2636–2640 Nucleic Acids Research, 1995, Vol. 23, No. 14

Editing of human α-galactosidase RNA resulting in a pyrimidine to purine conversion

F. J. Novo[1,2], A. Kruszewski[1,3,+], K. D. MacDermot[1], G. Goldspink[2] and D. C. Górecki[1,3,*]

[1]Department of Clinical Genetics, [2]Department of Anatomy and Developmental Biology, and [3]Molecular Neurobiology Unit, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, UK

Received April 20, 1995; Revised and Accepted June 12, 1995

ABSTRACT

During a study of the gene coding for α-galactosidase (EC 3.2.1.22), the lysosomal enzyme deficient in Fabry's disease, RT–PCR amplification of α-galactosidase mRNAs obtained from three different tissues isolated from males revealed a substantial number of clones with a U to A conversion at the nucleotide position 1187. Such a modification of the coding sequence would result in an amino acid substitution in the C-terminal region (Phe396Tyr) of the enzyme. Neither PCR analysis of the genomic sequence nor the RT–PCR amplification of RNA obtained by *in vitro* transcription of the wild-type cDNA showed this change in the sequence. Multiple genes, pseudogenes or allelic variants were excluded. Hence, we propose RNA editing as a mechanism responsible for this base change in the α-galactosidase RNA.

INTRODUCTION

Fabry's disease is an X-linked recessive disorder of the glycosphingolipid metabolism which results from the deficient activity of α-galactosidase (EC 3.2.1.22), a lysosomal enzyme (1). The gene has been assigned by physical mapping to Xq22.1 (2), the genomic sequence reveals 7 exons spanning ~10 kb (3) and its promoter shows the characteristics of a housekeeping gene. The processed mRNA of 1287 nucleotides (nt) codes for a protein of 429 amino acid (aa) residues, including a 31 aa signal peptide.

The sequence of the full-length cDNA encoding the mature enzyme was derived from clones obtained from a human fibroblast cDNA library (4,5). The most unusual feature found in all these clones was the absence of a 3′ untranslated sequence, with the polyadenylation signal included in the coding region and with the stop codon being followed by the poly-A tail (1). It is so far the only example of a human nuclear-encoded mRNA lacking a 3′ untranslated region.

Here we report a discrepancy between the genomic sequence and the sequence obtained by RT–PCR from various RNA sources, and we present data suggesting that this is the result of the editing of the α-galactosidase RNA. First described in the mitochondrial mRNA of trypanosomes (6), RNA editing has also been found in the mitochondrial tRNA of *A.castellanii* (7) and in mitochondrial (8) and cytoplasmic (9) mammalian tRNAs. Only three types of nuclear-encoded mammalian RNA have been found to undergo editing, namely those of human intestinal ApoB (10), several subunits of human and rat brain L-glutamate receptors (11–13) and the Wilms' tumor susceptibility gene (14), but the fact that Apo-B editing activity has been detected in non Apo-B producing cell lines (15) suggests that RNA editing could be a common biological process.

MATERIALS AND METHODS

Isolation of nucleic acids and first-strand synthesis

Genomic DNA was isolated from cerebellum and from blood of normal donors using standard procedures (16). Total RNA was isolated from muscle, cerebellum and a fibroblast cell line of male individuals not affected by Fabry's Disease, using an Ultraspec™ RNA isolation system (Biotecx). Isolation of mRNA from muscle and cerebellum was performed using a Fast Track mRNA isolation kit (Invitrogen).

In vitro transcription of an α-galactosidase cDNA clone (nucleotides 573–1296 of the cDNA sequence) was performed using the Riboprobe^{R} system (Promega) according to the manufacturer's protocol. Briefly, the cDNA cloned into a pCR™II vector (Invitrogen) was linearized with *Xba*I. After *in vitro* transcription using SP6 RNA polymerase (Stratagene), the samples were digested with *Rsa*I followed by RQ1 DNase I (Promega) in order to digest any plasmid DNA remaining in the sample, phenol–chloroform purified, precipitated and dissolved in nuclease-free water.

First-strand cDNA synthesis using the RNA samples from the different sources was carried out with Moloney Murine Leukemia Virus reverse transcriptase both by random hexamers and by oligo(dT) priming, with either a first-strand synthesis kit (Pharmacia) or enzyme and buffer from Gibco/BRL.

RT–PCR and amplification of genomic DNA

PCR reactions were carried out in 3 mM $MgCl_2$ using dNTPs at a final concentration of 0.5 mM with an annealing temperature of 60°C for 30–35 cycles and denaturation, annealing and extension steps of 1 min each, using either *Taq* polymerase (Perkin-Elmer)

* To whom correspondence should be addressed

+Permanent address: Department of Immunology, Central Clinical Hospital, Military School of Medicine, CSK WAM, Szaserów 128, 00-909 Warsaw, Poland

or a thermostable polymerase from *Thermus brockianus* (Dynazyme, Flowgen), reported to have a 2-fold higher fidelity than *Taq* polymerase.

Table 1. Nucleotide sequence and location of the primers used in this study

Primer	Sequence	5′-3′ Position (from start codon)
F7F	5′-GGCCACTTATCACTAGTT GC-3′	9749 to 9768 in genomic DNA
AK4	5′-AACATTTTAAAGTAAGT CTTTTAATGACATCTGC-3′	1296 to 1263 in cDNA
AK2	5′-GGTTAATCTTAAAAGCC CAG-3′	–59 to –40 in cDNA
AK3	5′-TTATGCTGTCCGGTCAC CG-3′	–23 to –5 in cDNA
AK8	5′-AATAGGACTGGCAGAA GCATTG-3′	574 to 595 in cDNA

The position of each primer is indicated relative to the first A in the initiation ATG.

The primers used in this study are summarized in Table 1. For amplification of genomic DNA we used primers AK4 and F7F. AK4 is complementary to the last 34 bases of the 3′ end of the α-galactosidase cDNA while F7F primes at the 3′ end of intron VI, so that only genomic DNA and not cDNA can be amplified. For the amplification of the first strand cDNAs we used primer pairs consisting of AK4 with one of the primers AK2, AK3 or AK8. Since all these primer pairs span at least 2 introns of the genomic α-galactosidase sequence, simple inspection of the PCR reaction products on a gel confirms that the products originate from first strand cDNA and not from any potentially contaminating genomic DNA. For the detection of possible pseudogene(s), amplification was performed under the same conditions with primers AK4–AK8, which prime in exons 7 and 4 respectively, and would yield a 2928 bp product on genomic DNA or a 723 bp fragment on cDNAs or intronless pseudogenes.

Cloning and sequencing of PCR products

For the cloning of PCR products 1 μl of each PCR reaction was ligated in a pCR™II vector (TA cloning kit, Invitrogen) following the manufacturer's protocol. From each transformation several clones containing an insert of the expected size were sequenced by the dideoxy chain-termination method with T7 polymerase (Pharmacia). Alternatively, PCR products were either gel-purified or column-purified (Quiaquick Spin, Qiagen) for direct sequencing by cycle sequencing (CircumVent, NEB) or by T7 polymerase after the conversion of the PCR products into single-stranded DNA with lambda exonuclease (PCR ssDNA template Prep kit, Pharmacia).

Restriction analysis with *Rsa*I

In vitro transcribed RNA was subjected to RT–PCR and the product was cloned as indicated above. Clones containing an insert of the expected size were digested overnight with *Rsa*I and the digests were electrophoresed in 2% MetaPhor agarose (Flowgen). A unique restriction pattern with an additional 200 bp band can be observed in clones with Adenine at position 1187 of the cDNA, but not in those containing Thymine at this position.

RESULTS

Amplification of the full-length α-galactosidase cDNA by RT–PCR from muscle and fibroblast revealed various mismatches when the sequence of the clones obtained was compared with the published sequence. However, since these base changes were present sporadically in a single clone out of several clones sequenced from each RT–PCR reaction, they were attributed to mistakes introduced either by the reverse transcriptase or by the polymerase. In contrast to this, however, a T to A transversion at nucleotide 1187 of the cDNA sequence, which creates a *Rsa*I restriction site, was present not only in several clones from the same PCR reaction but also in clones representing PCR reactions from different first-strand cDNAs. Specifically, five clones representing three different RT–PCR reactions (two from muscle RNA and one from fibroblast RNA) showed the normal sequence (T at position 1187), whereas 11 clones representing four different RT–PCR reactions (three from muscle RNA and one from fibroblast RNA) contained the T to A change. These results are summarized in Table 2.

Table 2. Number of clones/PCR reactions with T at position 1187 of the cDNA sequence (normal) and with the T1187A change (edited)

DNA/RNA source	No. of PCR reactions	No. of clones	Normal	Edited
Muscle RNA	5	8	2	6
Fibroblast RNA	2	8	3	5
Cerebellum RNA	2	14	10	4
Genomic DNA	9	Direct Seq	9	–

The source of RNA used is indicated.

This finding was confirmed when genomic DNA and RNA from the same tissue sample (human cerebellum) were isolated and first strand cDNA was prepared from the RNA. Amplification of the 3′ end of the α-galactosidase gene was performed using primer set AK4–F7F, specifically amplifying genomic DNA. Nine different PCR reactions were sequenced directly and all of them showed the normal reported sequence. At the same time, different first strand cDNA amplifications using primers AK4-AK8 yielded 10 clones with the normal sequence and four clones containing the T1187A change (without any other base changes in the rest of the sequence). Moreover, one of these RT–PCR reactions yielded four clones with the base change and one clone with the normal sequence (out of five clones sequenced), thus suggesting that the RNA used in the reverse transcription consists of a mixture of RNA species, some of them with the normal sequence and some bearing the base substitution. The product of another RT–PCR reaction sequenced directly also revealed this T1187A substitution.

In order to rule out the contribution of a possible pseudogene to the PCR product, we performed PCR with primers AK4–AK8 using genomic DNA as template. With this primer combination, an intronless pseudogene would have yielded a 723 bp fragment. This could not be detected in two unrelated normal males. In contrast, we could obtain low levels of the 2928 bp product expected when these primers amplify genomic DNA from exon 4 to exon 7 (results not shown).

As a control to exclude that the base change was introduced during any of the steps of the RT–PCR reaction, we used RNA transcribed *in vitro* from a cDNA clone containing the normal

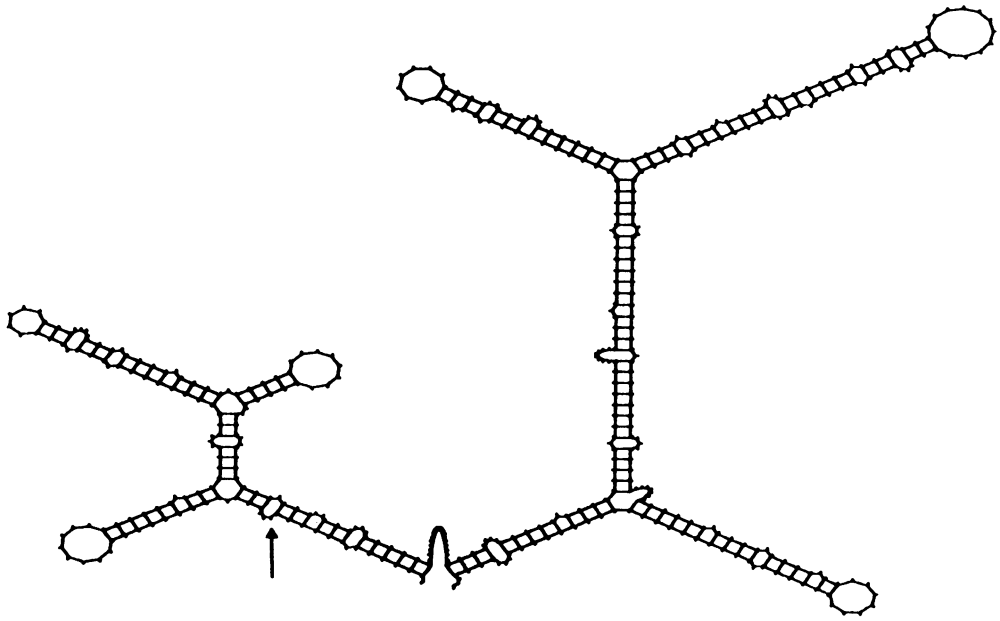

Figure 1. Putative secondary structure of the α-galactosidase primary transcript from nucleotide 1173 of the cDNA sequence and including 330 nt of 3′ flanking sequence. Calculated using FOLDRNA in the GCG package (36). The putative edited base is indicated with an arrow.

sequence at position 1187 (as described in the Methods section). This RNA was reverse transcribed into first strand cDNA using AK4 primer. The first strand obtained this way was amplified with primers AK4 and AK8 under the same conditions used in the other PCR reactions. The PCR product was cloned and 16 clones were analysed by digestion with *Rsa*I. The T to A transversion creates a *Rsa*I restriction site that results in a distinct restriction pattern with a 200 bp band in edited clones (see Fig. 2). None of the 16 clones studied showed the T1187A transversion.

DISCUSSION

There are a number of different mechanisms which could account for a site-specific difference between genomic and cDNA sequence. First of all, the presence of different copies of the gene, some of them bearing the mutation and some with the normal sequence, could explain the disparity found here. This could happen if α-galactosidase is a member of a family of highly homologous genes, or if allelic variants of the gene are present, or in the presence of a pseudogene in another location which is illegitimately transcribed. However, in any of these situations analysis of genomic DNA should reveal the differences in sequence, whereas our results clearly show a T in all the PCR reactions from genomic DNA. Furthermore, previous studies using *in situ* and Southern hybridization have established the occurrence of a single X-chromosomal gene for α-galactosidase and the absence of closely related sequences (1). We further excluded the possibility of illegitimate transcription of a pseudogene bearing the T to A change, since PCR priming off exonic sequences across exons 4 to 7 yielded no product from genomic DNA. As for the possibility of two different alleles, extensive mutation analysis of the human α-galactosidase gene in different individuals has never detected any base change at this position (17). Additionally, we always used samples from males to ensure that only one copy of this X-linked gene was present.

A second possibility is that the base change found here is an artifact introduced either by the reverse transcriptase or the DNA polymerase. However, under high fidelity conditions like the ones used in this study an error rate for *Taq* polymerase of 1/50 000 is probably a good estimate, so that the expected error frequency after 35 cycles of amplification is 1 in 2857 nt (18). With the *Tbr* polymerase used in this study this would become of the order of 1 in 6000, so that the possibility of the same base change being consistently introduced at the same position in different PCR reactions can be discounted. In reverse transcription the occurrence of a mismatch at one particular point of a sequence due to the formation of stem–loop structures in the RNA molecule is theoretically possible. However, such an event has never been reported and in any case the high incidence here is far above the reported error rate for the Moloney-MLV reverse transcriptase used, of between 1/3470 and 1/7372 as calculated using homopolymer templates (19). Moreover, when we used *in vitro* transcribed RNA as a template of the RT–PCR reaction, no base change was introduced at this nucleotide with respect to the normal sequence, indicating that neither the reverse transcriptase nor the polymerase are responsible for the T to A transversion found in such a significant number of RT–PCR reactions.

The observation that we have made is by the methods which in the other cases have been sufficient to deduce that RNA editing occurs. Here, this will be editing of the α-galactosidase RNA resulting in an adenine in the edited RNA instead of the uracil expected from the genomic sequence. The nature and the process of RNA editing is not fully understood and it is thought that different mechanisms are involved in different cases. However, the formation of a stem–loop seems to be a common feature underlying the editing in Apo B and L-glutamate receptor subunits RNAs. In the case of ApoB, the editing activity has been characterised as a cytidine deaminase (20,21) and various sequence requirements upstream and downstream of the edited region have been identified (22–24).

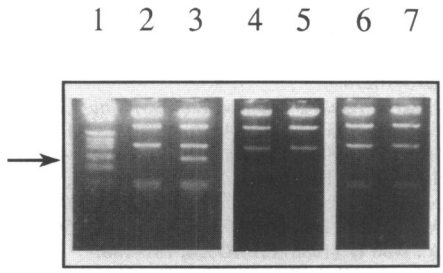

Figure 2. Restriction analysis of normal and edited clones with *Rsa*I. The T1187A transversion creates a *Rsa*I restriction site that results in a distinct restriction pattern with a 200 bp band in edited clones (arrow). Lane 1, 1 kb ladder (Gibco-BRL); lane 2, control clone (unedited); lane 3, control clone (edited); lanes 4–7, clones obtained from the *in vitro* transcription control experiment (see text for details).

The editing site in Apo B mRNA is included in a highly conserved stem–loop that binds to a 60 kDa protein (25). As for the L-glutamate receptor subunits, the editing seems to occur before the splicing of the RNA, since a specific intronic sequence downstream to the edited site is necessary for the editing to take place (26,27). This intron contains a 10 nt sequence complementary to the editing site (26), and the region between them includes an inverted repeat that can fold into a structure consisting of three helical elements (27). Recently, it has been shown that editing of GluR-B results from the conversion of adenosine to inosine by enzymatic deamination, and that several stem–loops close to the editing site are necessary for the targeting of the enzyme (28). When we analyzed the genomic sequence of α-galactosidase from intron VI and downstream, covering 1170 bp of 3′ flanking sequence, we found that several potential stem–loops are formed in this region, and the energy minimization algorithm of Zucker (29) also predicted the formation of stem–loops around the putative editing site when the RNA was folded (Fig. 1).

Of a total of nine RT–PCR reactions, five contained the T1187A base change and three did not, whereas one of the reactions resulted both in clones containing the mutation and in clones with the normal sequence. This clearly suggests that only a subpopulation of α-galactosidase RNA molecules undergo editing, and is in keeping with the fact that the frequency of editing of the human kainate receptors and of the Wilms' tumor susceptibility gene show tissue and/or age-related differences in the percentage edited (12,14). Also, the base change could be verified using RNA from three different sources (human muscle, human cerebellum and a human fibroblast cell line), indicating that the editing activity is present in a wide variety of tissues. This is in agreement with data suggesting that Apo B editing activity is found not only in enterocytes but also in cell types which do not produce Apo B (15).

The only example of RNA editing responsible for a U to A base change like the one reported here is the mitochondrial tRNAs of *A.castellanii*, where the 3′ half of the acceptor arm seems to act as an internal guide for the editing of the 5′ half (7). However, unconventional editing at a U site resulting in a pyrimidine to purine conversion has recently been described in man, namely in the RNA encoding a human glutamate receptor (GluR7), but this involves U to G, not U to A as here (13). The base change described here cannot be accounted for by deamination reactions such as those implicated in other instances of mammalian nuclear RNA editing. Rather, it must involve some form of base replacement in the RNA through as yet unknown mechanisms.

Whereas editing of Apo B, L-glutamate receptor subunits and the Wilms' tumor susceptibility gene introduces significant changes of functional relevance, in the case of the α-galactosidase the edited RNA will code for a tyrosine instead of a phenylalanine. This amino acid change is apparently conservative, although in some cases it may result in significant functional alteration of the protein product involved. For instance, a T to A transversion in the glucocerebrosidase coding sequence causing a Phe to Tyr change in the enzyme sequence resulted in Gaucher disease in one case (30). Similarly, a unique feature of the optic morphology mutant *Om(1D)* of *Drosophila ananassae* is that the phenylalanine in helix 3 of the BarH1 gene product, conserved in all homeodomains so far examined, is replaced by a tyrosine residue (31). Also, substitution of tyrosine for phenylalanine in the glycine receptor α1-subunit alters ligand discrimination and increases the efficacy of amino acid agonists (32). Since the crystal structure of α-galactosidase is not available, prediction of the conformational changes introduced by the amino acid conversion described here is difficult. In this regard, we have used several computer programs for the prediction of secondary structure and hydropathy of proteins (33–35) and could only observe a slight increase in the hydrophylicity of the molecule bearing the Phe396Tyr change. Further, computer analysis and sequence comparison of the region of the α-galactosidase protein encompassing the edited site did not reveal any homology with previously characterized protein domains. The elucidation of the effects of this amino acid change on the biological properties of α-galactosidase requires further study.

ACKNOWLEDGEMENTS

The authors wish to thank Professor Eric A. Barnard and Dr Paul Simons for helpful suggestions. This work was supported by the Research Trust for Metabolic Diseases, the Anderson Fabry Research Trust and the Vaincre les Maladies Lysosomales.

REFERENCES

1 Desnick RJ, Bishop DF. (1989) In Scriver C.R., Beaudet A.L., Sly W.S., Valle D. (eds): *The Metabolic Basis of Inherited Disease*, 6th ed. McGraw-Hill, New York, pp. 1751–1780.
2 Vetrie D., Kendall E., Coffey A., Hassock S., Collins J., Todd C. *et al.* (1994) *Genomics*, **19**, 42–47.
3 Kornreich R., Desnick R.J., Bishop D.F. (1989) *Nucleic Acids Res.*, **17**, 3301–3302.
4 Bishop D.F., Calhoun D.H., Bernstein H.S., Quinn M., Hantzopoulos P., Desnick R.J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4859–4863.
5 Tsuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V., Stubbleflied B.K., *et al.* (1987) *Eur. J. Biochem.*, **165**, 275–80.
6 Benne R., Van Den Burgh J., Brakenhoff J.P.J., Sloof P., Van Boom J.H., Tromp M.C. (1986) *Cell*, **46**, 819–26.
7 Lonergan K.M., Gray M. (1993) *Science*, **259**, 812–816.
8 Janke A., Paabo S. (1993) *Nucleic Acids Res.*, **21**, 1523–1525.
9 Beier H., Lee M.C., Sekiya T., Kuchino Y., Nishimura S. (1992) *Nucleic Acids Res.*, **20**, 2679–2683.
10 Powell L.M., Wallis S.C., Pease R.J., Edwards Y.H., Knott T.J., Scott J. (1987) *Cell*, **50**, 831.
11 Sommer B., Kohler M., Sprengen R., Seeburg P.H. (1991) *Cell*, **67**, 11–19.
12 Nutt S.L. and Kamboj R.K. (1994) *NeuroReport*, **5**, 2625–2629.
13 Nutt S.L., Hoo K.H., Rampersad V., Deverill R.M., Elliot C.E., Fletcher E.J. *et al.* (1994) *Receptors and Channels*, **2**, 315–326.
14 Sharma P.M., Bowman M., Madden S.L., Rauscher III F.J., Sukumar S. (1994) *Genes Dev.*, **8**, 720–31.

15 Bostrom K., Garcia A., Poksay K.S., Johnson D.F., Lusis A.J., Innerarity T.L. (1990) *J. Biol. Chem.*, **265**, 22446–22452.
16 Sambrook J., Fritsch E.F., Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor University Press, Cold Spring Harbor.
17 Eng C.M. and Desnick R.J. (1994) *Human Mutation*, **3**, 103–111.
18 Eckert K.A. and Kunkel T.A. (1991) In McPherson M.J., Quirke P., Taylor G.R.: *PCR. A Practical Approach*. IRL Press, Oxford, p. 225.
19 Gerard G.F. (1986) *Focus*, **8**, 12.
20 Teng B., Burant C.F., Davidson N.O. (1993) *Science*, **260**, 1816–1819.
21 Hodges P., Scott J. (1992) *Trends Biochem. Sci.*, **17**, 77–81.
22 Driscoll D.M., Lakhe-Reddy S., Oleska L.M., Martinez D. (1993) *Mol. Cell. Biol.*, **13**, 7288–7294.
23 Shah R.R., Knott T.J., Legros J.E., Navaratnam N., Greeve J.C., Scott J. (1991) *J. Biol. Chem.*, **266**, 16301–16304.
24 Backus J.W., Smith H.C. (1992) *Nucleic Acids Res.*, **20**, 6007–6014.
25 Scott J. (1994) In Latchman DS (ed.): From *Genetics to Gene Therapy*. Bioscientific Publisher Ltd., Oxford, pp. 1–23.
26 Higuchi M., Single F.N., Kohler M., Sommer B., Sprengel R., Seeburg P.H. (1992) *Cell*, **75**, 1361–1370.
27 Egebjerg J., Kukekov V., Heinemann S.F. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10270–10274.
28 Yang J.-H., Sklar P., Axel R., Maniatis T. (1995) *Nature*, **374**, 77–81.
29 Zuker M, Stiegler P. (1981) *Nucleic Acids Res.*, **9**, 133–148.
30 Beutler E, Gelbart T. (1990). *Ann. Hum. Genet.*, **54**, 149–153.
31 Koyima T., Ishimaru S., Higashijima S.-I., Takayama E., Akimaru H., Sone M. *et al.* (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4343–4347.
32 Schmieden V., Kuhse J., Betz H. (1993) *Science*, **262**, 256–258.
33 Jameson B.A., Wolf H. (1988) *CABIOS*, **4**, 181–186.
34 Gribskov M., Burgess R.R., Devereux J. (1986) *Nucleic Acids Res.*, **14**, 327–334.
35 Rost B., Sander C. (1994) *Proteins*, **19**, 55–72.
36 Devereux J., Haeberli P. and Smithies O. (1994) *Nucleic Acids Res.*, **12**, 387–395.