Inverse polymerase chain reaction mediated chromosome walking within the human glutamic acid decarboxylase gene

NEELIMA KHAIRATKAR-JOSHI and NASIR S AHMAD*

Department of Natural Product Biology, Research Centre, Hoechst Marion Roussel Ltd., LBS Road, Mulund (W), Mumbai 400 080, India

**Corresponding author (Fax, 91-22-564 1953; Email, ahmad@HOECHST RES.RPGMS.ems.vsnl.net.in).*

Using inverse polymerase chain reaction (PCR), we have cloned partial intronic sequences from human glutamic acid decarboxylase (GAD) gene. A small 153 bp core region was selected from the GAD cDNA sequence to design outward primers corresponding to its 3′ and 5′ ends. *Eco*RI digested human DNA which had been circularized by self-ligation and then linearized with *Sac*II was used as a substrate to carry out PCR. This gave a 900 bp long product which was cloned into pUC19. The sequence analysis of this fragment revealed the presence of introns in the region flanking the selected core DNA. In this work we used this technique to walk into the upsteam region of the GAD gene using sequence information from its cloned cDNA.

Introduction

Inverse polymerase chain reaction (PCR) provides the means to amplify unknown sequences that flank a known stretch of DNA referred to as a "core sequence". These sequences are not accessible in a conventional PCR where the primers anneal to the ends of the core sequence and mediate extension of only the enclosed region. In inverse PCR, the primers bind to the core region but are oriented "outwards". To get them in the right orientation, the flanking sequences are internalized by circularization of the pre-digested fragment. The circular DNA may be linearized with an enzyme which only cuts within the core region between the two primers as shown in figure 1. As a result, the two primers end up oriented towards each other and can now amplify the internalized sequence. Thus regions that lie outside the boundary of a known sequence are amplified. This technique was devised and implemented almost simultaneously by three labs (Ochman *et al* 1988; Triglia *et al* 1988; Silver and Kirikatte 1989). It has been used by others to clone different chromosomal regions such as immunoglobulin heavy chain variable region genes (Sinclair and Aitken 1995). Inverse PCR can be used as a convenient tool in chromosome walking if sequence information of the genomic or cDNA clone is available to serve as a core region.

Using inverse PCR, we have detected and cloned partial sequences from two introns within the human glutamic acid decarboxylase (GAD) gene. GAD has been considered as one of the candidate autoantigens involved in the early stages of development of human type I diabetes (Baekkeskov *et al* 1990; Kaufman *et al* 1993). Results of our inverse PCR and some of the parameters that were found to be crucial in our reaction are discussed.

2. Materials and methods

2.1 *Plasmid constructs*

Plasmid pUC18\C12 containing the full length cDNA sequence of the GAD65 gene was kindly provided by Dr Maruyama from Hoechst Japan.

2.2 *DNA preparation*

Human genomic DNA was prepared from WBCs (Hermann and Frischauf 1987). This DNA was manipulated as described in the following steps. (1) It was digested to completion with *Eco*RI. (2) The fragments were circularized by ligation in dilute solutions (Collins

Keywords. Introns; polymerase chain reaction; glutamic acid decarboxylase.

and Weissman 1984). About 50 ng digested DNA was self-ligated in 100 µl reaction volume using 0·5 Weiss units of T4 DNA ligase (Gibco BRL) at 4°C for 24 h. Forty such reactions were set up and DNA from these reactions was pooled to get approximately 2 µg of circularized DNA. An aliquot of about 300 ng was kept aside. (3) The rest of the cyclyzed DNA was phenol extracted, ethanol precipitated and digested with *Sac*II to completion in set 3(a) and partially in set 3(b). The digests were then concentrated by ethanol precipitation. About 150 ng of DNA from steps (2), (3a) and (3b) were used as substrate in three separate PCRs.

2.3 *PCR*

We selected a 154 bp long core region from GAD cDNA sequence starting from nucleotide + 118 to + 272 (A of ATG is + 1) in order to design outward facing primers. Nucleotide 134 to 118 formed the upstream primer and ntd 256 to 272 formed the downstream primer (both 5′ to 3′). The primers were obtained from the National Chemical Laboratory, Pune and the sequences were as follows:

Forward primer:
5′ GACAAGCTTCACAGTTTGTTTCCGAT 3′

Reverse primer:
5′ GACAAGCTTGATGTCAACTACGCGTT 3′

Both primers had a *Hind*III add-on site at 5′ end which was utilized in cloning. PCR was carried out in the thermal cycler of Perkin Elmer (Gene Amp 9600) using the thin walled tubes. Initial denaturation prior to cycling was carried out at 94°C for 30 min for *Sac*II undigested DNA and 1 min for the digested ones. Other parameters were same for the three sets of reactions as described below. We used 30 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and extension by Taq polymerase (Elongase system of Gibco BRL) at 68°C for 2 min. The reaction was carried out in 50 µl volume in the presence of 1·6 mM $MgCl_2$, 200 µM of each of the dNTPs and 50 *p*mol of each of the two primers.

2.4 *Cloning*

The PCR product was eluted from the gel on DEAE cellulose paper and cloned in pUC19. Elution of the PCR product was carried out as described by Maniatis *et al* (1987). The plasmid pUC18/C12 was digested with restriction enzymes *Eco*RI-*Sal*I, *Eco*RI-*Hind*III, *Eco*RI-*Xmn*I, and *Pst*I and subjected to electrophoresis on 1·4% agarose gel. The digested DNA was blotted onto nitrocellulose paper and hybridized to ^{32}PdCTP-labelled 900 bp PCR product. Southern blotting and hybridization was carried out as described by Maniatis *et al* (1987) except that hybridization was carried out in plastic boxes containing 15 ml of the hybridization mixture.

2.5 *Sequencing*

The sequencing of the PCR product was done by Taq polymerase using the automated DNA sequencing system from Applied Biosystems Inc. Sequencing and database search was courtesy of Dr K-P Koller, Hoechst AG, Germany.

3. Results and discussions

On the basis of published reports we attempted to carry out inverse PCR on three potential templates obtained from steps 2, 3(a) and (b) of the methods. We observed the following:

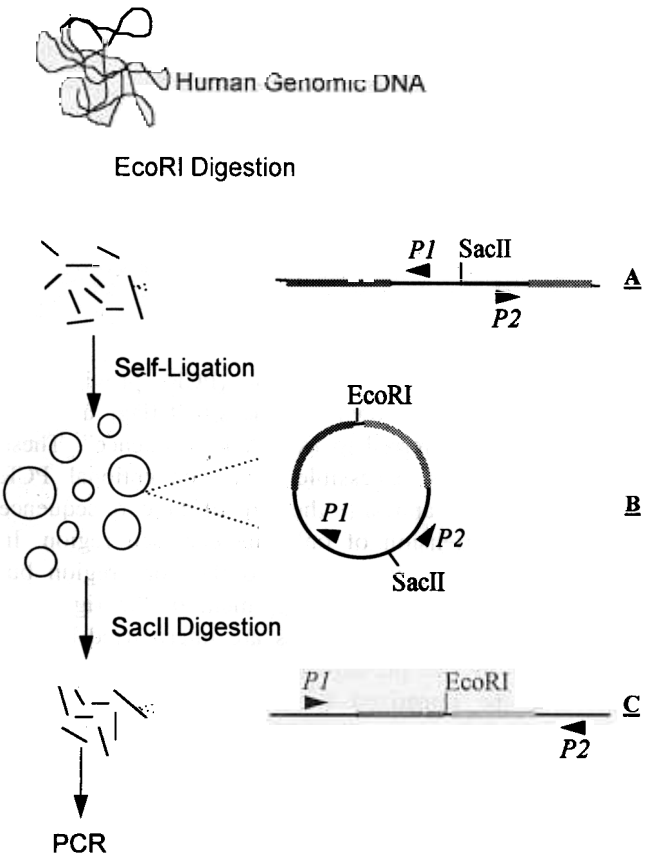

Figure 1. Strategy for inverse PCR: Genomic DNA is digested with *Eco*RI generating a large number of fragments some of which will contain required primer binding sequences. One such fragment is shown separately towards the right-side of the figure. (A) Note outward oriented primer binding site *p1* and *p2*. (B) The flanking sequences are internalized by self ligation of *Eco*RI ends to form a circular molecule. (C) The circularized DNA is digested with *Sac*II, so that the primer binding sites get oriented towards each other to which the primers anneal during PCR. The flanking sequences thus become accessible for amplification. (—), Leftward flanking sequence; (⋯), rightward flanking sequence; (—), core region; (►), primer binding sites.

(i) The circularized template DNA from step 2 did not give a PCR product. Our attempts to introduce nicks in the circularized DNA by heating did not produce the required result. This was in contrast to the observation of Triglia *et al* (1988).

(ii) We therefore used a linearized DNA substrate which was obtained by digesting with restriction enzyme *Sac*II as described in step 3(a) of the methods. Linearization was carried out so as to generate a DNA fragment capable of binding to the respective primers at its two ends (figure 1). In order to achieve this we chose the restriction enzyme *Sac*II as it is reported to have only one restriction site in the core region between the two primer sequences (Karlsen *et al* 1991). This substrate yielded no PCR product. One possible reason for not obtaining the PCR product could be the presence of additional *Sac*II site(s) in the flanking sequence. In that case, *Sac*II digestion would split the target molecule into two fragments, each binding to just one primer, as a result of which no PCR product would be formed. Since no other suitable restriction enzyme site was available in the core region inbetween the two primer sequences, we carried out partial digestion of the circularized DNA with *Sac*II. Here the idea was to generate at least a few intact target fragments which retain both primer binding sites at the ends to enable them to function as substrate in PCR. When such a manipulated template from step 3(b) was used, a 900 bp product was obtained.

(iii) In our experience, linearization of the circularized DNA template was crucial for the success of inverse PCR. The predicted *Sac*II site in the flanking sequence was confirmed by DNA sequencing as discussed below.

The 900 bp PCR fragment obtained from the above mentioned reaction was hybridized to restriction digests of pUC18/C12 . This was done to confirm that it arose from the GAD region and was not a product of nonspecific primer binding. Only the 5′ end fragments of the GAD cDNA viz., *Eco*RI-*Sal*I (2·4 kb), *Eco*RI-*Hind*III (2·1 kb), *Eco*RI-*Xmn*I (428 bp) and *Pst*I (290 kb) hybridized to the labelled PCR product (figure 2). The PCR product was then cloned in pUC19. At this stage we suspected the presence of introns in this fragment as the fragment

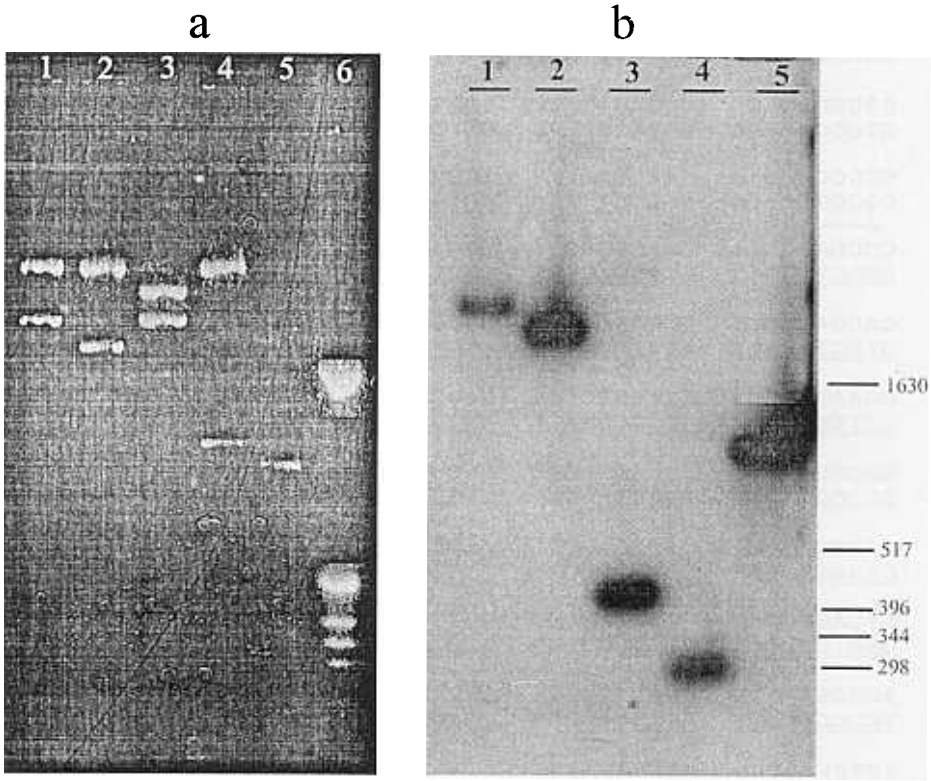

Figure 2. (**a**) Agarose gel electrophoresis of restriction digests of pUC18/C12 plasmid DNA. The digests were run on 1·4% agarose gel. The digests loaded were as follows: Lane 1, *Eco*RI-*Sal*I; lane 2, *Eco*RI-*Hind*III; lane 3, *Eco*RI-*Xmn*I; lane 4, *Pst*I; lane 5, 900 bp PCR product; lane 6, pBR322 DNA digested with *Hinf*I, this was used as DNA size standard. (**b**) An Autoradiogram obtained following Southern transfer and hybridization of DNA samples shown in (**a**). Only the restriction fragments present towards 5′ end of the GAD cDNA i.e., *Eco*RI-*Sal*I (2·4 kb), *Eco*RI-*Hind*III (2·1 kb), *Eco*RI-*Xmn*I (428 bp) and *Pst*I (290 kb) show the hybridization signal.

was obtained from *Eco*RI digestion of human chromosomal DNA and there is no reported *Eco*RI site in the human GAD cDNA (Karlsen *et al* 1991). This hypothesis was confirmed by sequencing the cloned fragment. The sequence is shown in figure 3. Being product of inverse PCR, the intronic regions contained in the cloned fragment

```
      agcttcacag  tttgtttccG  ATGCCGCCCG  TGAACTTCTG  AGCCACTTGG
      tcgaagtgtc  aaacaaaggC  TACGGCGGGC  ACTTGAAGAC  TCGGTGAACC

51    CACCAGGCTC  GCGCTGCAAG  GAAGAATAGG  CAGGCAGACA  GCGAGGCAGA
      GTGGTCCGAG  CGCGACGTTC  CTTCTTATCC  GTCCGTCTGT  CGCTCCGTCT

      AGGAGAACAA  CAGACGGACG  TGTGAAATTT  CGTTCTATGT  GAAAATCAAT
      TCCTCTTGTT  GTCTGCCTGC  ACACTTTAAA  GCAAGATACA  CTTTTAGTTA

151   CAGTCCGGGG  TCCTAGGAGG  GGCCGCCTTC  CTCCTGGGGG  CCGCTCAGGA
      GTCAGGCCCC  AGGATCCTCC  CCGGCGGAAG  GAGGACCCCC  GGCGAGTCCT

201   CTCGGTGAGG  CCGGCAGGTG  CAGAAGGCAA  GCGGGAGGCT  GGCGGCCCGG
      GAGCCACTCC  GGCCGTCCAC  GTCTTCCGTT  CGCCCTCCGA  CCGCCGGGCC

251   GCCAGCCCTC  ATTTCCTTTA  TCTGTTTCCC  GAGGAGAAGC  CCACACCGCT
      CGGTCGGGAG  TAAAGGAAAT  AGACAAAGGG  CTCCTCTTCG  GGTGTGGCGA

301   GCCCTTCCAG  GTGCCAATCG  GCTCAGGCAA  ATTCCTCTTC  CCTGCCTACC
      CGGGAAGGTC  CACGGTTAGC  CGAGTCCGTT  TAAGGAGAAG  GGACGGATGG

351   TTGGACGAGT  TTTCCTAAAC  ACTGAAGCTA  TCCATTTTTT  CGCGGACTGT
      AACCTGCTCA  AAAGGATTTG  TGACTTCGAT  AGGTAAAAAA  GCGCCTGACA

401   CTTTCTAGAT  AAAGaAATCT  CTCACCTTTC  CCATTTTTTT  TTTCTATCTT
      GAAAGATCTA  TTTCtTTAGA  GAGTGGAAAG  GGTAAAAAAA  AAAGATAGAA

451   CACGGTCACT  CCGTCCCACC  CCAGGGTCCA  GACACCCGGA  TCTTAGCAGC
      GTGCCAGTGA  GGCAGGGTGG  GGTCCCAGGT  CTGTGGGCCT  AGAATCGTCG
                              EcoRI
501   GCCCCAGGAC  CACGGAATTC  TCGGGCACGG  TATTTGGAAA  CTGAGTCCCA
      CGGGGTCCTG  GTGCCTTAAG  AGCCCGTGCC  ATAAACCTTT  GACTCAGGGT
      <-- Leftside Intron
551   CCCGAGGCCC  CAGCGGCCTA  GGGAAGCCAG  CCGTCACGAT  TCAGTGAGCG
      GGGCTCCGGG  GTCGCCGGAT  CCCTTCGGTC  GGCAGTGCTA  AGTCACTCGC

601   CACGANCGCC  TGGATCGTCG  GCCACCCGCC  ATCCCAAGGG  CCGCGCTGGG
      GTGCTNGCGG  ACCTAGCAGC  CGGTGGGCGG  TAGGGTTCCC  GGCGCGACCC

651   GGAAGAGCGC  CACGGCCTCA  GAGCCCTCTT  GGAGCTCCGA  GGCCAGGTTC
      CCTTCTCGCG  GTGCCGGAGT  CTCGGGAGAA  CCTCGAGGCT  CCGGTCCAAG

701   GGGGGCTTGG  ACCCAGCGGG  CAGGTCCCGC  ATTTCTCGCT  TTTATCTGAA
      CCCCCGAACC  TGGGTCGCCC  GTCCAGGGCG  TAAAGAGCGA  AAATAGACTT

751   GATTCTGGCA  GGTCCCGCCT  CCGCTGCCAC  CTCGACGCCC  CTGTCCGGCT
      CTAAGACCGT  CCAGGGCGGA  GGCGACGGTG  GAGCTGCGGG  GACAGGCCGA
                              SacII
801   CTCAGCGCCG  GACCCGCACC  GGCCGCGGTG  GGTGGGGCAG  AGAGAGGGGC
      GAGTCGCGGC  CTGGGCGTGG  CCGGCGCCAC  CCACCCCGTC  TCTCTCCCCG

851   AGGGGCGCCC  CAGGGCCCCC  CGCTGAGTCT  TTACCTGTTG  CATGGAGAaa
      TCCCCGCGGG  GTCCCGGGGG  GCGACTCAGA  AATGGACAAC  GTACCTCTtt
                              <-- Rightside Intron
900   cgcgtagttg  acatcaagct  920
      gcgcatcaac  tgtagttcga
```

Figure 3. Nucleotide sequence of the PCR product: Numbers in the left margin correspond to the nucleotide position. Primer sequences are shown in lower case letters. The *Eco*RI site created by recircularization of digested fragments is shown in a box. Bold face letters represent the cloned intronic sequences. The leftward intron spans nt 67 to 514, while the rightward intron shown underlined spans nt 521 to 884. The *Sac*II site as predicted is shown in a box within the rightward intron. Sequences flanking the introns represent core region selected from GAD coding portion and are shown in italicized uppercase letters.

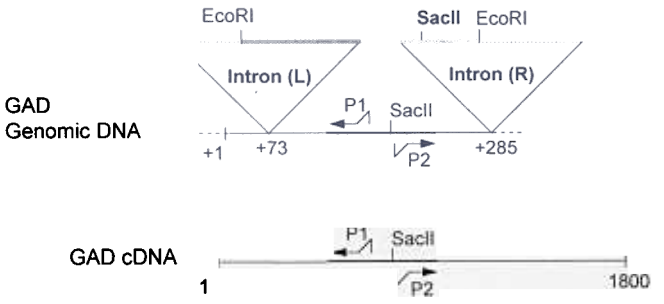

Figure 4. The genomic DNA organization of the PCR product was obtained upon inversion of the amplified sequences followed by its alignment with GAD cDNA sequence which is shown in the lower part of the figure. The leftward intron was present at +73 while the rightward intron being present at +285. Predicted *Sac*II site is shown in the right intron. (▬), Leftward intron; (▒▒▒), rightward intron; (——), extension of unknown intronic regions; (·····), extension of chromosomal DNA; (▬), core region; ( ), primers with *Hind*III add-on site.

were aligned in opposite orientation as compared to the original genomic DNA configuration. In order to deduce the correct orientation as expected in the genomic DNA, these sequences had to be inverted and are shown in figure 4. From this exercise it became evident that there were two introns present flanking the core region on either side. The left-side intron started at nt + 74 in the genome and the PCR product contained the last 446 bp from its 3′ end, while the other intron towards the right was encountered at nt + 287 and the construct included its first 363 bp. This intron contained the predicted *Sac*II site as shown in the sequence in figures 3 and 4.

Although we do not have the full length introns in our PCR product, the available partial sequences of both were put through the database TRANSFAC. Interestingly we were able to detect the presence of some putative transcription factor binding sites for AP2, SP1 and others (unpublished data). The sequence information obtained may now be used for designing new primers and walking along further upstream or downstream within the GAD gene.

Acknowledgment

We are grateful to Dr K-P Koller of Pharma Research, Hoechst AG, for help provided in DNA sequencing and database search.

References

Baekkeskov S, Aanstoot H-J, Christgau S, Reetz A, Solimena M, Cascalho M, Folli F, Olesen R-H and Camilli P-D 1990 Identification of the 64 K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase; *Nature (London)* **347** 151–156

Collins F S and Weissman S M 1984 Directional cloning of DNA fragments at a large distance from initial probe: A circularization method; *Proc. Natl. Acad. Sci. USA* **81** 6812–6816

Herrmann B G and Frischauf A-M 1987 Isolation of Genomic DNA; *Methods Enzymol.* **152** 180–183

Karlsen A E, Hagopian W A, Grubin C E, Dube S, Disteche C M, Adler D A, Barmeier H, Matewes S, Grant F J, Foster D and Lernmark A 1991 Cloning and primary structure of human islet isoform of glutamic acid decarboxylase from chromosome 10; *Proc. Natl. Acad. Sci. USA* **88** 8337–8341

Kaufman D L, Michael C S, Jide T, Forsthuber T, Ting G S P, Robinson P, Atkinson M A, Sercarz E E, Tobin A J and Lehmann P V 1993 Spontaneous loss of T-cell tolerance to glutamate decarboxylase in murine insulin-dependent diabetes; *Nature (London)* **366** 69–72

Maniatis T, Fritsch E F and Sambrook T 1989 Molecular cloning—A laboratory manual (New York: Cold Spring Harbor Lab)

Ochman H, Gerber A S and Hartl D L 1988 Genetic applications of inverse polymerase chain reactions; *Genetics* **120** 621–623

Silver J and Keerikatte V 1989 *J. Cell. Biochem.* (Suppl.) 13E, Abstract No. WH 239, 306

Sinclair C M and Aitken Robert 1995 PCR strategies for isolation of 5′ end of an immunoglobulin-encoding bovine cDNA; *Gene* **167** 285–289

Triglia T, Peterson M G and Kemp D J 1988 A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences; *Nucleic Acids. Res.* **16** 8186

MS received 7 February 1998; accepted 19 May 1998

Corresponding editor: SEYED E HASNAIN