn Loewenstein for help in the preparation of this manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Walser, M. (1983) in Metabolic Basis of Inherited Disease, Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L., and Brown, M.S. Eds., pp.402-438, McGraw-Hill, New York.
2. Schimke, R.T. (1964) J. Biol. Chem. 239, 136-145.
3. Jacoby, L.B. (1978) Somatic Cell Genetics 4, 221-231.
4. Su, T.-S., Beaudet, A.L. and O'Brien, W.E. (1981) Biochemistry 20, 2956-2960.
5. Su, T.-S., Bock, H.-G.O., O'Brien, W.E. and Beaudet, A.L. (1981) J. Biol. Chem. 256, 11826-11831.
6. Beaudet, A.L., Su, T.-S., Bock, H.-G.O., Freytag, S.O. and O'Brien, W.E. (1983 in Recombinant DNA Applications to Human Disease, Caskey, C.T., White, R., ed., pp. 97-103, Cold Spring Harbor Laboratory, New York.
7. Su, T.-S., Bock, H.-G.O., Beaudet, A.L. and O'Brien, W.E. (1982) J. Clin. Invest. 70, 1334-1339.
8. Su, T.-S., Beaudet, A.L. and O'Brien, W.E. (1983) Nature 301, 533-534.
9. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
10. Ratner, S. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 5197-5199.
11. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
12. Proudfoot, N.J. and Brownlee, G.G. (1974) Nature 252, 359-362.
13. Hudson, L.D., Erbe, R.W. and Jacoby, L.B. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 4234-4238.
14. Keller, E.B. and Calvo, J.M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 6186-6190.
15. Yanofsky, C. (1981) Nature 289, 751-758.
16. Platt, T. (1981) Cell 24, 10-23.
17. Kolter, R. and Yanofsky, C. (1982) Ann. Rev. Genet. 16, 113-134.
18. Andreadis, A., Hsu, Y.-P., Kohlhaw, G.B. and Schimmel, P. (1982) Cell 31, 319-325.
19. Hay, N., Skolnik-David, H. and Aloni, Y. (1982) Cell 29, 183-193.
20. Takeshita, K., Scarpa, A., Chan, L. and Benz, E.J.,Jr. (1983) Clin. Res. 31, 479A.