# Characterization of the promoter region of the gene for the rat neutral and basic amino acid transporter and chromosomal localization of the human gene

NING YAN<sup>\*†</sup>, RACHEL MOSCKOVITZ<sup>\*</sup>, LOUISE D. GERBER<sup>\*</sup>, SUSAN MATHEW<sup>‡</sup>, V. V. V. S. MURTY<sup>‡</sup>, SURESH S. TATE<sup>§</sup>, AND SIDNEY UDENFRIEND<sup>\*¶</sup>

\*Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110; <sup>‡</sup>Program of Cell Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY 10021; and <sup>§</sup>Department of Biochemistry, Cornell University Medical College, New York, NY 10021

Contributed by Sidney Udenfriend, April 29, 1994

ABSTRACT The promoter region of the rat kidney neutral and basic amino acid transporter (NBAT) gene has been isolated and sequenced. The major transcription initiation site was mapped by primer extension. The entire promoter region and a set of 5' deletions within it were expressed at a high level in LLC-PK1 cells using the luciferase indicator gene. Positive and negative regulatory elements in the promoter region were observed. A human genomic clone of the transporter was also obtained and was used to localize the NBAT gene at the p21 region of chromosome 2.

The cDNA for a rat kidney protein has been cloned that, when expressed in Xenopus oocytes, mediates sodiumindependent transport of neutral amino acids, cystine, and dibasic amino acids (1, 2). The protein appears to be a broad spectrum neutral and basic amino acid transporter (NBAT) resembling the  $b^{0,+}$  transporter (3). Additional cDNA clones, highly homologous to NBAT, were isolated from rabbit and human kidney (4-6). Previous studies on the distribution of mRNA had shown that NBAT is present in only a limited number of rat tissues (7). Immunocytochemical localization studies, using site-directed antisera raised against two different synthetic peptides representing epitopes in NBAT, showed this transporter to be localized to microvilli of epithelial cells lining renal proximal tubules and small intestinal epithelia of the rat (8, 9). Surprisingly, in the small intestine intense staining was seen within enteroendocrine cells and submucosal neurons (9). More recently, NBAT was shown to be present in rat brain and rat adrenal medulla. The material in the brain is localized in central autonomic nuclei (M. J. Nirenberg and V. M. Pickel, personal communication). The cellular localization of NBAT in renal and intestinal microvilli is in accord with its role as a transporter of amino acids across epithelial cells in these tissues. Its presence in both the central and peripheral nervous systems as well as in adrenal medulla suggests that these tissues have high requirements for amino acids transported by NBAT or that NBAT, or a closely related protein, is involved in neuronal functions other than transport. The limited distribution of NBAT mRNA in rat tissues suggests that the tissue-specific expression of NBAT may be regulated transcriptionally. To help answer some of these questions we have cloned, sequenced, and characterized the promoter region of the rat NBAT gene.  $\|$  We also determined the position of the NBAT gene in the human chromosomal map.

## MATERIALS AND METHODS

**Materials.**  $[\alpha^{-32}P]dCTP$  and  $[\gamma^{-32}P]ATP$  were from Amersham; cDNA probes were radiolabeled by using an oligola-

beling kit from Pharmacia; nitrocellulose filters were from Schleicher & Schuell; genomic DNA blots (Zoo-Blot and rat Geno-Blot) were obtained from Clontech. The gene light reporter vectors pGL2-Basic and pGL2-Control were from Promega; luciferase assay reagents and the Erase-a-Base system were from Promega; restriction enzymes were purchased from different commercial sources; superscript reverse transcriptase was from BRL; cell lines were from American Type Culture Collection; the CaHPO<sub>4</sub> transfection kit was from 5 Prime  $\rightarrow$  3 Prime, Inc.

Isolation of Rat Genomic Clones. Approximately  $2 \times 10^6$ bacteriophage plaques from an adult rat liver genomic library in EMBL3 SP6/T7 (Clontech RL1022j) were screened by hybridization with a <sup>32</sup>P-labeled NBAT full-length cDNA. Hybridizations were performed at 65°C for 24 h in  $6 \times$  SSC (15 mM sodium citrate, pH 7.0/0.15 M NaCl) containing 0.25% nonfat dry milk and the radiolabeled probe. Filters were washed with  $1 \times SSC/0.1\%$  SDS at 56°C for 1 h. Subsequent screenings of the genomic library with a 403-bp fragment that corresponds to the nucleotide sequence 1-403 of NBAT cDNA (5' probe) were performed under the same stringency. The genomic fragments from positive clones were released from EMBL3 by BamHI and subcloned into the Bluescript plasmid vector. A series of 5' unidirectional deletions of the genomic fragments using the Erase-a-Base System was made and these truncated constructs were sequenced by the dideoxynucleotide chain-termination method (10). Sequencing for each construct was performed on both strands. DNA sequence analysis was carried out on an Applied Biosystems DNA sequencing system (model 373A). Sequence comparisons were made using GenBank and the European Molecular Biology Laboratory nucleotide data bases.

**Primer Extension.** A 27-base oligonucleotide complementary to bases 22–48 in the 5' region of the rat NBAT cDNA was end-labeled with  $[\gamma^{-32}P]ATP$  and T4 polynucleotide kinase. Primer extension was performed as described (11). The labeled primer (10<sup>6</sup> cpm) was coprecipitated with either 50 µg of total RNA or 4 µg of mRNA at 30°C for 18 h, and the precipitate was resuspended in a buffer containing 40 mM Pipes (pH 6.4), 400 mM NaOAc, 1 mM EDTA, and 80% formamide. Extension was performed with 40 units of superscript reverse transcriptase for 90 min at 42°C in a solution containing 50 mM Tris·HCl (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, 1 mM each deoxynucleotide, and 1 unit of RNasin per ml. RNA templates were degraded with

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NBAT, neutral and basic amino acid transporter; CMV, cytomegalovirus;  $\beta$ -Gal,  $\beta$ -galactosidase; FISH, fluorescence *in situ* hybridization.

<sup>&</sup>lt;sup>†</sup>Present address: Department of Molecular Drug Mechanism, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ 08854.

To whom reprint requests should be addressed.

The sequence reported in this paper has been deposited in the GenBank data base (accession no. U10110).

1  $\mu$ g of RNase A. Extension products were extracted with phenol/chloroform, precipitated with ethanol, and then separated on an 8% acrylamide/urea gel. Sequencing of the NBAT genomic clone, utilizing the 27-base primer, was carried out for calibration.

Luciferase Plasmid Construction. The 3.6-kb BamHI/Msp I genomic fragment corresponding to nucleotides -3633 to +63 of the rat NBAT gene was subcloned in front of the luciferase gene into the Xho I site of the pGL2-Basic vector. A series of 5' unidirectional deletions of the constructed plasmid was created and sequenced as described above. As a positive control for transfection efficiency, we used the CMV- $\beta$ -Gal ( $\beta$ -galactosidase) vector that contains the cytomegalovirus promoter and the Escherichia coli lacZ indicator gene.

Transient Transfection. LLC-PK1 cells (derived from pig kidney), OK cells (derived from opposum kidney proximal tubules), and CHO cells were cultured according to the American Type Culture Collection recommendations. At 20 h before transfection, cells were seeded at a density of  $\approx 3 \times$ 10<sup>5</sup> cells per 3.5-mm well in a six-well plate. To each well 4  $\mu g$  of plasmid was added and transfection was carried out by the CaHPO<sub>4</sub> precipitation technique with a transfection kit (5 Prime  $\rightarrow$  3 Prime, Inc.). No glycerol shocking procedure was used. For cotransfection experiments, 0.1  $\mu$ g of the reporter CMV- $\beta$ -Gal was mixed with 4  $\mu$ g of each construct. All transfections were carried out in duplicate. At 48-72 h after transfection, cells were rinsed with PBS and then lysed with 100  $\mu$ l (per well) of lysis buffer (Promega kit for luciferase assay). For enzymatic reactions, 20  $\mu$ l of each lysate was mixed with 100  $\mu$ l of luciferase assay reagent (Promega) at 25°C. Luciferase activity was measured for 10 sec using a luminometer (Monolight 2010; Analytical Luminescence Laboratory, San Diego). Results are expressed as relative luciferase activity (light units) per mg of protein.

**Isolation of Human Genomic Clones.** A human genomic library in EMBL3 SP6/T7 (Clontech) was initially screened with the rat NBAT full-length cDNA (as described above). Positive clones, when rescreened with the rat NBAT 5' cDNA probe, yielded three positive clones. Genomic fragments were released from the vector by *Bam*HI, subcloned into Bluescript, and partially sequenced. One clone was obtained that contained a 2.2-kb fragment, which included the nucleotide sequence 150–476 of human NBAT cDNA (6). This fragment was used for chromosomal mapping.

Southern Analysis of a Somatic Cell Hybrid Panel. The 2.2-kb genomic insert of the human NBAT was labeled with  $[\alpha^{-32}P]dCTP$  and then used in Southern analysis following standard methods. The somatic cell hybrid DNA panels were derived from 35 hybrid cell lines consisting of various combinations of human chromosomes against a hamster background (Bios, New Haven, CT).

Fluorescence in Situ Hybridization (FISH). The human NBAT genomic insert was labeled with biotin-11-dUTP using the bio-nick labeling system (Life Technologies, Grand Island, NY) for chromosomal mapping. Metaphase chromosome preparations from phytohemagglutinin-stimulated and 5-bromodeoxyuridine-synchronized lymphocyte cultures, hybridization of biotin-labeled probe, and detection of signals by indirect immunofluorescence were performed as described (12, 13).

#### **RESULTS AND DISCUSSION**

Cloning of the Rat NBAT Promoter Region. NBAT cDNA was cloned from rat kidney (1, 2) and the corresponding mRNA was shown to be present in kidneys from rat, mouse, and rabbit as well as in rat intestine, brain, heart, and human colon carcinoma (CaCO-2) cells (7). Southern blot analysis of genomic DNAs from several eukaryotic species using rat

NBAT cDNA as the probe has now revealed that the NBAT gene is present in rat, mouse, rabbit, chicken, cow, dog, monkey, and human (data not shown). We have used Southern analysis of rat genomic DNA to show that the NBAT gene is >20 kb and that it is a single copy gene (data not shown).

To isolate genomic clones corresponding to the 5' end of the NBAT gene, we used full-length NBAT cDNA as a hybridization probe to screen a rat liver genomic library. Twelve positive clones were isolated and rescreened with the NBAT 5' cDNA probe. Two positive clones were isolated, sequenced, and found to be identical. One of them (T6-3), containing a 5.1-kb insert, was selected for further studies. Its restriction map is shown in Fig. 1A. Sequencing of T6-3 showed it to contain exon 1 and a portion of the first intron. In exon 1, the NBAT sequence +33 to +492 (Fig. 1B) corresponds to positions 1-459 of the previously published cDNA sequence (1). Exon 1 from position +72 (ATG) encodes the first 140 amino acid residues of NBAT and contains a typical transmembrane domain (1, 2). The first intron starts at position +493 with a GT consensus splicing donor sequence (14).

Sequence Analysis of the T6-3 Promoter Region. Fig. 1B shows the nucleotide sequence of the 5' flanking region of the rat NBAT gene. This sequence contains a typical TATA box (positions -31 to -26) that is known to modulate the efficiency of transcription by RNA polymerase II (15). However, there is no CCAAT box. The 3.6-kb region upstream of exon 1 also contains consensus sequences for the binding of many transcription factors. However, the functions of these putative regulatory elements in the expression of the NBAT gene remain to be determined. Fig. 1B also shows some DNA motifs that are repeated a few times in the NBAT promoter region. The TGGCA motif serves as a basal transcription factor for viral and cellular enhancer/promoter elements (16). The LBP-1 site (T/ACTGG) is known to bind the leaderbinding protein (17) and the CACCC site is present in the 5' flanking region of many other genes (18). None of the DNA motifs that were found in the NBAT promoter region are known to be kidney-specific regulatory elements (19, 20).

A search for sequence homology between the NBAT promoter region and other genes, using GenBank and European Molecular Biology Laboratory data bases, showed that the sequence -1471 to -1171 is highly homologous (80%) to type 2 *Alu*-like repetitive elements (21). These elements are present at  $\approx 10^5$  copies in the rodent genome (21) and have been claimed to be gene markers for brain, embryonic, and oncogenic proteins (21).

**Determination of Transcription Start Sites.** The transcription start site of the rat NBAT gene was determined by primer-extension analysis. A synthetic oligonucleotide, corresponding to the NBAT cDNA sequence 22–48, was used as primer. As shown in Fig. 2, primer extension against rat kidney RNA revealed one major extension product of 81 nt and smaller amounts of two larger products (asterisks). Traces of additional start positions, upstream and downstream of the major site, may represent premature terminations during the reverse transcriptase reaction or artifactual hybridization of the primer.

Functional Analysis of the NBAT Promoter. The promoter activity of the cloned 5' end of the NBAT gene was studied by transient expression assays with the luciferase gene as a reporter. Initially, the *BamHI/Msp* I fragment that contains 3570 nt of upstream sequences and 63 nt of exon 1 [T6-3/ (-3633)] was cloned in front of the luciferase gene in the plasmid pGL2-Basic. This construct and a series of deletion constructs were introduced by calcium phosphate-mediated transfection into three different cell lines, and the luciferase activity in the cell extracts was analyzed. Two different plasmids, pGL2-Control for luciferase activity and CMV- $\beta$ -Gal for transfection efficiency, were used as positive con-

### 7550 Biochemistry: Yan et al.



2	CCTGGAGCCC	GCGTGGTAGA	AGGAGAGAGG	TGATTTCCAC	GGGTGGACCT	CGGACCTCTA	AGTGCATTCT	GTGATGCACC	TGTACACAGG	TGTACCCAAC	-3471
	ACACCGTGCA	GTTTAAAAAC	TGCTTCACTC	AACTGAACAC	ACATTACTCC	CGGCATTTTG	GGAGGCAGAG	GCAGGCAGAT	CTCTGAGTTC	AAGGCCAGCT	-3371
	TGGCCTACAG	AGTGAGTTCC	AGAACAGGCA	GGGCTCCACA	CACAAAACCT	GTCTTGAAAC	ACACACACAT	TCTACTAGCT	GCTCTTGTAG	TGGTCCAGGG	-3271
	TGCTATTCCC	AGTACCTACA	TGGTGGCAAC	CATCTGTGAT	GCCAGCCCCA	GAGGATCTGA	TGTTCTGTGC	TTGCACTCCC	GTCTTCCATG	GTACCTGACA	-3171
	TGTACGTACA	TGGTGCACAG	ATAAGGGTGC	ATGCAAAGTG	CCCGTAAAAT	GGAAAATTAC	TAAAATGTGA	GTAGAGAGGA	CTGGGCTCTG	GTCTGGGTAC	-3071
	ACCTGGCACG	TTACTGACAG	CTCCCTGCAG	CATCTCAGAG	GTTGACCTAA	TGTGGCTTCC	AGCTCTTTCG	TTCTGAGTGA	AGTTCTTAAC	CACTTGTATA	-3071
	CCCTTTTCCC	TGTATTGCCT	CCGTATTCCC	TTTGTTCCTT	TGTCAACCTT	TCTCTCTCAC	TTTCTCATAC	ACATTTCTAC	ACTOTONACT	CTCACATCTC	-29/1
	ACACATATTA	Chickhocce	GCGIAIIGCC	CTCTCTCTCC	TTTTCCCTTA	TTTAACATTT	CONTRACTOR	CATTTANCC	AGICIGANGI	TOTOTOTO	-28/1
	AGAGAIAIIA	CAACAAACGI	BIICCICIAG	ACARCCORCE	CACTERES	A COLORADO	GCATATITAT	GATITIATEC	ACAIGGGIGC	CIGIGIICA	-2771
	GACACCICAG	AGAGCATCAG	AICCCGCIIC	AGAIGGGIGI	GAGICACCAC	AIGGIGACGG	GGGAAGAICC	ACIGCICICA	GCIGCGGAGC	CATCTCTGCA	-2671
	GACCTTCCTA	TCTAAGGTGG	ACCCCCCCCCCC	CCCATCCCCT	AGGCTCTAGG	ATTTATATT	GGGTTGGTAT	TTAGGAAAGC	TITTCTACTC	CAGGATTGTT	-2571
	TTTATGTATT	TTTTAAAAAT	CCTTAAATTT	TACCTTTTTT	TTACTTTTAA	AATCGGTGTA	CATAGAAATT	TTTTTTTTTT	AAGGAGTAAG	GTTAGCTTTA	-2471
	TGTTTTATGA	TCTTCCCAGA	TAGCCAGTGG	TCCAACACTA	CGTATTGTCT	CTGTATTAAG	TTCTCTTTTA	TGCTTTATAT	GCATTTATAT	ATGCACATAT	-2371
	ATGCATAGTC	ACATAGTCAT	GATCCCAAAT	ATTTCCACAT	TCCTTTGTCT	CAAGTCCAAA	CTCTCATTTT	GTGTAGTTTC	TGTTTCAATA	TAGGACTACC	-2271
	CTCAGTTACC	CCTACAAATG	TCAGAGCTCT	CGGCACTCAT	TGCCAAATTA	CTTGGTAGGT	TCCGTTTGTC	AAGTTACATT	GTACCCTTAT	CTGAAGAAAC	-2171
	ATATACATAA	AATTCACCAG	CATACACTTT	GATATTAAAG	GTAAGTGCTT	TGGTTGCTTG	TGGCAGACAT	ATGCCTCAAC	TTGTGTCTGC	TCTCAATGAT	-2071
	TGCCTTGGGA	TATAGAAAAT	TGAAAGGTGT	TAGTCTAGTG	GGTCAAGCAC	TGAGCTCATG	ATTTGTTTAA	TTGTTAAAAG	CACCCTCCAG	TGGGTTCAAT	-1971
	CCCCAGCTCC	GAAAAAAAGA	ACCAAAAAAA	алалалала	AAAGAAAAGA	AAAAAAGCAC	CCTCCAGGCC	AAAGCCCAGC	TGATTAGTCC	AACTCATACA	-1871
	TTCTGCATGC	TGAAGGGTTG	TGTTGTCTCA	GTCAAATTCG	GGGAGGATGG	ACATGGATGT	TTCTATGGTG	TATTTATTAT	GTGTCTGACA	GATAATGCAG	-1771
	TATCAAAACA	TGGAAGGTAA	CTCACTTCAT	TAGCTACCTC	TGAATTCAAG	GCTAGTGTTG	TCCATCCACT	CTAGAGACTT	CCAAGCCAGC	CAGGGATACA	-1671
	TCGTGAGACT	TGGTCAAAAC	AAACAATAGC	ACACAACTGC	TTCTTTACAA	AAGTTTTGGG	TATAATTAAT	AATGGTGAGA	AACAACATTG	TAGTTTATGC	-1571
	AAGTTTCTGA	GAAGGGATTC	ACATTTCTGT	GTTTTCATGA	CATTAAACCA	GAACAGGTCA	CTGGTTACAC	AAGATGCCCT	TCAGAAGCCC	CTGTGGTGTT	-1471
	GGGCACCCAG	AGAAGACAGT	GGAGTCTGAA	CGTCTCATGC	ACCCCGCACC	CCGCACCCCA	CCCAACACAA	GAATGATAAA	AACATTTTAA	TTAAGAAAAC	-1371
	ACCTCCTGGG	CTGGGGAGAT	GGCTCAGCGG	CTAAGAGCAC	TGACTGCTCT	TCCAGAGGTC	CTGAGTTCAA	ATCCCAGCAA	CCACATGGTG	GCTCACAACC	-1271
	ATCTGTAATG	GGATCTGATG	CCCTCTTCTG	GTGGGTCTGA	AGACAGCTAC	AGTGTACTTA	ТАТАТААТАА	ATAAATAAAT	CTTTAAAAAA	AAAAGAAAAC	-1171
	ACCTCCTGTC	TAATCAAAGC	CCTCAAGTGT	GACTGATCAT	CTTATTCTT	GAGGCCTGGG	GAATCTCTCA	GGCCAGCCTT	TCAACAGTTT	CTGGGGACTT	-11/1
	CONNECTOR	TATCARAGE	TCANTCATTA	ANCANCTACT	TACTATCACC	ATTAACCACT	CARTECCETC	ACTABABACAC	TTCAACAGIII	ANCOCTONCO	-1071
	GCAAAIGIGI	TCTACACITG	ICARIGATIA	AACAACIACI	TAGTATGAGG	ATTAAGGACT	GITIGCCIGI	AGIAAAACAC	TIGAACTIGC	AAGGCICAGC	-9/1
	ATCCCATTCC	CCCACCTTTT	CTGGAAGTGC	TTACAAGAAT	GITAAACCAC	ATTCTTCTCT	GAAATCGGGA	TCAAAGGCCT	GAGGIIGAAC	TGAGCACACA	-871
	CATGCCTTGC	TTAGTCATAC	ACAAGGGCTG	TCTGGAGTCG	GCCTGAGCTG	ACATGAAATG	AGCTTCAGCG	ACAAAATGCA	GTCCCCTGTC	TCCAGAGCTG	-771
	CTGCTGTGGA	TCTGTAGGGG	AACACCTGCA	AACAAAATGG	ACAGCCAGCT	TCTCCTTAGA	AAGTTTGAGA	TAATTTTTCT	ACCTAGCTAA	AGAAAGAAAA	-671
						LBP-I	TGGCA				
	TGTCAATGTT	TTTCAACCAT	TGTGAAACAA	CACTATTATA	CCCACTTAAT	ACTGGCTTG	G <b>TGGCA</b> GTGT	GTCTGCATTA	TTGTAGTACC	ACAATTGTCA	-571
	CGTCTCTAAA	ACTCATCAAA	ATCCCTAGCT	AGAATTCTAT	TCCATTGGCC	AACACTTTCC	CACAGTTCCT	CCTABACCCC	CAACCACTGT	TCTGTTCTGT	-471
	CGTCTCTAAA	ACTCATCAAA	ATCCCTAGCT	AGAATTCTAT	TCCATTGGCC	AACACTTTCC	CACAGTTCCT	CCTAAACCCC	CAACCACTGT	TCTGTTCTGT	-471
	CGTCTCTAAA	ACTCATCAAA	ATCCCTAGCT	AGAATTCTAT	TCCATTGGCC	AACACTTTCC	CACAGTTCCT	CCTAAACCCC	CAACCACTGT	TCTGTTCTGT	-471
	CGTCTCTAAA GCTTCTGTGA	астсатсааа асатадаттс	атссстадст сттстстаад	AGAATTCTAT TGGAACCACA	tccattggcc caatatttgt	AACACTTTCC GTGGCTGTGG	CACAGTTCCT TTATTTTCCA	CCTAAACCCC CTCAGCGTAA	CAACCACTGT TGCCCTTGAG	TCTGTTCTGT GTCCATCCAT	-471 -371
	CGTCTCTAAA GCTTCTGTGA	ACTCATCAAA ACATAGATTC	атссстадст сттстстаад	AGAATTCTAT TGGAACCACA	TCCATTGGCC CAATATTTGT	AACACTTTCC GTGGCTGTGG	CACAGTTCCT TTATTTTCCA	CCTAAACCCC CTCAGCGTAA	CAACCACTGT TGCCCTTGAG	TCTGTTCTGT GTCCATCCAT BP-1	-471 -371
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT	ACTCATCAAA ACATAGATTC GTGTCAGACT	атссстадст сттстстаад тддсттссст	AGAATTCTAT TGGAACCACA TTCAGGGAAT	tccattggcc caatatttgt aatttctcct	AACACTTTCC GTGGCTGTGG GTACGTCCAT	CACAGTTCCT TTATTTTCCA TTCACATTCA	CCTAAACCCC CTCAGCGTAA CATGATGATG	CAACCACTGT TGCCCTTGAG LI GGTGCATCTC	TCTGTTCTGT GTCCATCCAT BP-I TGGGTCACCT	-471 -371 -271
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT	ACTCATCAAA ACATAGATTC GTGTCAGACT	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT	AGAATTCTAT TGGAACCACA TTCAGGGAAT	TCCATTGGCC CAATATTTGT AATTTCTCCT	AACACTTTCC GTGGCTGTGG GTACGTCCAT	CACAGTTCCT TTATTTTCCA TTCACATTCA	CCTAAACCCC CTCAGCGTAA CATGATGATG	CAACCACTGT TGCCCTTGAG GGTGCATCTC	TCTGTTCTGT GTCCATCCAT BP-1 TGGCTCACCT	-471 -371 -271
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC	CAACCACTGT TGCCCTTGAG GGTGCATCTC CCGCCTCTTC	TCTGTTCTGT GTCCATCCAT BP-1 TGGCTCACCT TCCCCAGTCT	-471 -371 -271 -171
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG	атссстадст сттстстаад тддсттссст аатддаатта	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p	CCTANACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC rotein	CAACCACTGT TGCCCTTGAG GGTGCATCTC CCGCCTCTTC	TCTGTTCTGT GTCCATCCAT BP-I TGGGTCACCT TCCCCAGTCT	-471 -371 -271 -171
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC	АТСССТАВСТ СТТСТСТААБ ТБССТТСССТ ААТБСААТТА ТБТАБСТБАБ	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT	тссаттбосс саататттбт аатттстсст ттббаастса ббтбтааба	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCC	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TCCCCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATCTC CCGCCTCTTC CCCCTTTACT	TCTGTTCTGT GTCCATCCAT BP-1 TGGGTCACCT TCCCCAGTCT AAACTTTACC	-471 -371 -271 -171
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT	астсатсала асатадаттс дтдтсадаст тадатттдтд дтастдттас	АТСССТАСТ СТТСТСТААС ТССТТСССТ АЛТССАТТА ТСТАССТСАС	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC rotein TCCCCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCCTTTACT	TCTGTTCTGT GTCCATCCAT BP-1 TGGTTCACCT TCCCCAGTCT AAACTTTACC	-471 -371 -271 -171 -71
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT	астсатсала асатадаттс дтдтсадаст тадатттдтд дтастдттас	АТСССТАСТ СТТСТСТААС ТССТСТСТААС ТССТСССТ АЛТССААТТА ТСТСТСАССС	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p A <u>CACACCC</u> CC	CCTAAACCCC CTCAGCGTAA CATGATGATGATG GTTTGGAAGC rotein TCCCCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATCTC CCGCCTCTTC CCCCTTTACT	TCTGTTCTGT GTCCATCCAT BP-1 TGGTTCACCT TCCCCAGTCT AAACTTTACC	-471 -371 -271 -171 -71
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC CAAAGCTCAT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCCC-binding p ACACCCCC TAATAAACCA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC Totein TCCCCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATCTC CCGCCTCTTC CCACCCACCT	TCTGTTCTGT GTCCATCCAT BP-1 TGGCTCACCT TCCCCAGTCT ANACTTTACC TTTCACTCTG	-471 -371 -271 -171 -71 30
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC CAAAGCTCAT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTOIN TCCCCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG	-471 -371 -271 -171 -71 30
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT ANAGCACTGG	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC CAAAGCTCAT GACAAAGACA ACD UR APPL	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTC	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCAG CAGCACCCAG	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCCGGAC	-471 -371 -271 -171 -71 30 130
1	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA AspLysAspL	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTC ysArgAspSe	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LysG	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC LYCysArgTh	-471 -371 -271 -171 -71 30 130
1	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGGAG	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASPLYSASPL CCAGGGACAC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTC YSATGASPSe TCCACAGTCC	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC Odein TCCCCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT ACTATGAAGG SerMet Lysg GTATCCCTGC	TCTGTTCTGT GTCCATCCAT BP-1 TGGTTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC LYCYSArgTh TCCAGAGGAG	-471 -371 -271 -171 -71 30 130 230
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rASNASNGIY	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT ANAGCACTGG TTTGTCCANA PheValgina	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT snGluAspl1	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATT ACCGGATAGA CCAGGAGCAG eGInGluGin	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAGCCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGCACA AsplysAspl CCAGGGACAC erArgAspTh	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACCACCCC TAATAAAACCA AGAGAGACTC YSATGASPSe TCCACAGTCC rProGinser	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCAGATG rlleGlnMet AACGCTGTTA AsnAlaVals	CAACCACTGT TGCCCTTGAG GGTGCATCTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LysG GTATCCCTGC erIleProAl	TCTGTTCTGT GTCCATCCAT BP-1 TGGDTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC LYCysArgTh TCCACAGGGAG aProGluGlu	-471 -371 -271 -171 -71 30 130 230
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rASnASnG1y CCTCAACTAA	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValgina	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SING1UASP11 GCCCTATGCA	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG eGInGluGin	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnCCAGCT AspProAspS AGGAGGTGCT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLysASpL CCAGGGACAC erArgASpTh GTTCCAGTTC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTCC ysArgAspSe TCCACAGTCC TProGINSer	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTO TCCCCCTGAC 1 CAGCACCCAG TATCCAGATG r11eG1nMet AACGCTGTATA ASDA12V1S CTCGCTACCG	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCCTTTACT CAAGCCACCT ACTATGAAGG GTATCCCTGC GTATCCCTGC GTGCCCTCGG	TCTGTTCTGT GTCCATCCAT BP-I TGG TCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC I YCYSATGTh TCCAGAGGAG aProGluGlu	-471 -371 -271 -171 -71 30 130 230
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG CTCAACTAA	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGIAA AGGTGGGGGG ysvalValAr	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SPG1uASp11 GCCCTATGCA GPF0TYFA1a	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG GGATGCCCA GIMGEPFOL	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT YSCIuValLe	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC CACAAGTAGC CAAAGCTCAT GACAAAGACA ASpLySASpL CCAGGACAC CCAGGACAC CCAGGACAC GTTCCAGTTC UPheGIPPhe	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding r ACACACCCCC TAATAAACCA AGAGAGACTC ysArgAspSe TCCACAGTCC TProGInSer TCTGGCCAGG SerGlyGInA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCAG TATCCAGATG r1leGlnMet AASnAlaValS CTGGCTACCG laArgTyrAr	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LySG GTATCCCTGC GTATCCCTGC GYAIPTOATg QVAIPTOATg	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1YCYSArgTh TCCAGGGAG BProGluGlu GAGATCCTCT GUILIELEUP	-471 -371 -271 -171 -71 30 130 230 330
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG CTCAACTAA ProGInLeuL	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGGTGCG ysValValAr	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT snGluAsp11 GCCCTATGCA GProTyrAla	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG eGInGluGIn CGGGATGCCA	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG Met AsnGlu GACCCAGACT AspProAspS AGGAGGTGCT ysGluValle	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLySASpL CCAGGGACAC erArgAspTh GTCCCAGTCC UPheGInPhe	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTC ysArgAspSe TCCACAGTCC rProGinSer TCTGGCCAGG SerGiyGinA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTO TCCCCCTGAC CAGCACCCAG TATCCAGATG r1leG1nMet AACGCTGTTA AsnA1aVa1S CTCGCTACCG 1aArgTyrAr	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LysG GTATCCCTGC erIleProAl GGTGCCTCGG gValProArg	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC IYCYSATGTh TCCAGAGGAG aProGluGlu GAGATCCTCT GluILeLeuP	-471 -371 -271 -171 -71 30 130 230 330
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG CCTCACTAA ProGInLeuL TCTGGCTCAC	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG PheValGlnA AGGTGGTCCG ysValValAr CGTGGTCCCG	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT snGluAsp11 GCCCTATGCA GProTyrAla GTGTTCCTGC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGACCAC egingiugin GGGATGCCCA GlyMetProL TCATCGGAGC	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TAAACCATT CATGAATGAC MetAsnGlu GACCCAGACT ASPProASPS AGGAGGTGCT ysGluValLe CACCATAGCC	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA AspLysAspL CCAGGGACAC erArgAspTh GTTCCAGTTC uPheG1nPhe ATCATCATCA	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTC ysArgAspSe TCCACAGTCC rProGInSer TCTGGCCAGG SerGIyGInA TCTCTCCAAA LeSerProLy	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCAG TATCCAGATG r1leG1nMet AACGCTGTA ASDA1aVa1S CTCGCTGACCG laArgTyrAr ATGCCTGACC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LysG GTATCCCTGC erIleProAl GGTGCCTCGG QValProArg TGGTGCGCAGG	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1yCysArgTh TCCACAGGGAG aProGluGlu GAGATCCTCT GluIleLeuP CAGGTCCCAT	-471 -371 -271 -171 -71 30 130 230 330 430
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rASNASNCIY CCTCAACTAA ProGInLeuL TCTGGCTCAC	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCANA PheValGINA ACGTGGTGCC YsValValSer CGTGGTCTCC rValValSer	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT snGluAspl1 GCCCTATGCA gProTyrAla GTGTTCCTGC ValPheLeuL	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATT ACCGGATAGA CCAGGAGCAG eGInGluGin GGGATGCCCA GlyMetProL TCATCGGAGC euIIGIyAI	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TTAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGGCT ysGluValLe CACCATAGCC aThrIleAla	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA AspLysAspL CCAGGGACAC erArgAspTh GTTCCAGTC uPheGInPhe ATCATCATCA	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAAACCA AGAGAGACTC YSATGASPSe TCCACAGTCC rProGinser TCTGGCCAGG SerGiyGinA 1cSerProLy	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCTGAC TATCCAGATG r1leG1nMet AACGCTGTAA ASIN1aVa1S CTCGCTACCG laArgTyrAr ATGCCTGAC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LysG GTATCCCTGC erileProAi GGTGCCTCGG QValProArg TGGTGGCAGG TrpTrpGInA	TCTGTTCTGT GTCCATCCAT BP-1 TGGCTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC lyCysArgTh TCCAGAGGAG aProGluGlu GAGATCCTCT GluIleLeup CAGGTCCCAT laGlyProMe	-471 -371 -271 -171 -71 30 130 230 330 430
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG CCTCAACTAA ProGInLeuL TCTGGCTCAC heTrpLeuTh GTACCAGATL	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCANA PheValGlnA AGGTGGTCCC rValValAr CGTGGTCTCC rValValSer TACCCCAGGT	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SnGluASpl1 GCCCTATGCA GProTyrAla GTGTTCCTGC ValPheLeuL CTTTTAAGGA	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATT ACCGGATAGA CCAGGACCAG egingiugin GGGATGCCCA GlyMetProL TCATCGGACAG euileGlyAl	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT ysGluValLe CACCATAGCC aThrIleAla GATCGCATAGCC	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLysAspL CCAGGGACAC erArgAspTh GTTCCAGTTC UPheGInPhe ATCATCATCAT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTCC YSATGASPSe TCCACAGSTCC TProGInSer TCCTGCCCAGG SerG1yG1nA TCCTCTCCAAA leSerProLy AGGTGGGTGT	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTGGAAGC TCCCCCTGAC 1 CAGCACCCAG TATCCAGATG TIleGInMet AACGCTGTAA ASDAIaVaIS CTCGCTACCG laArgTyrAr ATGCCTTGAC SCYSLEUASP TCGGCGCTTA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG GTATCCCTGC eTILeProAI GGTGCCTCGG gValProArg TGGTGGCAGG TrpTrpGINA GGACCAGTGG	TCTGTTCTGT GTCCATCCAT BP-I TGG TCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC IyCySATGTh TCCACAGGAC aProGluGlu GAGATCCTCT GluIleLeup CAGGTCCCAT IaGlyProMe AGGGTGGTTT	-471 -371 -271 -171 -71 30 130 230 330 430 530
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rAsnAsnGly CCTCAACTAA ProGlnLeuL TCTGGCTCAC heTrpLeuTh GTACCAGATC tTyrGlnILe	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTCG TTTGTCCAAA PheValGINA AGGTGGTGCC ysValValAr CGTGGTCTCC rValValSer TACCCGAGGT TyrProArgS	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SNGLUASPI1 GCCCTATGCA GPTOTYTA1a GTGTTCCTGC ValPheLeuL CTTTTAAGGA erPheLysAs	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG eGInGluGin GGGATGCCCA GlyMetProL TCATCGGAGC euileGlyAi CAGTGACAAG pSerAspLys	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT ysGluValLe CACCATAGCC aThrIleAla GATGGGAATG AspGlyAsnG	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLySASpL CCAGGGACAC CCAGGGACAC UPheGINPhe ATCATCATCA Ileileilei GAGACCTGAA lyAspLeuLy	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTC ysArgAspSe TCCACAGTCC TProGInSer TCTGCCCAG SerGlyGInA TCTCTCCCAAA leSerProLy AGTGGGTGT	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TCCCCCTGAC CAGCACCCAG TATCCAGATG rIleGInMet AACGCTGTA ASINAIaVaIS CTCGCTACCC IaArgTyTAY ATGCCTTGAC SCYSLEUASP TCGGCGCTTA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCCTTTACT CAAGCCACCT AGTATGAAGG GTATCCCTGC GTATCCCTGC GTATCCCTGC GValProArg TGGTGCCACGG YalProArg GGACCAGTGG	TCTGTTCTGT GTCCATCCAT BP-I TGG TCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1YCYSATGTh TCCAGAGGAG GAPCGIUGIU GAGATCCTCT GlulleLeup CAGGTCCCAT laGlyProMe AGGGTGGTTT	-471 -371 -271 -171 -71 30 130 230 330 430 530
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG CTCAACTAA ProGInLeuL TCTGGCTCAC heTrpLeuTh CTACCAGATC tTyrGInILe GTGGGCCCAC	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT ANAGCACTGG TTTGTCCANA PheValGINA AGGTGGTGCG yValValAr CGTGGTCTCC rValValSer TACCCGAGGTT TyrProArgS	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SNGLUASPI1 GCCCTATGCA GPTCTYTA1a GTGTTCCTGC ValPheLeuL CTTTTAAGGA erPheLysAs CTGGAACCA	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG GGGATGCCA GGGATGCCA GINGLIGINAI CAGTGACAAG pSerAspLys CTTTGTGCCT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TAGAACCATT ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT YSGLUValLe CACCATAGCC aThrIleAla GATGGGAATG AspGIJASnG GGCTTGTGTG	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLySASpL CCAGGGACAC erArgAspTh GTTCCAGTTC UPheGInPhe ATCATCATCA IleIleIleI CAGACCTGAA lyAspLeuLy GTAAACACAT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAAACCA AGAGAGACTC YSATGASPSe TCCACAGTCC TCTGGCCAGG SerG1yG1nA TCTCTCCCAA AGSTGGGTCT TCCAGGATGG	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TCCCCCTGAC CAGCACCCAG CAGCACCCAG TATCCAGATG r1leGInMet AACGCTGTA ASNA1aVa1S CTCGCTACCG IAArgTyrAr ATCCCTGAC SCYSLEUASP TCGGCGCTTA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LySG GTACTCCTGC erIleProAl GGTGCCTCGG gValProArg TGGTGGCAGG TrpTrpGINA GGACCAGTGG	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1yCysArgTh TCCACAGGAG aProGluGlu GAGATCCTCT IaGlyProMe AGGGTGGTTT AGTTCATTCA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG CCCACCACTCA TCTGGCTCAC TCTGGCTCAC TTTPLeuTh GTACCAGATC TTTGGCCCCAC	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGGTCCC CGTGGTCTCC CGTGGTCCCAGGTT TYPPTOArgS TCCCCAGGTTC AGCAGAAGCC	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC GAAGAGCTGC GAAGAGCCT GCCTATGCA GTGTTCCTGC GTGTTCCTGC CTTTTAAGGA erPheLysAs CTGGAAACCA TGCAAGCCT	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG eGInGIUGIn GGGATGCCCA GIYMETPTOL TCATCGAACA pserAspLys CTTTGTGCCT GCTCCTGATG	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TAGAATGAG ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT CACCATAGCC CACCATAGCC CACCATAGCC CACCATAGCC CACGATCAGCATC AspClyAsnG GCCTTGTGTG CAAGGCATCT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACCA ASpLySASPL CCAGGGACAC erArgASpTh GTTCCAGTTC UPheGInPhe ATCATCATCA IleIleIleI GAGACTGAAGA IyASpLeuLy GTAAACACAT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAAACCA AGAGAGACTC YSATJASPSe TCCACAGTCC rProGInSer TCTGCCAGG SerGIYGINA TCTCTCCAAA AGTGGGTGT TCCAGGATGG TTCCAGGATGG	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCAG CAGCACCCAG TATCCAGATG TILeGInMet AACGCTGTA ASNAIaVaIS CTCGCTGCCTGAC I aArgTyrAr ATGCCTGACC SCYSLEUASP TCGGCGCTTA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LysG GTATCCCTGC erIleProAl GGTGCCCTCGG gValProArg TGGTGGCAGGG TrpTrpGInA GGACCAGTGG	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1YCYAATGTh TCCAGAGGAG aProGluGlu GAGATCCTCT GluIleLeuP CAGGTCCCATT AGGTCGTTT AGGTCATTCA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG CCTCAACTAA ProGInLeuL TCTGGCTCAC heTrpLeuTh GTACCAGATC CCACGACCACTC CCACGACCACTC	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGGTCCC rValValSer TACCCAGGTC TyrProArgS TCCCAGCTC TCCCAGCTC TCCCAGCTC	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SnGluASpl1 GCCCTATGCA GProTyrAla GTGTTCCTGC ValPheLeuL CTTTAAGGA erPheLySAS CTGGAAACCA TGCAAGCCTT CCATCAATCC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATT ACCGGATAGA CCAGGACAG eGInGIuGIn GGGATGCCCA GIyMetProL TCATCGGACAG ceuileGiyAi CAGTGACAAG CTTTGTGCCT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT ysGluValLe CACCATAGCC aThrIleAla GATGGGAATG AspGlyAsnG GGCTTGTGGG CAAGCCTCT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLysAspL CCAGGGACAC erArgAspTh GTTCCAGTTC UPheGInPhe ATCATCATCAT IleIleIleI GAGACCTGAA IyAspLeuLy GTAAACACAT TGCTGAAGAC	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTCC YSATGASPSe TCCACAGTCC TProGInSer TCCTCCCAAA leSerG1yG1nA TCCTCCCCAAA leSerFroly ACCTCGCGTGT TCCAGGATGG TTCCAGGATGG	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG ASDA1401S CTCGCTGAC SCYSLEUASP TCGGCGCTTA CTGGCACCCG AGGGTTGTT GCTCTCTGGG	CAACCACTGT TGCCCTTGAG GGTGCATQTC CGGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG GTATCCTGC GVAIProArg TGGTGGCAGG TTCCCCTGGG TAATCTCTTG GACAGGCACA	TCTGTTCTGT GTCCATCCAT BP-I TGG TCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC IYCYSATGTh TCCACAGGAC aProGluGlu GAGATCCTCT GluIleLeup CAGGTGCCAT IAGIYPTOME AGGGTGGTTTT AGTTCATTCA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 890
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG CTACAACTAA ProGInLeuL TCTGGCTCAC heTrpLeuTh GTACCAGATC tTyrGInIle GTGGCCCAC	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGIAA AGGTGGTGCG ys'alValAr CGTGGTCTCC rValValSer TACCCGAGGTT TyrProArgS TCCCAGCTTC AGCAGAAGCC ATCATTCAC	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT GCCCTATGCA GPTOTYTAIa GTGTTCCTGC ValPheLeuL CTTTTAAGGA erPheLysAs CTGGAACCA TGCAAGCCA TGCAAGCCA	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG GGATGCCCA GIMGLYAI CAGTGACAAG pSerAspLys CTTTGTGCCT GCTCCTGATG GCTCCTGATG	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TAGAACCATT ATAAACCATT CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT YSGluValLe CACCATAGCC aThrIleAla GATGGGAATG AspClyAsnG GCATGTGTGGCA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLySASpL CCAGGACAC GTTCCAGTTC UPheGInPhe ATCATCATCA IleileIleI GAGACCTGAA IyAspLeuLy GTAAACACAT TGCTGAAGGA	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAAACCA AGAGAGACTC ySATGASpSe TCCAGGTCC TCTGGCCAGG CCCCCCAA CCCCCCCCC ProGINSer TCCGGCCAGG TCCGGCAGG TCCGGGGTCT SCCGCGGTCT TCCAGGACCG TCCGGAACCA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC CAGCACCCAG TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG CTGCCTGAC SCYSLEUASP TCGGCGCTTA CTGGCACCCG AGGGTGTT TCGGCCCCG	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG SerMet LySG GTATCCCTGC erIleProAr GGTGCCTCGG GYAIProArg TGGTGGCAGG TTCCCTGGG TATCCTTG GGACCAGTGG TAACCTCTG GACAGGCATAAGA	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1YCYSArgTh TCCAGGGAG aProGluGlu GAGATCCTCT GIUI1eLeuP CAGGTCCCAT IAGUYPIOME AGGGTGGTTT AGTCCATCAA CCCACTAAAC GTGTGTTCTT GAAGGCTAAA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 830 930
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCAT CTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG GAATAACGGG CACACACTCA TCTGGCTCACAG GGCTCACCACAC GGCTCACCACAC	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGCTCC CGTGGTCTCC TValValSer TACCCAGGTTC AGCAGCAGCC ATCCATCAC ATCATCACC	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC GAAGAGCTGC GAAGAGCACA GTGTTCCTGC CTTTTAAGGA erPheLysAs CTGGAAACCA TGCAAGCCTT CCATCAATCC AATACATCGT	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG eGInGIUGIn GGGATGCCCA GIYMETProL TCATCGGAGC eulieGIyAl CAGTGACAAG pserAsplys CTTTGTGCCT GCTCCTGATG TTAAATCTAA CTCCACAGGG	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TAGATCAATA CATGAATGAG MetAsnGlu GACCCAGACT AspProAspS AGGAGGTGCT CACGATAGCC CACCATAGCC CACCATAGCC CACGCATAGCC GGCTTGTGTG CAAGGCTCTT TTTTGTGGCACA GCCCCATACCC	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACCA ASpLySASPL CCAGGGACAC erArgASpTh GTTCCAGTTC UPheGInPhe ATCATCATCAT Ile1le1le1 CAGACCTGAAGA CCGTGTTCA CGTGAAGAA CCGTGTTCA	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCC TAATAAAACCA AGAGAGACTC ysArgAspSe TCCACAGTCC rProGInSer TCCAGCAGG SerGIyGInA TCTCCCAAA tCTCCCAAA TCCAGGATGG TTCCAGGATGG TCCCAGGATGG TCCCAGGATGG TCCCAGGATGG TCCCAGCACAA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TTTGGAAGC TCCCCCTGAC CAGCACCCAG TATCCAGATG TILeGInMet AACCCTGTA ASNAIaVaIS CTCGCTGAC SCYSLEUASP TCGGCGCTTA CTGGCACCCG AGGGTTGTT GCTCTTGGC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CCAAGCCACCT AGTATGAAGG SerMet LysG GTATCCCTGC erIleProAl GGTCCCTCGG gValProArg GGACCAGTGG TTCTTGGTGGCAGG TTCCCTGGG GACCAGTGG TTCCCCTGGG TATCTCTG GGACCAGTGG	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1YCYAATGTh TCCAGAGGAG aProGluGlu GAGATCCTCT GluIleLeuP CAGGTGCGTTT AGGTCGTCCTT AGGTCGTTCT AGGTCGTTCATTCA CCACCTAAAC GTGTGTTCTTA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 830 930
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rASnASnGly CCTCAACTAA ProGlnLeuL TCTGGCTCAC AGCACCACTC CCACGACAT AGCGGCCACA AGTTCTGTAT	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGGTCTCC rValValSer TACCCAGCTTC AGCAGACCC TTTATTAATA GTCATCACTTCACT	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SnGluASpl1 GCCCTATGCA GProTyrAla GTGTTCCTGC ValPheLeuL CTTTAAGA CTGGAAACCA TGCAAGCCTT CCATCAATCC ATAAAGGCTT	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGACCAG egingiugin GGGATGCCCA GlyMetProL TCATCGGACGAG CUILEGIYAI CAGTGACAAG CTTTGTGCCT GCTCCTGATG TTAAAATCTGA GTTATACAAA	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAACCATT CATGAATGAG MetAsnGlu CACCCAGACT AspProAspS AGGAGGTGCT ysGluValLe CACCATAGCC aThrIleAla GATCGGAATG GGCTTGTGGG GGCTTGTGGGG GCACCCATAT	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLysAspL CCAGGGACAC erArgAspTh GTTCCAGTTC UPheGInPhe ATCATCATCAT IleIleIleI GAGAACTGAA IyAspLeuLy GTAAACACAT TGCTGAAGAA AGGGATGACT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAACCA AGAGAGACTCC TYATAAACCA AGAGAGACTCC TCCAGCAGTCC TCCAGCAGTCC TCCAGCAGTCG TCCCAGGATGG TTCCAGGATGG GTCTTCAAGT TCCAGGACGCG GTCTTCAAGATGCG ATGGAACCAA	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTTGGAAGC TOTTGGAAGC TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG CTGGCCCGTA ACGCTGAC AGGGGTTGTT GCTCTCTGGG AGGGGTTGTT GCCCTGGACGA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CGGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG GTATCCCTGC GVAIProArg TGGTGGCAGG TTCCCCTGGG TAATCTCTG GACCAGGCACA AGAGTAAGA AGACATAGG	TCTGTTCTGT GTCCATCCAT BP-I TGG TCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC IyCySArgTh TCCAGAGGAC aProGluGlu GAGATCCTCT GluIleLeuP CAGGTGGTTTT AGGGTGGTTTT GAGGCTGATCT GTGTGTTCATTCA GTCGTCATCA GATGCCTACA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 830 930 1030 1130
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rASnASnGly CCTCAACTAA ProGInLeuL TCTGGCTCAC heTrpLeuTh GTACCAGATC tTyrGInIle GTGGCCCAC AGCACCACTC CCCCCGCCAT TAGCTGAGAG GGCTCACCACA AGTTCTGTAT ATTGGCTCAG	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGINA ACGTGGTGCGC ysValValSer TACCCGAGTTC TYProArgS TCCCAGCTTC AGCAGAACCC ATCATTCAT TAGTGCCAA	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT STGUASPI1 GCCCTATGCA GPF0TYFA1a GTGTTCCTGC VaIPheLeuL CTTTTAAGGA erPheLysAs CTGGAACCA ATGCAAGCCT TCCATCAATCC AATACATCGT ATTACATCGT ATTACTACATC ATTACTACATC	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG GINGLUGIN GGGATGCCA GIMELPIOL TCATCGGAGC eulleGlyAI CAGTGACAAG STATACAAAG TTGACTATA ACAAGAAAGT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA ATAAQCCATT CATGAATGAG MetAsnGlu GACCCAGACT ysGluValLe CACCATAGCC aThrIleAla GACGCGATGTGTG CAAGGCTGTT TTTTGGGCT TTTTGGGCT TTTTGGGCATAT ATATATGTC	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLySASpL CCAGGACAC CTCCAGTC UPheGInPhe ATCATCATCA Ileileilei GAGACCTGAA IyAspLeuLy GTAAACACAT TGCTGAAGGA AGGCATGACT	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCC TAATAAAACCA AGAGAGACTC YSATGASPSe TCCACAGTCC TCTGGCCAGG SerG1YGIAA TCTCTCCCAAA LeSerProly AGTTGGGTGT TCCAGGATCG TTCAGGAACCA ATTAAGATG	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTTGGAAGC TCCCCCTGAC CAGCACCCAG TATCCAGATG TATCCAGATG TATCCAGATG TATCCAGATG ASINAIAVAIS CTCGCTACG IAACGTGCT AGCGCTGAGA AGCGCTGTGG ACGTTGGACCG ACGTTGGACCG CCCCAGACAC	CAACCACTGT TGCCCTTGAG GGTGCATQTC CGGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG GTATCCCTGC GTATCCCTGC GValProArg TGGTGGCAGG GGACCAGTGG GAACGACTAGA TTTCTTTG GACAGGCACACG GGCTCAGCAG	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC 1YCYSATGTh TCCAGAGGAG aProGluGlu GAGATCCTCT GluIleLuP CAGGTCCCAT IAGUYPIOME AGGGTGGTTT AAGTCCATCAA TTAAGCTAAA TTAAGCTAAA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 830 930 1030 1130 1230
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTCTTCCT TCCCTACTGG GAATAACGGG GAATAACGGG GAATAACGGG CTCAACTAA ProGlnLeuL TCTGGCTCAC heTrpLeuTh GTACCAGATC TYYGInIle GTGGGCCCAC AGCACACTA AGTCTGAAA GGCTCACCAT AGCTGAGAG	ACTCATCANA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC GTACTGTTAC TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGGTGCG ysValValAr CGTGGTCTCC TValValSer TACCCAGGTTC AGCAGAACCC ATTCATCAC TTTATTATA	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC GAAGAGCTGC ATGAAGACAT SNG1uASp11 GCCCTATGCA GProTyrA1a GTGTTCCTGC Va1PheLeuL CTTTTAAGGA erPheLysAS CTGGAAACCA TGGAAACCA TGCAAGCCT ATTAAAGGCT ATTAAAGGCT	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATTE ACCGGATAGA CCAGGAGCAG eGInGIUGIN GGGATGCCA CIYMetProL TCATCGGAGC eUIIeGIYAI CAGTGACAAG pSerAspLys CTTTGTGCCT GCTCCTGATG GTTATACAAA CTCACACAGG ACAGAAAGT	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA TAGACCATT CATGAATGAG MELASDGL GACCCAGACT ASPProASPS AGGACGTGCT CACGCATAGCC CACCATAGCC CACCATAGCC CACCATAGCC CACGCATAGCC CACGCCATAGCC CACGCCATAGCC CACGCCATAGCC CACGCCATAGCC CACGCCATAGCC CACGCCATAGCC CACGCCATAGCC TGTCTGCGCA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTAGC TCATTTGTGC CAAAGCTCAT GACAAAGCTCAT GACAAAGCCCAT CCAGGGACAC erArgAspTh GTTCCAGTTC UPheGInPhe ATCATCATCAT IJelleileI GAGACCTGAA UPheGINPHEJULY GTAAACACAT TGGTGAAGGA CGGTGTTCA AGGATGACT TGATAGAAAA	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCCC TAATAAAACCA AGAGAGACTC YSATGASDS TCCACACGTCC TProGINSer TCCGGCAGG SerGIYGINA TCTCTCCAAA 1eSerProLy TCCAGGATGG TTCCAGGATGG TTCCAGGATGG TTCTCAGCTG GTCTTCAAGTTG GTTTCAAGTTG	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC TOTTGGAAGC TCCCCCTGAC CAGCACCCAG CAGCACCCAG CAGCACCCAG CAGCACCCAG CAGCACCCAG CAGCACCCG IAArgTyrAr ATGCCTGAC CTGGCACCCG AGGGCTGAT CTGGCACCCG TCAATCCTG ACCTTTGAC CTCGCGCTTA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CCGCCTCTTC CCCCTTTACT CCAGCCACCT AGTATGAAGC SerMet LySG GTATCCCTGC erIleProAl GGTGCCTCGG gValProArg TGCTGGCAGG TTCCCCTGGG TATCTCTTG GGACCAGTGG GACAGGCACA AGACATAGA	TCTGTTCTGT GTCCATCCAT BP-I TGGTCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC IYCYSATgTh TCCAGAGGAG aProGluGlu GAGATCCTCT GluIleLeup AGGGTGGTTT AGGTCATTCA CCACTAAAC CCACTAAAC CGAGTCCCAT GAGGCTGATT GAAGCTAAA TTAAGCTAAA TTAAGCTAAA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 830 930 1030 1130 1230 1230
	CGTCTCTAAA GCTTCTGTGA CCTGTAGTGT CCACCTTTGG CCTCCCCCAT CTTTCTTCCT TCCCTACTGG GAATAACGGG rASNASNG1y CCTCAACTAA ProG1nLeuL TCTGGCTCAC AGCACCACTC CCACGACAT GTGGGCCCAC AGCACCACTC CCACGACAT AGCTCCACCAG AGCTCTGTAT ATTGGCTCAT	ACTCATCAAA ACATAGATTC GTGTCAGACT TAGATTTGTG GTACTGTTAC TGGCTGGACT TGGCTGGACT AAAGCACTGG TTTGTCCAAA PheValGlnA AGGTGGTCTCC rValValSer TACCCGAGCTTC AGCAGAACC ATTCATTCAC TTATTAATA TTGTTGCCTA GTGATCCCAT GTGATCCCAT GTCATTACT	ATCCCTAGCT CTTCTCTAAG TGGCTTCCCT AATGGAATTA TGTAGCTGAG TTGTCCAGCC GAAGAGCTGC ATGAAGACAT SnGluASpI1 GCCCTATGCA GProTyrAla GTGTTCCTGC ValPheLeuL CTTTAAGA CTGGAAACCA TGCAAGCCTT CCATCAATCCC AATACATCGT ATAAAGGTT ATAAAGGTT ATAAAGGTT	AGAATTCTAT TGGAACCACA TTCAGGGAAT CTGTAGGCAG TCAACCACCT AGGATTATT ACCGGATAGA CCAGGACCAC eGInGIuGIn GGGATGCCCA GIYMetProL TCATCGGACAG cuileGIYAI CAGTGACAAG CTTTGTGCCT GCTCCTGATG TTAAAATCTGA GTTATACAAA ACTACAGAACTC	TCCATTGGCC CAATATTTGT AATTTCTCCT TTGGAACTCA GGTGTAAAGA CATGAACGA ATAAACCATT CATGAATGAG MetAsnGlu CACCCAGACT AspProAspS AGGAGGTGCT ysGluValLe CACCATAGCC aThr IleAla GATCGGAATG GGCTTGTGTG GGCTGTGTG GGCAGCCATAT ATATAGTTC GGCGAGCGCGT TTTCTTCACA	AACACTTTCC GTGGCTGTGG GTACGTCCAT CACAAGTCCAT TCATTTGTGC CAAAGCTCAT GACAAAGACA ASpLysAspL CCAGGGACAC erArgAspTh GTTCCAGTTC UPheGInPhe ATCATCATCATCA IleIleIleI GTAACACAT TGCTGAAGA AGGGATGATT AGGGATGATTA AGGGATGACTA	CACAGTTCCT TTATTTTCCA TTCACATTCA ATCTTACGGT CACCC-binding p ACACACCCC TAATAAACCA AGAGAGACTCC TProGInSer TCCAGCAGTCC TProGInSer TCCAGCAGTCC TCCAGGATGG TTCCAGCCCC	CCTAAACCCC CTCAGCGTAA CATGATGATG GTTTGGAAGC rotein TCCCCCCTGAC 1 CAGCACCCAG TATCCAGATG TATCCAGATG TIleGInMet AACGCTGAC AACGTTGAC CTGGCACCCG AGGGGTTGTT GCTCTCTGGC ACGTTGGAT AACCTGGCACCCG AGGGGTTGTT GCCCTGGAGA GCGCGAGACAG AAAACCAAGA	CAACCACTGT TGCCCTTGAG GGTGCATQTC CGGCCTCTTC CCCCTTTACT CAAGCCACCT AGTATGAAGG GTATCCCTGC gValProArg TGGTGCCTCGG gValProArg TGGTGCCAGGG TTCCCCTGGG TAATCCTTG GAACAGCACG AGAGATAAGA AGACAACAG TGGTGGGTAGGA AGACACACAG TGGTGGTGGTA	TCTGTTCTGT GTCCATCCAT BP-I TGG TCACCT TCCCCAGTCT AAACTTTACC TTTCACTCTG GATGCCGGAC JYCySArgTh TCCACAGGAG aProGluGlu GAGATCCTCT GluIleLeuP CAGGTCGATT AGTCATTCA CCCACTAAACTAG GATGCCTAAA TGTAAACTAG GATGCCTAAA	-471 -371 -271 -171 -71 30 130 230 330 430 530 630 730 830 930 1030 1130 1130 1230

FIG. 1. Organization and nucleotide sequence of the 5' end of the rat NBAT gene. (A) Restriction map of the 5.13-kb BamHI fragment (clone T6-3) containing exon 1 and part of intron 1. Sites for restriction enzymes starting from the 3' end are Nde I (N), Bsu36I (Bu), Apa I (A), Hae II (H), Msp I (M), BstEII (Bt), Ssp I (S), EcoRI (RI), Bcl I (BI), Dra I (D), Xba I (X), HindIII (HIII), Kpn I (K), and Bgl I (Bg). (B) Sequence of the 5.13-kb BamHI fragment. Nucleotides are numbered with respect to the major transcription initiation site (+1); exon 1 is boxed and the translation initiation codon (ATG) is double underlined. Arrowhead indicates cytosine at the transcriptional start point. Consensus sequences for some DNA motifs—TATA, CACCC, LBP-1, and TGGCA—are enclosed in small boxes.

#### Biochemistry: Yan et al.

trols. Luciferase activity was normalized to  $\beta$ -Gal activity. All transfected cells gave positive results, with the lowest at ≈70 times above background (pGL2-Basic). Since the construct T6-3/(-3633)Luci yielded much higher levels of luciferase activity in LLC-PK1 cells than in OK or CHO cells (data not shown), we carried out analysis of the NBAT promoter in those cells. A series of 5' deletions of T6-3/ (-3633)Luci was made to search for regulatory elements and to determine the minimum sequence needed for promoter activity (Fig. 3). A marked increase in luciferase activity was observed when the sequence -2966 to -2783 was deleted, suggesting the presence of negative regulatory elements in this region. Further deletions between -2783 to -573 did not have much effect on the high level of luciferase activity. However, an additional deletion from -573 to -158 led to a marked decrease in luciferase activity, indicating the presence of positive regulatory elements within this sequence. It should be noted, however, that the activity of the T6-3/ (-158)Luci construct was still  $\approx 70$  times higher than background. These findings indicate that the upstream sequence -573 to -2783 contains a highly efficient promoter. Furthermore, appreciable regulatory activity still resides in the proximal 5' 158 nucleotides. Clearly, additional studies on structure-function relationships of upstream sequences in the NBAT gene are needed to evaluate their importance in tissue-specific expression.

Chromosomal Localization of the Human NBAT Gene. Mapping of the NBAT gene to a human chromosome may help in understanding the role of NBAT in amino acid transport both in normal and abnormal tissues and in human





FIG. 3. Functional analysis of NBAT promoter in LLC-PK1 cells by 5' deletions. (A) Schematic representation of the T6-3 deletions/ Luci constructs. The 5' deletion fragments were generated by an ExoIII digestion and then linked to the luciferase gene. Number of nucleotides cloned in front of the luciferase gene is indicated. (B) Relative luciferase activity observed after transfection in LLC-PK1 cells. Data are expressed as the increase of luciferase activity above background per mg of protein and represent the mean of five experiments.

inherited disorders of transport such as cystinuria (22) and Hartnup disease (23). In this context, it should be noted that



FIG. 2. Primer-extension analysis of NBAT RNA from rat kidney. Arrow indicates major extension product of 81 nt. Asterisks indicate additional possible extension products. Primer used was a 27-mer oligonucleotide located at +55 to +81 in exon 1. Size marker on the left was used to determine the size of the extended products. DNA sequence generated with the 27-mer oligonucleotide on clone T6-3 is shown for calibration.

FIG. 4. Localization of human NBAT gene to chromosome 2p21. Partial metaphase showing propidium iodide staining (a) and 4',6diamidino-2-phenylindole (b) following hybridization of the probe. Arrow indicates hybridization signals and arrowhead shows band 2p21. (c) Idiogram of chromosome 2 showing hybridization signals at p21. Arrowhead represents 26 hybridization events with signals on both chromatids (doublets) and 10 hybridization events with signals on one chromatid.

NBAT is thought to be the major cystine transporter in rat, rabbit, and human kidneys (4-6). Our initial attempts to map the NBAT gene to a human chromosome with the rat genomic clone did not succeed because the signal obtained was very weak. We were, however, able to use rat NBAT cDNA to isolate a human genomic clone. A 2.2-kb insert from the human genomic clone was used as a probe for Southern analysis of the somatic cell hybrid panel of hamster-human DNAs. Hybridization revealed signals syntenic with chromosome 2 with 100% concordance (data not shown). To determine the subregional localization of the NBAT gene in chromosome 2 we used the FISH assay with a biotin-11dUTP-labeled probe to normal metaphase chromosomes. The resulting hybridization signals were highly specific to chromosome 2. A total of 36 metaphase spreads with signals on chromosome 2 were available for analysis, either on both chromatids (26 hybridization events) or on one chromatid (10 hybridization events). All the hybridization signals clustered at the p21 region of chromosome 2 (Fig. 4).

Recently, Lee *et al.* (6) were able to assign the human NBAT gene to chromosome 2 by somatic cell hybrid analysis. In the present studies, we have precisely assigned this gene to segment p21 on chromosome 2 by the FISH technique. The significance of this localization is still unclear. Additional genes that were mapped to 2p21 are the human luteinizing hormone/choriogonadotropin receptor gene (24), the forebrain cleavage gene (25), the human follicle-stimulating hormone receptor gene (26), the interferon-inducible double-stranded RNA-dependent protein kinase gene (27), and the cardiac sarcolemmal Na<sup>+</sup>-Ca<sup>2+</sup> exchanger gene (28).

- Tate, S. S., Yan, N. & Udenfriend, S. (1992) Proc. Natl. Acad. Sci. USA 89, 1-5.
- Wells, R. G. & Hediger, M. A. (1992) Proc. Natl. Acad. Sci. USA 89, 5596-5600.
- Van Winkle, L. J., Campione, A. L. & Gorman, J. M. (1988) J. Biol. Chem. 263, 3150-3163.
- Bertran, J., Werner, A., Moore, M. L., Stange, G., Markovich, D., Biber, J., Testar, X., Zorzano, A., Palacin, M. & Murer, H. (1992) Proc. Natl. Acad. Sci. USA 89, 5601–5605.
- Bertran, J., Werner, A., Chillaron, J., Nunes, V., Biber, J., Testar, X., Zorzano, A., Estivill, X., Murer, H. & Palacin, M. (1993) J. Biol. Chem. 268, 14842–14849.
- Lee, W.-S., Wells, R. G., Sabbag, R. V., Mohandas, T. K. & Hediger, M. A. (1993) J. Clin. Invest. 91, 1959–1963.

- Yan, N., Mosckovitz, R., Udenfriend, S. & Tate, S. S. (1992) Proc. Natl. Acad. Sci. USA 89, 9982–9985.
- Mosckovitz, R., Yan, N., Heimer, E., Felix, A., Tate, S. S. & Udenfriend, S. (1993) Proc. Natl. Acad. Sci. USA 90, 4022– 4026.
- Pickel, V. M., Nirenberg, M. J., Chan, J., Mosckovitz, R., Udenfriend, S. & Tate, S. S. (1993) Proc. Natl. Acad. Sci. USA 90, 7779-7783.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 11. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- 12. Pinkel, D., Straume, T. & Gray, J. W. (1986) Proc. Natl. Acad. Sci. USA 83, 2934–2938.
- Mathew, S., Murty, V. V. V. S., Hunziker, W. & Chaganti, R. S. K. (1992) Genomics 14, 775–779.
- 14. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472.
- 15. Breathnach, R. & Chambon, P. (1981) Annu. Rev. Biochem. 50,
- 349-383.
  Rupp, R. A. W. & Sippel, A. E. (1987) Nucleic Acids Res. 15, 9707-9726.
- 17. Jones, K. A., Luciw, P. A. & Duchange, N. (1988) Genes Dev. 2, 1101-1114.
- Schule, R., Muller, M., Otsuka-Murakami, H. & Reinkawit, R. (1988) Nature (London) 332, 87-90.
- Horiuchi, M., Nakamura, N., Tang, S.-S., Barrett, G. & Dzau, V. J. (1991) J. Biol. Chem. 266, 16247–16254.
- 20. Fouser, L. & Avner, E. D. (1993) Am. J. Kidney Dis. 21, 64-70.
- 21. Weiner, A. M. (1986) Annu. Rev. Biochem. 55, 631-661.
- Segal, S. & Thier, S. O. (1989) in *The Metabolic Basis of Inherited Disease*, eds. Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (McGraw-Hill, New York), pp. 2479-2496.
- Bergeron, M. & Scriver, C. R. (1992) in *The Kidney: Physiology and Pathophysiology*, eds. Seldin, D. W. & Giebisch, G. (Raven, New York), pp. 2947–2969.
- Rousseau-Merk, M. F., Misrahi, M., Atger, M., Loosfelt, H., Milgrom, E. & Berger, R. (1990) Cytogenet. Cell Genet. 54, 77-79.
- Hecht, B. K., Hecht, F. & Munke, M. (1991) Am. J. Med. Genet. 40, 130.
- Rousseau-Merk, M. F., Atger, M., Loosfelt, H., Milgrom, E. & Berger, R. (1993) Genomics 15, 222-224.
- Hanash, S. M., Beretta, L., Barcroft, C. L., Sheldon, S., Glover, T. W., Ungar, D. & Sonenberg, N. (1993) Genes Chromosomes Cancer 8, 34-37.
- Shieh, B. H., Xia, Y., Sparkes, R. S., Klisak, I., Lusis, A. J., Nicoll, D. A. & Philipson, K. D. (1992) Genomics 12, 616–617.