

# Cloning and Sequencing of the 5' Region from Human Spinocerebellar Ataxia 2 Gene

Jorge Aguiar,<sup>1</sup> María Vázquez,<sup>1</sup> Julio Fernández,<sup>1</sup> Yssel Mendoza,<sup>1</sup> George Auburger<sup>2</sup>

<sup>1</sup>Division of Pharmaceutics. Center for Genetic Engineering and Biotechnology, PO Box 6162, Cubanacán, Havana 10600, Cuba. Phone: (53-7) 21 8008; Fax: (53-7) 21 8070; E-mail: farma3@cigb.edu.cu

<sup>2</sup>Division of Neurology. University Hospital, D-40225 Duesseldorf, Germany.

## ABSTRACT

A restriction fragment containing the 5' region of the spinocerebellar ataxia type 2 (SCA2) gene was subcloned from PAC65i22 into pBluescript and totally sequenced using an ABI 373A automated sequencer, in combination with the direct sequencing of the mentioned PAC genomic clone through the DOP-vector PCR method. A typical CpG island with several potential sites for Sp1 transcription factor has been found, where the promoter activity of the SCA2 gene can be situated, maybe like it occurs with the gene for Huntington disease.

**Keywords:** CpG island, disease, spinocerebellar ataxia 2, transgenic mouse

*Biotecnología Aplicada 1999;16:165-168*

## RESUMEN

Un fragmento de restricción que contiene la región 5' del gen de la ataxia espinocerebelosa 2 (SCA2), fue subclonado en el vector pBluescript a partir del PAC65i22, y secuenciado completamente mediante el empleo de un secuenciador automatizado ABI 373A, en combinación con la secuenciación directa del mencionado clon genómico de PAC mediante la técnica de PCR basada en el vector DOP. Se encontró un islote de CpG típico con varios sitios potenciales de unión al factor de transcripción Sp1, donde puede estar situada la actividad promotora del gen de la SCA2, como ocurre posiblemente para el gen de la enfermedad de Huntington.

**Palabras claves:** ataxia espinocerebelosa 2, enfermedad, islote de CpG, ratón transgénico

## Introduction

Spinocerebellar ataxia type 2 (SCA2) is an autosomal dominant genetic disease with marked anticipation. Anticipation is defined as the decrease of age for disease onset and the increase of disease severity in successive generations. The SCA2 gene was recently identified and cloned in different laboratories [1-3]. The abnormality at DNA level was elucidated as the intergenerational expansion of a CAG repeat in the coding region of this gene. Other seven different genes for neurodegenerative genetic diseases caused by (CAG)<sub>n</sub> tract expansions were also cloned in the past few years [4-10], and the expanded repeats were described as *dynamic* due to the meiotic and mitotic instability of the sequence (CAG)<sub>n</sub> [11].

In Holguín, Cuba, more than 1000 individuals are affected by this disease, and the reported frequency of 133:100 000 is the highest worldwide caused by a founder population effect [12].

The recent progress in the generation and characterization of transgenic mice expressing the genes containing expanded CAG repeats associated with four of eight disorders—X-linked recessive spinal and bulbar muscular atrophy (SBMA), spinocerebellar ataxia 1 (SCA1), Machado-Joseph disease (MJD/SCA3), and Huntington disease (HD) — is beginning to provide insight into the underlying mechanisms of these neurodegenerative disorders [13]. The best animal model of neurodegenerative disorders caused by a (CAG)<sub>n</sub> expansion is based on the expression of HD gene exon 1 carrying the mutated (CAG)<sub>n</sub> sequence under the physiological promoter of this gene [14].

To further use the SCA2 promoter and the truncated gene with the same purpose, the 5'-end of the

latter was sequenced and cloned. In this paper we report a 5.366 kb sequence of this region. We have noted that the 5' region of the SCA2 gene is composed of a typical CpG island, which could contain the putative promoter elements of this gene.

## Materials and Methods

### Cloning of the 5'-end from the SCA2 gene into pBluescript

PAC65i22 genomic clone [15] was digested with a combination of several restriction enzymes (Boehringer Mannheim, Germany). The resultant fingerprint was transferred to a Hybond N<sup>+</sup> membrane (Amersham, United Kingdom), and a Southern blot [16] was performed using the radioactive labelled probe CS (163 bp), which hybridizes approximately 800 bp upstream of the *SacI* site in exon 1 (Figure 1). A *HincII-SacI* fragment 3.3 kbp long was cloned into pBluescript SK(+) (Stratagene, USA) to obtain the clone 1BS.

The PAC65i22 vector was digested with *SacI* restriction endonuclease, and then ligated to itself. Ten recombinant clones were then digested with *EcoRI* (Boehringer Mannheim, Germany), and the fingerprint was transferred to a Hybond N<sup>+</sup> membrane. A Southern blot was performed with the radioactive labelled probe B (169 bp), which hybridizes with the region of the (CAG)<sub>n</sub> sequence within SCA2 gene exon 1 (Figure 1). An *EcoRI* fragment (2.4 kbp) including the rest of exon 1, from the *SacI* site to the *EcoRI* site in intron 1, was cloned into pBluescript SK(+) to produce the clone 3BS.

1. Pulst SM, Nechiporuck A, Nechiporuck T, Gispert S, Chen X-N, López-Cendes, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 1996; 14:269-76.

2. Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, et al. Cloning of the gene for spinocerebellar ataxia type 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* 1996; 14:285-91.

3. Sampei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, et al. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 1996;14:277-84.

4. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971-83.

5. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991;352:77-9.

6. Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolouysian atrophy (DRPLA). *Nat Genet* 1994;6:9-13.

7. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 1994;8:221-7.



Clone 1BS was digested with *SacI* and dephosphorylated with alkaline phosphatase (Boehringer Mannheim, Germany). A *SacI* fragment from the clone 3BS was cloned into the clone 1BS to produce the 5.366 kbp-long SCA2 gene 5' region (clone 9BS) (Figure 1).

**Sequence of the 5' region of the SCA2 gene**

The DOP-vector PCR method [17] was used to sequence from PAC65i22 genomic clone. This protocol is based on a two-step PCR: The first step was developed with the 6MW primer [18] and a sequence-specific remote primer. Then, a nested PCR was performed using 3 to 5 mL of the first reaction, the 6MW primer and a sequence-specific nested primer. Primers used to sequence from the 5' region of the SCA2 gene are listed in the Table. The products from several nested PCRs and DNA from the clone 9BS were sequenced using a Prism Ready Reaction DyeDeoxy Terminator Kit (Applied Biosystem, USA). Sequencing reactions were performed in a Perkin Elmer 9600 thermocycler (Perkin Elmer, USA): 25 cycles of 96 °C for 30 s, 50 °C for 15 s, and 60 °C for 4 min. Sequencing reactions were processed on an ABI automated sequencer 373A. Data was collected and analyzed on a Macintosh IIfx or Centris 650 using the ABI 373A Analysis and Data Collection software.

**Computer algorithms for promoter predictions**

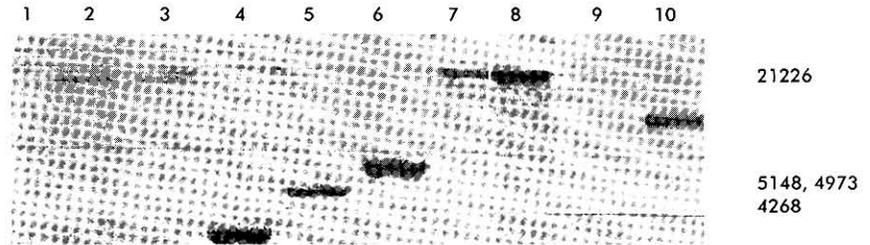
Computer algorithms such as TSSW and TSSG were used in order to predict a promoter sequence from the 5'-end of the SCA2 gene. Both TSSW and TSSG use the same underlying algorithm but TSSW is based on TRANSFAC, whereas TSSG is based on the promoter data file derived from Transcription Factor Database (TFD) [19].

**Results and Discussion**

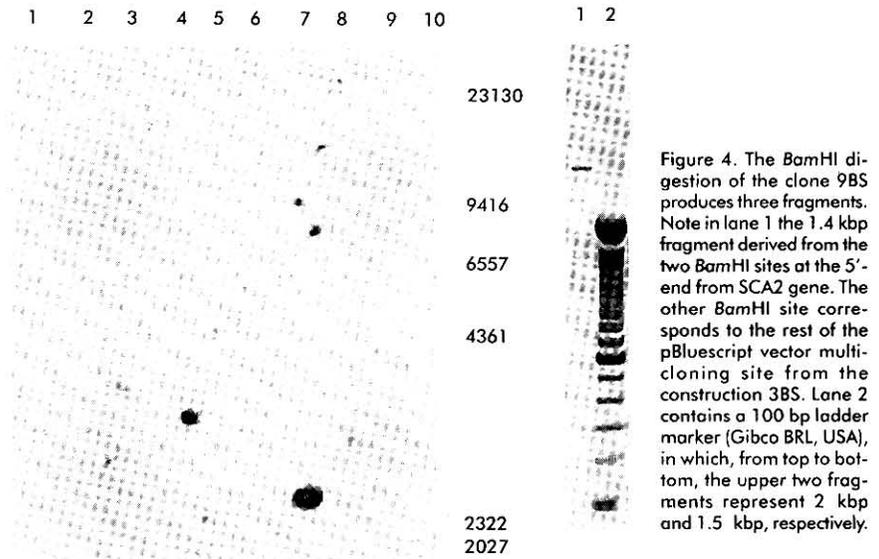
In the last few years, a (CAG)*n* instability was found for eight neurodegenerative disorders in the sequences of their genes. The complete nucleotide sequence of the 5.366 kbp fragment from the SCA2 gene was obtained. The cloning steps to produce the clone 9BS are shown in Figures 2, 3 and 4. The truncated proteins that include the expanded polyglutamine domain from huntingtin [14, 20-24], atrophin 1 [23, 24], androgen receptor protein (AR) [24], and ataxin 3 [24]

**Table. Primers used to obtain the total sequence of the 5'-end region from the SCA2 gene.**

Primer	Sequence
DAN1-reverse	5'-cccataccggctgcacg-3'
J1	5'-cgctgggtgctttctcg-3'
J2	5'-ccacatctgggtacctttcc-3'
J3	5'-atgatcattgtattctgcaatg-3'
J4	5'-agcagaggcaaaaggctcg-3'
J5	5'-cagaatacaatcttaoagaacaag-3'
J6	5'-aggcctgaagctgacatagc-3'
J1-forward	5'-cgagaagcaaccagcg-3'
DAN1	5'-cgtgcgagccggtgtatggg-3'
J7	5'-acggctcctctcgggtgctc-3'
J8	5'-aggctgtgggtggcgtttc-3'
J9	5'-aaatcgggtgggtgaggctg-3'
J10	5'-agactcattctcattgtggc-3'



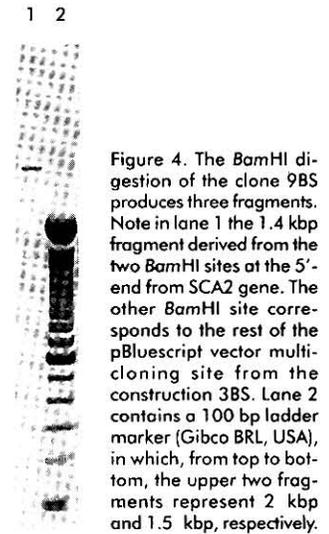
**Figure 2.** Southern blot with the radiolabelled probe CS on a fingerprint from the genomic clone PAC65i22. A combination of different restriction endonucleases with *SacI* is represented. Lane 1 to 10: *AccI*, *Clal*, *EcoRV*, *HincII*, *HindIII*, *PstI*, *Sall*, *SpeI*, *Tth1111* and *XbaI*. Note in lane 4 the signal on the *HincII-SacI* fragment corresponding approximately to the size of 3.3 kbp. This fragment was cloned to obtain the clone 1BS. Molecular weight marker III (Boehringer Mannheim, Germany) is represented to the right.



**Figure 3.** Southern blot using the radiolabelled probe B on a fingerprint from ten clones derived from PAC65i22 lacking *SacI* fragments, which were digested with *EcoRI*. The radioactive signal representing approximately 2.4 kbp is shown in lane 7. The fragment corresponding to this size was cloned *EcoRI* into pBluescript to produce the clone 3BS. Molecular weight marker II (Boehringer Mannheim, Germany) is represented to the right.

are able to produce the neurodegeneration observed in the specific neurons affected in these disorders. Major pathological steps of these related diseases involve a neurodegeneration caused by an induced apoptosis due to abnormal protein aggregates in the affected neurons [24, 25].

For the 5'-end of the SCA2 gene (Figure 1), we have found a CpG island (% G+C > 50 and CpG obs/exp > 0.6) with multiple G/C boxes that include 436 bp upstream of the transcription start site [1, 26], 162 bp from the 5'UTR [1, 26], the complete exon 1 carrying a CAG repeat with 22 units, and the beginning of intron 1. G/C boxes (composed of the sequence GGGCGG or its reverse complement CCGCCC) are located upstream of and downstream from the transcription start site of genes with CpG islands [27]. Sp1 transcription factor binds to some sequences that contain a G/C box, and promot-



**Figure 4.** The *BamHI* digestion of the clone 9BS produces three fragments. Note in lane 1 the 1.4 kbp fragment derived from the two *BamHI* sites at the 5'-end from SCA2 gene. The other *BamHI* site corresponds to the rest of the pBluescript vector multi-cloning site from the construction 3BS. Lane 2 contains a 100 bp ladder marker (Gibco BRL, USA), in which, from top to bottom, the upper two fragments represent 2 kbp and 1.5 kbp, respectively.

8. Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 1993;4:221-6.

9. Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the  $\alpha_1$ -voltage-dependent calcium channel. *Nat Genet* 1997;15: 62-9.

10. Davis G, Abbas N, Stevanin G, Duerr A, Yvert G, Cancel G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* 1997;17:65-70.

11. La Spada AR. Trinucleotide repeat instability: Genetic features and molecular mechanisms. *Brain Pathol* 1997;7:943-63.

ers based on Sp1 transcription factors often contain multiple binding sites [28].

Both TSSW and TSSG algorithms predicted the 120 bp stretch upstream of the *SacI* site in exon 1 as the strongest promoter into the 5.366 kbp region. This region contains four G/C boxes that possibly constitute target sites for the Sp1 transcription factor.

The best transgenic mouse model of neurodegeneration caused by an expansion of polyglutamine tracts was constructed using a fragment from the 5'-end of the HD gene from an ill patient [14]. We analyzed the complete coding region predicted for HD gene [4] and we noted that the structure of its 5'-end presents both the average % G+C > 50 and the CpG obs/exp ratio > 0.6, which is typical for CpG islands [27]. House-keeping genes and widely expressed genes have a 5' CpG island involved in their promoter activity and lack the

TATA box [27]. SCA2, like HD, shows a wide expression in different tissues. It is supposed that the clone 9BS contains the regulatory sequences involved in the expression of the SCA2 gene. This clone containing the (CAG)<sub>n</sub> tract expansion could be useful for the generation of a transgenic model for SCA2 disease.

In conclusion, further experiments using a reporter gene and deletion constructions are necessary to determine, in *in vitro* systems, which part of the 5' region of the SCA2 gene is responsible for the promoter activity.

## Acknowledgments

The authors thank Nicola Bergmann for her excellent technical assistance. Jorge Aguiar was supported by the Deutscher Akademischer Austauschdienst (DAAD) and by the Deutsche Forschung Gemeinschaft.

14. Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 1996;87:493-506.

15. Ioannou PA. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* 1994; 6:84-9.

16. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning. A laboratory manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 1989.

17. Wu C, Zhu S, Simpson S, de Jong PJ. DOP-vector PCR: a rapid isolation and sequencing of insert termini from PAC clones. *Nucl Acid Res* 1996;24:2614-5.

18. Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primer PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 1992; 13:718-25.

19. Fickett JW, Hatzigeorgiou AG. Eukaryotic promoter recognition. *Genome Res* 1997; 7:861-78.

20. Li SH, Li XJ. Aggregation of N-terminal huntingtin is dependent on the length of its glutamine repeats. *Hum Mol Genet* 1998; 7:777-82.

21. Cha JH, Koninski CM, Kerner JA, Alsdorf SA, Mangiarini L, Davies SW, *et al.* Altered brain neurotransmitter receptors in transgenic mice expressing a portion of an abnormal human huntingtin disease gene. *Proc Natl Acad Sci USA* 1998;95:6480-5.

22. Cooper JK, Schilling G, Peters MF, Herring WJ, Sharp AH, Kaminsky Z, *et al.* Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Hum Mol Genet* 1998;7:783-90.

23. Becher MW, Kotzok JA, Sharp AH, Davies SW, Bates GP, Price DL, *et al.* Intranuclear neuronal inclusions in Huntington's disease and dentatorubral and pallidolysian atrophy: correlation between the density of in-

clusions and IT15 CAG triplet repeat length. *Neurobiol Disease* 1998;4:387-97.

24. Wellington CL, Ellerby LM, Hackam AS, Margolis RL, Trifiro MA, Singaraja R, *et al.* Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *J Biol Chem* 1998; 273:9158-67.

25. Lunke A, Mandel JL. A cellular model that recapitulates major pathogenic steps of Huntington's disease. *Hum Mol Genet* 1998; 7:1355-61.

26. Nechiporuk T, Huynh DP, Figueroa K, Sahba S, Nechiporuk A, Pulst SM. The mouse SCA2 gene: cDNA sequence, alternative splicing and protein expression. *Hum Mol Genet* 1998;7:1301-9.

27. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol* 1987;196:261-82.

28. Kadanaga JT, Jones KA, Tjian R. Promoter-specific activation of RNA polymerase II transcription by Sp1. *TIBS* 1986;11:20-3.

Received in September, 1998. Accepted for publication in January, 1999.