

Short Sequence-Paper

cDNA sequence for rkST1, a novel member of the sodium ion-dependent glucose cotransporter family¹

Kiyotaka Hitomi *, Norihiro Tsukagoshi

Department of Applied Biological Sciences, School of Agriculture, Nagoya University, Chikusa, Nagoya, 464-01, Japan

(Received 10 November 1993)

Abstract

A full length cDNA for rkST1, a novel member of the Na⁺/glucose cotransporter family, was cloned from rabbit kidney and sequenced. The coding sequence comprised 2022 base pairs and 674 amino acids. rkST1 beared 50–60% amino acid identity to the other cotransporters and was characteristic in respect of its expression in brain in addition to kidney among the cotransporters.

Key words: Sodium ion/glucose cotransport; cDNA; Nucleotide sequence

Many organic solutes are cotransported with Na⁺ in mammalian cells against a concentration gradient. Several Na⁺-coupled transporters have been characterized from various sources and some of them appeared to have significantly similar structures. The cDNAs for Na⁺/glucose cotransporters from both rabbit small intestine (SGLT1) [1] and other sources [2–5] have been isolated and sequenced. They exhibit no sequence homology with those of Na⁺-independent facilitative glucose transporters but do share a high homology with those of the rabbit kidney Na⁺/uridine cotransporter (SNST1) [6], the canine Na⁺/myo-inositol cotransporter (SMIT) from MDCK cell line [7] and the porcine Na⁺/amino acid cotransporter (SAAT1) from LLC-PK₁ cell line [8]. Furthermore, these transporters have several conserved amino acid sequences in common with the bacterial Na⁺ cotransporters for proline [9], panthothenate [10] and glutamate [11], indicating they comprise a gene family related to SGLT1.

To characterize genes other than previously cloned cotransporters, the reverse transcriptase-polymerase chain reaction (RT-PCR) have been used to clone novel cDNAs of the gene family using degenerated primers that could anneal to consensus sequences.

Although the amino acid sequences are fairly homologous among Na⁺/glucose cotransporters, there existed several highly conserved regions even at a nucleotide level. A pair of degenerated primers (sense; 5'-CCTCTCTGTTTGCCAGYAATATYGGGRAGTG-3', anti-sense; 5'-GTCTGTAGKGTRTCYGTGTASATCA-3') was designed and synthesized chemically based on the nucleotide sequences highly conserved among SGLT1, SNST1, and SMIT (nt 238–267 and nt 621–646 in SGLT1). Total RNA was isolated by guanidium-thiocyanate method from rabbit kidney, since this organ has been shown to contain a number of Na⁺-dependent cotransporters related to SGLT1 [12]. Poly(A)⁺ RNA purified using oligo (dT)-cellulose column was reverse transcribed in an appropriate buffer by the addition of the anti-sense primer and reverse transcriptase. The resulting cDNAs were subjected to amplification by PCR with a pair of degenerated primers. The PCR products with the expected size (approx. 410 bp) were cloned into pUC118, classified based on the fragments generated by restriction enzyme digestion and sequenced. In addition to SGLT1 and SNST1, two novel cDNA clones were isolated. One clone (ST-A) was highly homologous (90% amino acid identity) to SMIT, indicating the presence of the Na⁺/myo-inositol cotransporter in rabbit kidney. The other clone (ST-B) showed rather low similarity (50–60% amino acid identity) with SGLT1, SNST1 and SMIT. Therefore, a full length cDNA for ST-B (rkST1) was then isolated by screening 2 · 10⁶ phages of a

* Corresponding author. Fax: +81 52 7814447.

¹ The nucleotide sequence reported in this paper has been submitted to the DDBJ data bank under the accession number of D16226.

rabbit kidney cDNA library with the radiolabeled ST-B cDNA insert as a probe and sequenced on both strands. The complete nucleotide sequence of 2157 bp and the deduced amino acid sequence are shown in Fig. 1. A single open reading frame (nt 28–nt 2049) encodes a polypeptide of 674 amino acid residues with the non-glycosylated molecular weight of 73 629. ATG at nt 28 is considered as a start codon, since there is high homology between the putative coding region of rkST1 downstream of this ATG and those of other members of Na⁺/cotransporter family. Furthermore, this is con-

tained in the consensus initiation sequence, YN-NAUGG [13]. A poly(A) tail is observed at the 3' end and a consensus polyadenylation sequence (AATAAA) is located 26 nt upstream from the poly(A) stretch. The length of the rkST1 cDNA was in good agreement with the mRNA size of approximate 2.2 kb analyzed by Northern blotting as shown in Fig. 3.

The deduced amino acid sequence of rkST1 was compared with previously cloned Na⁺ cotransporters (Fig. 2). The sequences shows significant identity to those of SGLT1 (57%), SNST1 (51%), SMIT (51%)

```

10      20      30      40      50      60      70      80      90      100     110     120
AAGGAGTCGGAGCGCGTGTTTCAGGACCATGGAGAGCAGCAGCAGTCTCAGCCCGCGTGTCCGATCCCTTGGATCCCTTTCCCGAGGAGGCGCTGGAGCCAGGTGACATTGCAGTG
M E S S T S S P Q P P L S D P L D P F P Q R S L E P G D I A V

130     140     150     160     170     180     190     200     210     220     230     240
CTGGTCTGTACTCTCTCTTTGTCCTGGCTGTTGGATTATGGTCCACAGTGAAGACCAAGAGACACAGTGAAGAGCTACTTCTGGCTGGAGGGACATGGTGTGGTGGCGGTGGG
L V L Y F L F V L A V G L W S T V K T K R D T V K G Y F L A G G D M V W W P V G

250     260     270     280     290     300     310     320     330     340     350     360
GCCTCTGTTCGCCAGCAACGTTGGGAGTGGGCACTTCGTGGGCTGGCAGGCTGGGGC GCTGCAACAGGAATTTCTGTAGTGTCTTACAGAGTTTAAACGGCATGTTTCCGTGCTGATG
A S L F A S N V G S G H F V G L A G S G A A T G I S V A A Y E F N G M F S V L M

370     380     390     400     410     420     430     440     450     460     470     480
TTGGCTGGATCTTCTCCCTATCTACATCGCGGGCCAGGTACACAGATGCCGAATACCTGCGGAGGCGCTTCGGCGGCAGCAGGATCGCCATCACCTGGCGGCTCTACCTGTTT
L A W I F L P I Y I A G Q V T T M P E Y L R R R F G G S R I A I T L A V L Y L F

490     500     510     520     530     540     550     560     570     580     590     600
ATCTACATCTTACCAAGATCTCGTGGACATGTACACCGCGCCATCTTTATTACAGCAGTCCCTGCACCTGGATCTGTACCTGTCCGTAGTCCGGCTGTGGCGGTACAGCGCTGTAC
I Y I F T K I S V D M Y T G A I F I Q Q S L H L D L Y L S V V G L L A V T A L Y

610     620     630     640     650     660     670     680     690     700     710     720
ACTGTTGCAAGTGGCTGGCGCTGTGATCTACACGATGCTTGCAGACTCTGATCATGCTCGTCGGAGCGCTCACCTTGATGGGTACAGTTTGTGTCAGTTGGCGGCATGGAAGGC
T V A G G L A A V I Y T D A L Q T L I M L V G A L T L M G Y S F A A V G G M E G

730     740     750     760     770     780     790     800     810     820     830     840
CTGCAGGAAGTACTTCTGGCCCTGCCAGCAACCGGAGTGAAGACAGCAGTCCGGGCTGCCCGGGAGAGCGCTTCCATCTCTTCGGTATCCGCTGACGTCTGATCTGCCGTGG
L Q E K Y F L A L P S N R S E N S S C G L P R E D A F H L F R D P L T S D L P W

850     860     870     880     890     900     910     920     930     940     950     960
CCCGGGATCTGTTGGGATGTCCATCCCATCCCTGGTACTGGTGCACCGACCAAGGTG ATCGTCCAGCGGAGTCTGCCCGCAAGAACCTGTCTCACGCCAAAGAGGGTCTCTAATG
P G I L F G M S I P S L W Y W C T D Q V I V Q R S L A A K N L S H A K G G S L M

970     980     990     1000    1010    1020    1030    1040    1050    1060    1070    1080
GCCGCTCTGAAAGTGTGCTCCCTTTTCATCATGGTGTTCCTGGAGTGGTGCAGCGCATCTCTTCCAGATCAAGTGGCTTGTGTCAGAGCCAGAAACCTGCCAGAGGGTCTGCAAC
A A Y L K V L P L F I M V F P G H V S R I L F P D Q V A C A D P E T C Q R V C N

1090    1100    1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
AACCCTCCGGCTGTTCGACATTGCTATCCCAAGCTGGTCTGGAATCCTGCCACA GGACTCCGGTGGGCTCATGATGGCTGTGATGGTGTGCTGCGCTCATGCTCCCTCACCTCC
N P S G C S D I A Y P K L V L E L L P T G L R G L M M A V M V A L M S S L T S

1210    1220    1230    1240    1250    1260    1270    1280    1290    1300    1310    1320
ATCTTTAAGAGTGCAGCACCATTCTTACCATTGGACCTCTGGAACCACTCCGGCTCGGGCATCTGAGAAGGAGCTCATGATGTTGGGAGGGTATTGTGCTTCTGCTGGTGTGGTC
I F N S A S T I F T M D L W N H V R P R A S E K E L M I V G R V F V L L V L V

1330    1340    1350    1360    1370    1380    1390    1400    1410    1420    1430    1440
TCCGTGCTCTGGATCCCTGTGTCAGGCCAGCCAGGAGGCACTGTTGTCTATATCAAGCCATCAGCTCCTACCTGCAGCGGCTGTGGCCATGGTCTCTGCTTGGGGTGTGTTT
S V L W I P V V Q A S Q G G Q L F V Y I Q A I S S Y L Q P P V A M V F V L G C F

1450    1460    1470    1480    1490    1500    1510    1520    1530    1540    1550    1560
TGGAAACGGGCCAATGAAAGGGGGCTTCTGGGGCTGGTCTTGGGCTGCTCCTGGGCTTCATCAGGCTGATCTGGACTTACATTACGTGGAGCTGCCTGCCACGACCCGACGAG
W K R A N E K G A F W G L V L G L L L G F I R L I L D F I Y V E P A C H Q P D E

1570    1580    1590    1600    1610    1620    1630    1640    1650    1660    1670    1680
CGCCCTCTGTGGTGAAGAACGTCACCTACTCTTCCATGATCTGCTTCCGGTCTACTGTCTCAGCGTACCGGTGACGGTTCATGAGCTGCTCACAGAGCCACCCTCAAGGAGATGATC
R P S V V K N V H Y L Y F S M I L S S V T V L T V T V M S L L T E P P S K E M I

1690    1700    1710    1720    1730    1740    1750    1760    1770    1780    1790    1800
AGCCACCTGACCTGGTCTACTCGCGCGACCCGCTGGTCCAGAGGCACAGTGCACGACGCCACGCGCTGCCCTGTCTCACACGGGACCGGGAAGCCAAACAGCGCCAGC
S H L T W F T R R D P V V Q K A Q V P A A T P L P P A L S H N G T T A E A N S A S

1810    1820    1830    1840    1850    1860    1870    1880    1890    1900    1910    1920
ATCCAGCTGGAGACGATTACAGGAAGCGCGCTCAAAGCCACAGCAGTACGTGACCCCGAAGCAGTCCAGAGTGGTGGAGCCCTGCTGTGGCTCTGTGGATGGAGGGCAAGAGCAG
I Q L E T I Q E G A S K A H S S D V T P K Q S R V V R A L L W L C G M E G K S T

1930    1940    1950    1960    1970    1980    1990    2000    2010    2020    2030    2040
GAGCAGCGCCGAGGCCAGAGGCCCTTGGCTTCCATAGAAGAAACCCCGTCGTG AAGACCTCCTGGACGTCAACTGCTGCTGATCTGTGCGCTTCTCTCTGGGGC
E Q A P R P A E P V L A S I E E N P V V K T L L D V N C L L C I C C A F F L W G

2050    2060    2070    2080    2090    2100    2110    2120    2130    2140    2150
TACTTCGCTAGCGCTGCGCTACGCCCGCACTCCTTCCCGTGCCTGTTGTTTATA ATGAAGAAAGATAATAATAATAAGCTTTTGTGTTTACCGTGAAAAA
Y F A

```

Fig. 1. Complete nucleotide and deduced amino acid sequences of rkST1. Polyadenylation signal is underlined.

```

rkST1 1: MESSTSSQPPLSDPLDPFPQSL-EPGDIAVLVLYFLVLAFLVSTVTKRDTVKGYFLAGDMVWMPVGASLFASNVSGHFLVGLAGSAAATGISVA
SGLT1 1: D...L..LT-T.TAAPLESYERIRNAA..S.I.I..V.M....AMFS.N.G..G.F...RS....I.....I.....T...S..ATG
SNST1 1: --EHMEAGS-RL-G--GDHGALIDN.A...IAA...L.IG....MCR.N.G..G....RS.....I.....I.....T...N..LA..
SMIT 1: -----MRAV--TA...IVA...IL.MCI.FFAMW.SN.S.S....RS.T.VAI...V..I..E..I.....S.FA.G
SAAT1 1: --A..L..STVTKT.GP.EISERIQNAA..S.I.I..VV.M....AMLR.N.G..G.F...R.VT...M.....I.....T...S..AI.
          **      * * * * *      * * * * *      * * * * *      * * * * *

rkST1 100: AYEFNMGFVSLMLAWIFLPIYIAGQVTTMPEYLRRRFGGSRIATLAVLYLFIYIFTKISVDMYTGAIFIQSSLHLDLYLSVVGGLAVTALTYTVAGGLAA
SGLT1 100: GF.W.ALIM.VV.G.V.V...RAG.V.....QK...K..Q.Y.SI.S.LL.....A.IFS.....LT.G..I.VAIII..VI.G...IT....
SNST1 97: GF.W.AL.V..L.G.L.A.V.LTAG.I...Q...K...H..RLY.S..S..L.....FS..V...A.GWNI.A..IA..GI.MV...T....
SMIT 81: W...ALLLQL.G.V.I...RSG.Y...SK...H..QVYF.A.S.IL...L...L.S.L..E..GWN..V..IL.IGM...L.T...V..
SAAT1 100: F.W.ALLLL.V.G.F.V...KAG.M.....K...K.IQ.Y.SI.S...CVALR..S.IFS.....KLA.G....AIFS..I..I..IT....S
          * *      * * * * *      * * * * *      * *      * * * * *      * * * * *

rkST1 200: VIYTDALQTLIMLVGALTIMGYSFAAVGMEGLQEKYFLALP---S--[N-R-SPNSS]CGL-PREDAPHLFRDPLTSDLPWPGILFGMSIPSLMYWCTDQ
SGLT1 200: ...T...A..M..SVI.T.FA.HE...Y.AFT...MR.I.SQI.YGN---T.IPKR..YT.....I...AI.G.I...LV.....LT.....
SNST1 197: LM...TV..FVIA..PI.T.A.HE...YS..FD..MG.MTSLVSEDPAVGNIS...YR..P.SY..L...V.G...AL.L.LT.V.G...S..
SMIT 181: ...T..A.L.I.....II.MMEI..F.EVKRR.M..S.NVT..ILLTYNL.NTN..NVH.KK..LKML.N.TDE.V...FVL.QTPA.V...A..
SAAT1 200: ...T...I...I.SPI...FA.VE...Y.SFT...MN.I.TIVEGDN---LTISPK..YT.QG.S..I...AV.G.I...MI...TVVAA.....
          *** *      * *      * *      * *      * *      * *      * *      * *      * *      * *      * *

rkST1 291: VIVQRSLAAN[SHARGGS]MAAYLKVLPLFIMVFPGMVSRILFPDQVACADPETCORVCN[SPSCSD]IAYPKLVLELLPTGLRGLMAVMVAALMSSLT
SGLT1 296: ...C.S....V.A.CILCG...M.M.LI.MM....YT.K...VV.SE.E.Y.GTRV..TN..F.T.V..M.N....LS..M.S.....
SNST1 296: ...C..GR..T.I.A.CILCG...LT.M.L.M...I...Y..E...VA..V.K...GTEV..N..R.VK.M.N....L...L....A
SMIT 281: ...V....IA...ST...GF..L.M..I.V...I...A.DI..IN..H.MQ..GSRA...N...R.MK.V.V.....I....D.D
SAAT1 296: ...C.SG.DM..V.AACI.CG...L.M.L.M...I...YTEK...VV.SE.VKH.GTEV..NY...L.M..M.S....LS..L.S....
          ***** *      * *      * * * * *      * * * * *      * * * * *      * * * * *      * * * * *

rkST1 391: SIFNSASTIFTMDLNHVRPRASEKELMIVGRVFLVLLVLSVLMIPVVOASQGGOLFVYIQAISSYLQPPVAMVFLVLCGFWRKANERKGAFLVGLVGLLL
SGLT1 396: ...L...L...IYTKI.KK....A..L.M.F.IGI..IA.V.I..SA.S...D...S.T..G..I.A..AI...V..P.....F.I
SNST1 396: ...S..L...IYT-L...G.G..LL..LW.VFI..A..A.S....D...SV...A...A...VAL.VP.V.....IG...M
SMIT 381: ...L.VYKLI.RS..SR.....I..AFM.VI..IA.V.IIVEM...MYL...EVAD..T...AL.L.AI...C..Q...Y.GMA.FV.
SAAT1 396: ...L...L...YTKI.RQ....L.A..L.II..IVI..IV.V.L.VA.N...H..ES...G..I.A..L.AI.C..V..Q....II..FVM
          ***** * *      * *      * *      * *      * *      * *      * *      * *      * *      * *

rkST1 491: GFIRLILDFIYVEPACHQPDERSVVKVNHVLYFSMILSSVTVLTVVMSLITEPPSKEMISHLT-W----F-TR----RDPVVQKAQVPAATPLPAL
SGLT1 496: .IS.M.TE.A.GTGS.ME.SNC.TIICG....AI..FVISII..V.V..F.K.IPDVHLYR.CWSLRNS---KEERID-L.AGEEDIQEAPEEATDTEV
SNST1 495: .IA...PE.SFGTGS.VR.SAC.AFLCR....AIV.FPCSG.LIIIV..C.A.IPRKHLR.VFSLRHS---KEERID-L.ADELEAPASPPVQNGRPE
SMIT 481: .AV..T.A.A.RA.E.D...N..GFI.DI..M.VATA.FW..G.ITVIV...P..T..Q.RIT.F.SKSLVVKESCSK.P.EPYKMQEKSILRCSENSE
SAAT1 496: .L..M.AE.V.GTGS.LAASNC.QIICG....AL..FF.SI.V.LAI...K.IPDVHLYR.CWALRNS---KEERID-L.AEKKRHEEADGDVEDNP
          * *      * *      * *      * *      * *      * *      * *      * *      * *      * *

rkST1 580: SH[NT]EANSASIOLETFIQEGASKAHSSDVTPKQSRVVRALLWLCMEGKSTEQAPRAEPVLAIEENPVVKTLLDVNCLLICCAFFLWGYFA
SGLT1 593: PKKKKGFFRAYDLFCGLDQKQPKMKTEEAAMKILKLTDTSEHLNRTVVNINGVILLAVAVFCYAYFA
SNST1 592: HAVEMEPEQAPGPGFLFRQCLLWFCGMNRGRAGGPAPPTQEEERAAARLEINEDPMSRVVNLNALLMAVAMFFWGFYA
SMIT 581: AINHIIPNGKSEDSIKGLQPEDVNLVTCREEGNPFVSLGHSEAEPTVDAYSNGQALMGEXKERRKETEDGGRYWKFIDWFCGFKSKLSKRSIRDLME
SAAT1 593: EETRGCLRKAYDLFCGLQKRGPKLSKEEEAQKRKLDTTSEKPLNKTIVNIRAILLAVAVFVGYFA
          * *      * *      * *      * *      * *      * *      * *      * *      * *      * *

SMIT 681: EAVCLQMLEEPPQVKLLILNIGLFAVCSLGIIMFVYFSL

```

Fig. 2. Alignment of deduced amino acid sequences for rkST1 with SGLT1, SNST1 from rabbit kidney, SMIT from MDCK cell line, SAAT1 from LLC-PK₁ cell line. Dashes indicate gaps introduced to obtain an optimal alignment. Amino acid identity to rkST1 is indicated with dots. The asterisks indicate the conserved amino acids among all Na⁺ cotransporters. The predicted N-glycosylation sites in rkST1 are boxed.

and SAAT1 (52%). N-terminal portions are generally conserved and most of the divergence in sequence is in C-terminal portions. This characteristic is commonly observed among all these transporters. The 'SOB motif' [11] (G₃₇₅-A₄₁₂-XXXX-L₄₁₇-XXX-G₄₂₁-R₄₂₂; X denotes any amino acids), which may play an important role in sodium ion binding, is also well conserved. Furthermore, there are five potential N-glycosylation sites, two (N₂₄₃ and N₂₄₆) of which are situated at similar positions of all the other cotransporters (Fig. 2). The hydropathy profile (data not shown) is nearly identical to those of the other Na⁺ cotransporters suggesting rkST1 has twelve membrane-spanning regions. All these data indicate rkST1 is a novel member of Na⁺/glucose cotransporter family.

Tissue distribution of rkST1 was investigated by Northern blot analysis using a specific probe in order

to prevent the cross-hybridization between rkST1 and the other Na⁺ cotransporters. This probe was a 796-bp cDNA fragment (nt 1331–nt 2126) which was the most divergent region in sequence among the Na⁺ cotransporters. mRNA for the rkST1 gene was detected at a relatively low level in both kidney and brain, but was undetectable in lung, liver, intestine, heart and spleen (Fig. 3). All Na⁺ cotransporters isolated so far were expressed in kidney. Furthermore, SGLT1, SNST1 and SAAT1 were expressed in intestine, heart and intestine/spleen, respectively [1,6,8]. Therefore, the tissue distribution of rkST1 was quite characteristic among Na⁺ cotransporters. Works are in progress to identify the substrate of the transporter, rkST1.

We are grateful to Dr. Ernest M. Wright for providing valuable information for designing degenerated primers.

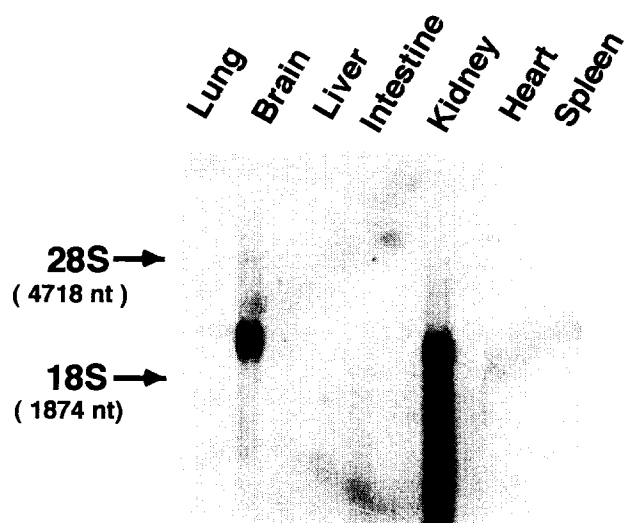


Fig. 3. Tissue distribution of rkST1 mRNA. 5 μ g of poly(A)⁺ RNA prepared from indicated rabbit tissues was fractionated by electrophoresis on formaldehyde gels, transferred to Nylon filter, hybridized with a rkST1 cDNA probe as described in the text. The 28S and 18S rRNA are indicated.

1. References

- [1] Hediger, M.A. et al. (1987) *Nature* 330, 379–381.
- [2] Coady, M.J. et al. (1990) *Am. J. Physiol.* 259, C605–C610.
- [3] Hediger, M.A. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86, 5748–5752.
- [4] Ohta, T. et al. (1990) *Mol. Cell. Biol.* 10, 6491–6499.
- [5] Wells, R.G. et al. (1992) *Am. J. Physiol.* 263, F459–F465.
- [6] Pajor, A.M. et al. (1992) *J. Biol. Chem.* 267, 3557–3560.
- [7] Kwon, H.M. et al. (1992) *J. Biol. Chem.* 267, 6297–6301.
- [8] Kong, C.-T. et al. (1993) *J. Biol. Chem.* 268, 1509–1512.
- [9] Nakano, T. et al. (1987) *Mol. Gen. Genet.* 208, 70–75.
- [10] Jackowski, S. et al. (1990) *J. Bacteriol.* 172, 3842–3848.
- [11] Deguchi, Y. et al. (1990) *J. Biol. Chem.* 265, 21704–21708.
- [12] Coady, M. et al. (1990) *Arch. Biochem. Biophys.* 283, 130–134.
- [13] Kozack, (1987) *Nucleic Acids Res.* 15, 8125–8132.