RESEARCH COMMUNICATION

Neuropathy target esterase and a homologous *Drosophila* neurodegeneration-associated mutant protein contain a novel domain conserved from bacteria to man

Michael J. LUSH*, Yong LI*, David J. READ*, Anthony C. WILLIS† and Paul GLYNN*[1]

*MRC Toxicology Unit, University of Leicester, Leicester LE1 9HN, U.K., and †MRC Immunochemistry Unit, University of Oxford, Oxford OX1 3QU, U.K.

The N-terminal amino acid sequences of proteolytic fragments of neuropathy target esterase (NTE), covalently labelled on its active-site serine by a biotinylated organophosphorus ester, were determined and used to deduce the location of this serine residue and to initiate cloning of its cDNA. A putative NTE clone, isolated from a human foetal brain cDNA library, encoded a 1327 residue polypeptide with no homology to any known serine esterases or proteases. The active-site serine of NTE (Ser-966) lay in the centre of a predicted hydrophobic helix within a 200-amino-acid C-terminal domain with marked similarity to conceptual proteins in bacteria, yeast and nematodes; these proteins may comprise a novel family of potential serine hydrolases. The Swiss Cheese protein which, when mutated, leads to widespread cell death in *Drosophila* brain [Kretzschmar, Hasan, Sharma, Heisenberg and Benzer (1997) J. Neurosci. **17**, 7425–7432], was strikingly homologous to NTE, suggesting that genetically altered NTE may be involved in human neurodegenerative disease.

INTRODUCTION

Organophosphorus esters (OPs) have long been recognized as relatively specific irreversible inhibitors of serine proteinases and esterases by virtue of their covalent reaction with the serine residue at the active site of these enzymes [1,2]. Various OPs can cause two human neurotoxicity syndromes by reaction with different esterases [3]. Inhibition of acetylcholinesterase by OPs results in acute, often fatal, toxicity. Covalent modification of neuropathy target esterase (NTE) by certain OPs leads, after a delay of several days, to a degeneration of long axons in the spinal cord and peripheral nerves [4,5]. Induction of this neurodegenerative process by selective modification of a single protein suggests that NTE, which in the adult nervous system is confined to neurons [6], may play key roles in neuronal and axonal maintenance; thus study of NTE is of fundamental biological interest.

NTE was originally identified and defined by its ability to both bind a radiolabelled OP and hydrolyse a (non-physiological) ester substrate [7,8]. By the mid-1970s, pioneering work by Johnson showed that inhibition by OPs of NTE's esteratic activity *per se* did not itself initiate neuropathy; the latter required a rapid secondary intramolecular change in the neutrally charged covalently bound OP, which left a negatively charged form of the OP still attached to the active-site serine of NTE [9]. Subsequent progress in this area was hampered by difficulty in isolating NTE from brain tissue. This was finally achieved by the synthesis of S9B [1-(saligenin cyclic phospho)-9-biotinyldiaminononane], a biotinylated OP that could be used for selective labelling of the active-site serine of NTE, and affinity purification by avidin–Sepharose chromatography [10]. We now report the use of S9B to obtain information on the location of the active-site serine and to initiate cDNA cloning of NTE.

EXPERIMENTAL PROCEDURES

Protein biochemistry

NTE, specifically biotinylated at its active-site serine residue by the S9B reagent, was isolated from pig brain microsomal membrane preparations and was digested with *Staphylococcus aureus* endoproteinase Glu-C, as described previously [10]. Digests were run on Tricine/SDS/PAGE gels (16.5%) [11], electroblotted on to Immobilon-P membranes and the N-terminal sequences of individual peptide bands were determined by automated Edman degradation. In addition, some excised peptide bands were digested secondarily with trypsin, the resulting digests fractionated by HPLC and the N-terminal sequence of resolved peptide peaks were determined [12]. To obtain a partial peptide map around the biotinylated active-site serine, endoproteinase Glu-C digests of S9B-labelled NTE were fractionated on avidin–Sepharose (cf. [10]) before running on Tricine-SDS/PAGE, as above. Synthetic peptides corresponding to three portions of deduced sequence from a putative human NTE cDNA clone (see Figure 1) were conjugated to keyhole limpet haemocyanin and used to raise rabbit antisera.

Molecular biology

Standard procedures detailed by Sambrook et al. [13] were used throughout. A human-expressed sequence tag clone (IMAGE Consortium id 204484), similar to one of the pig NTE peptides (DLGLPYFNVVTDI; 11/13 identical), was obtained from Research Genetics Inc., Huntsville, AL, U.S.A. and used to screen a λDR2 adult human brain cDNA library (kindly given by Dr Wilhelm Schwaeble, University of Leicester, U.K.), from which a clone with a 2.2 kb insert was isolated. The insert DNA was used to probe a Northern blot (Clontech) of human tissue

Abbreviations used: NTE, neuropathy target esterase; OP, organophosphorus ester; PKA, cyclic AMP-dependent protein kinase; S9B, 1-(saligenin cyclic phospho)-9-biotinyldiaminononane.

[1] To whom correspondence should be addressed (e-mail pg8@le.ac.uk).

The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL Data Bank with the accession number AJ004832.

poly(A)$^+$ (RNA) and hybridized to a broad band of 4.5–5.0 kb in brain, placenta, kidney and skeletal muscle, suggesting that this cDNA represented a truncated clone (results not shown). Additional 5′-cDNA sequence was obtained by two approaches. First, adult human brain cDNA was used as a template for PCR amplification of the 5′-cDNA ends (Clontech Marathon cDNA amplification kit; Clontech, Heidelberg, Germany); the data obtained (not shown) indicated that the 5′-end of NTE may be encoded by at least two different transcripts. In the second approach, a λZAP human foetal brain cDNA library (from W. Schwaeble) was screened with the original 2.2 kb insert DNA and four new clones were isolated; partial sequencing of these supported the view that isoforms of NTE with some variations in the 5′-coding sequence may exist. One clone, designated D16, with a 4.4 kb insert was fully sequenced (ABI automatic sequencer) and included a polyadenylation signal and a poly(A)$^+$-tail. The D16 sequence was translated and compared with protein sequences on EMBL and SwissProt data bases by BLAST searching [14]. Deduced polypeptide sequences of NTE and related proteins were analysed using PROSITE, O-glycbase [15] and TMpred [16] programmes.

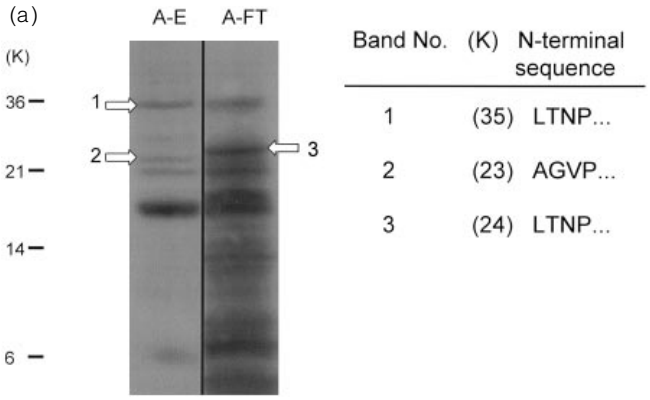
RESULTS AND DISCUSSION

Proteolytic digests of purified S9B-labelled NTE were fractionated and N-terminal sequences were determined for a total of 11 peptide fragments. A human-expressed sequence tag homologous to one peptide sequence was used to initiate the process of cloning NTE (see the Experimental procedures section). This culminated in the isolation, from a foetal human brain cDNA library, of a clone designated D16. Translation of the D16 clone from the first methionine predicts a 1327 amino acid polypeptide of 146 kDa with multiple potential sites for N- and O-linked

```
1    MEAPLQTGMV LGVMIGAGVA VVVTAVLILL VVRRLRVPKT PAPDGPRYRF RKRDKVLFYG
61   RKIMRKVSQS TSSLVDTSVS ATSRPRMRKK LKMLNIAKKI LRIQKETPTL QRKEPPPAVL
121  EADLTEGDLA NSHLPSEVLY MLKNVRVLGH FEKPLFLELC RHMVFQRLGQ GDYVFRPGQP
181  DASIYVVQDG LLELCLPGPD GKECVVKEVV PGDSVNSLLS ILDVITGHQH PQRTVSARAA
241  RDSTVLRLPV EAFSAVFTKY PESLVRVVQI IMVRLQRVTF LALHNYLGLT NELFSHEIQP
301  LRLFPSPGLP TRTSPVRGSK RMVSTSATDE PRETPGRPPD PTGAPLPGPT GDPVKPTSLE
361  TPSAPLLSRC VSMPGDISGL QGGPRSDFDM AYERGRISVS LQEEASGGSL AAPARTPTQE
421  PREQPAGACE YSYCEDESAT GGCPFGPYQG RQTSSIFEAA KQELAKLMRI EDPSLLNSRV
481  LLHHAKAGTI IARQGDQDVS LHFVLWGCLH VYQRMIDKAE DVCLFVAQPG ELVGQLAVLT
541  GEPLIFTLRA QRDCTFLRIS KSDFYEIMRA QPSVVLSAAH TVAARMSPFV RQMDFAIDWT
601  AVEAGRALYR QGDRSDCTYI VLNGRLRSVI QRGSGKKELV GEYGRGDLIG VVEALTRQPR
661  ATTVHAVRDT ELAKLPEGTL GHIKRRYPQV VTRLIHLLSQ KILGNLQQLQ GPFPAGSGLG
721  VPPHSELTNP ASNLATVAIL PVCAEVPMVA FTLELQHALQ AIGPTLLLNS DIIRARLGAS
781  ALDSIQEFRL SGWLAQQEDA HRIVLYQTDA SLTPWTVRCL RQADCILIVG LGDQEPTLGQ
841  LEQMLENTAV RALKQLVLLH REEGAGPTRT VEWLNMRSWC SGHLHLRCPR RLFSRRSPAK
901  LHELYEKVFS RRADRHSDFS RLARVLTGNT IALVLGGGGA RGCSHIGVLK ALEEAGVPVD
961  LVGGTSIGSF IGALYAEERS ASRTKQRARE WAKSMTSVLE PVLDLTYPVT SMFTGSAFNR
1021 SIHRVFQDKQ IEDLWLPYFN VTTDITASAM RVHKDGSLWR YVRASMTLSG YLPPLCDPKD
1081 GHLLMDGGYI NNLPADIARS MGAKTVIAID VGSQDETDLS TYGDSLSGWW LLWKRLNPWA
1141 DKVKVPDMAE IQSRLAYVSC VRQLEVVKSS SYCEYLRPPI DCFKTMDFGK FDQIYDVGYQ
1201 YGKAVFGGWS RGNVIEKMLT DRRSTDLNES RRADVLAFPS SGFTDLAEIV SRIEPPTSYV
1261 SDGCADGEES DCLTEYEEDA GPDCSRDEGG SPEGASPSTA SEMEEEKSIL RQRRCLPQEP
1321 PGSATDA
```

Figure 1 Polypeptide sequence of human NTE from translation of D16 cDNA

Positions of the N-terminal sequences of the 11 pig NTE peptides determined by Edman degradation are shown by single underlines with single residue differences shown above in lower-case letters. The position of the putative active-site serine residue (Ser-966) is marked with a diamond (◆; see Figure 2). Further evidence that this translation encoded NTE was obtained when synthetic peptides corresponding to three portions of the predicted human NTE sequence (shown by broken underlines) were used to raise rabbit antisera which reacted with S9B-labelled pig NTE on Western blots (results not shown). Putative transmembrane helices indicated by the TMpred programme are boxed, and potential sites for N-linked and O-linked glycosylation are marked with arrowheads and asterisks respectively. Note that the Edman sequencing data indicate that Asn-1040 in pig NTE does not bear a carbohydrate group.



Theoretical Fragment	Size (aa)	Size K	Putative identity with Band No.
LTNP → DLWL	(306)	34	1
LTNP → AGVP	(228)	25	3
AGVP → VVKS	(211)	23.5	2

Figure 2 Peptide mapping around the active-site serine residue of S9B-labelled NTE

(**a**) Blot from Tricine SDS/PAGE showing biotinylated (avidin-eluate; A-E) and putative non-biotinylated (avidin flow-through; A-FT) pig S9B-labelled NTE fragments, of which the N-terminal sequences of bands numbered 1–3 are indicated. The intense band at 16 kDa in the eluate fraction is avidin monomer. Only two NTE-related N-terminal sequences, LTNPAS (band 1, 35 kDa) and AGVPV (band 2, 23 kDa), were found for biotinylated peptides; LTNPAS was also the N-terminal sequence of a 24 kDa peptide (band 3) found in the avidin flow-through fraction. (**b**) Theoretical peptide map of the C-terminal half of human NTE (i.e., translation of D16 cDNA) cleaved by endoproteinase Glu-C to generate fragments with N-terminal sequences observed by Edman degradation of pig NTE peptides (cf. Figure 1).

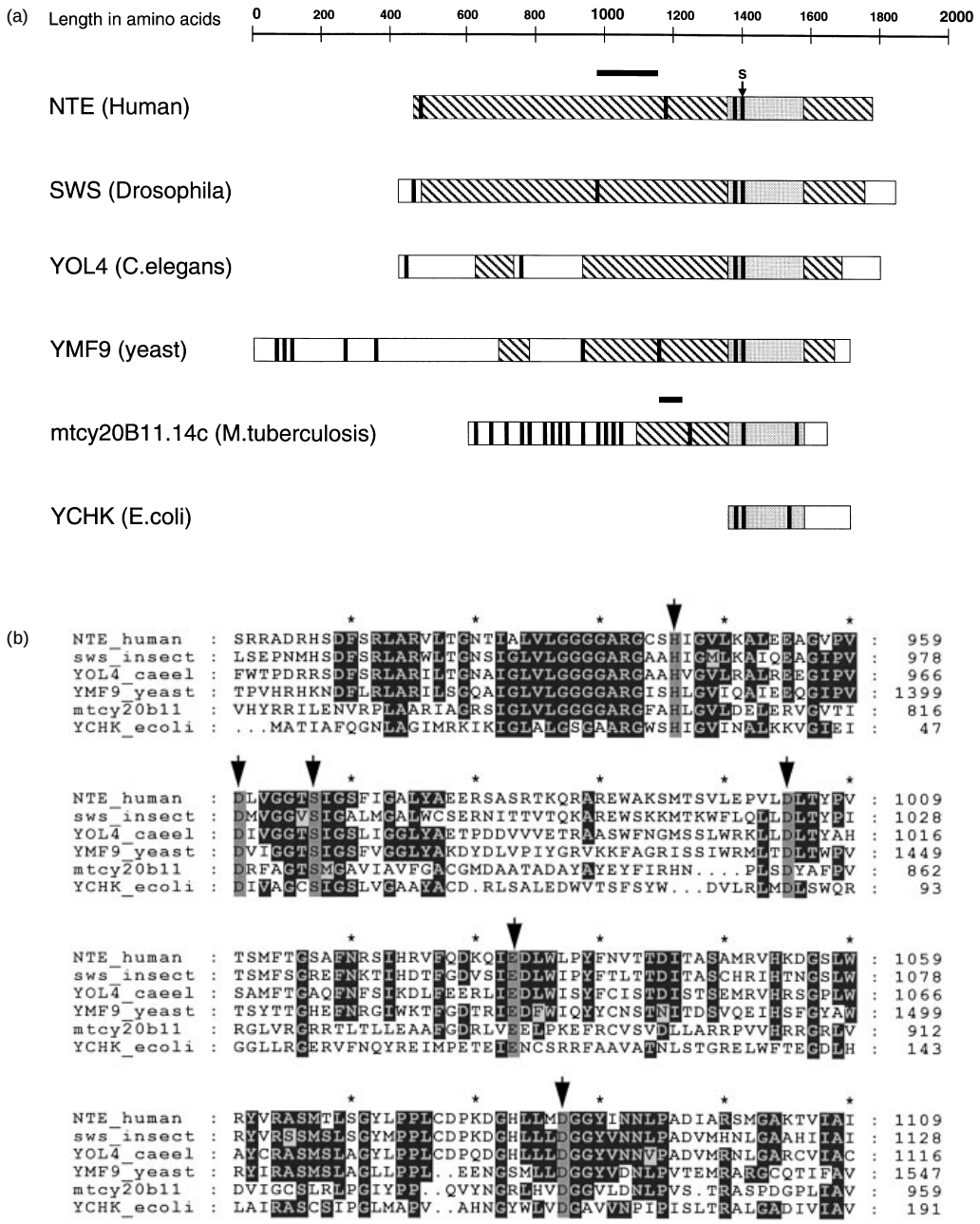


Figure 3 NTE is a member of a novel protein family

Representative conceptual proteins homologous to NTE are shown; apart from Sws, these proteins are predicted from gene-sequencing projects, and their functions are unknown. Database accession numbers are: NTE (AJ004832), Sws (Z97187), YOL4 (Q02331), YMF9 (Q04958), mtcy20B11.14c (Z95121) and YCHK (P37053). Other highly significant sequence similarities were found for a predicted protein in *Bacillus subtilis* (P54513), as well as additional proteins in *Caenorhabditis elegans* (Q21534; Q20023), yeast (P36165; S61643), *M. tuberculosis* (Q50733) and *E. coli* (P39407). Several human-expressed sequence tags encoding part of NTE have been mapped to chromosome 19p13.3 [24]. Hatched areas indicate regions with sequence homology to NTE. Thick black vertical bars indicate the positions of putative transmembrane helices analysed by the TMpred programme. Horizontal bar indicates the region of similarity to PKA regulatory subunit. The highly conserved C-terminal domain of about 200 residues is shown in grey. The position of the putative active-site serine (S) in NTE is indicated by an arrow. (**b**) Multiple sequence alignment for proteins in (**a**) within the conserved C-terminal domain. Residues identical in at least four of the proteins are shown white-on-black. Positions of the putative active-site serine and of conserved His, Asp and Glu residues are shown black-on-grey, and are indicated by an arrow.

glycosylation (Figure 1). On SDS/PAGE, OP-labelled NTE migrates with a molecular mass of about 155 kDa [10,17] and on Western blots the periodate-oxidized polypeptide reacts with labelled hydrazide reagents, indicating that NTE is a glycoprotein (P. Glynn, unpublished work). In addition, NTE is known to be firmly associated with microsomal membranes, and appears to be an integral membrane protein [18,19]. Analysis of the D16 clone translation sequence with the TMpred programme found four potential transmembrane domains, of which the most N-terminal domain was the strongest candidate (Figure 1).

To gain information on the position of the active-site serine residue, S9B-labelled NTE was digested with endoproteinase

Glu-C and fractionated on avidin–Sepharose to isolate fragments containing the biotinylated active site. Avidin eluate and flow-through fractions were then run on SDS/PAGE, electroblotted and the N-terminal sequence of excised peptide bands was determined (Figure 2a). These experimental data were compared with a theoretical peptide map of endoproteinase Glu-C cleavage of the C-terminal half of the translated D16 sequence, which would give rise to the observed N-terminal sequences for pig NTE fragments (Figure 2b). This comparison allows the conclusion that the active-site serine lies between two determined sequences, AGVPVDL and DLGLPYFNVVTDI. A total of nine serine residues are found in the D16 translated region between these points, but Ser-966 is particularly interesting, because it is located within the motif GXSXG, which is found at the active site of many serine esterases and proteases [20]. This is especially striking since database searching failed to reveal similarity between D16 and any known serine hydrolases and suggests that NTE is a new type of serine esterase.

Database searching revealed that NTE was similar to conceptual proteins from nematode, yeast and bacteria and also to the recently cloned Swiss Cheese (Sws) protein of *Drosophila* [21]. Six representative members of this group of proteins, which all contain a conserved C-terminal domain of about 200 amino acids, are shown in Figure 3(a). Upstream of this common C-terminal domain, the predicted proteins show increasingly extended similarity to NTE with evolutionary progression and, within this region, all the eukaryotic members of the family have a domain (equivalent to residues 487–744 of NTE), which resembles the regulatory subunit of cyclic AMP protein kinase (PKA) (Figure 3a). This similarity even partially extends to the mtcy20b11.14c polypeptide of *Mycobacterium tuberculosis*, which has a 54-residue sequence with homology to the cyclic-nucleotide-binding motif of PKA (Figure 3a). The N-terminal domains of the homologues in yeast and mycobacteria that contain multiple putative membrane-spanning domains are clearly different from those in the multicellular eukaryotes (Figure 3a); the N-terminal domain of mtcy20b11.14c is similar to antibiotic-resistance- and efflux-proteins [22].

The sequences of NTE and the Sws protein of *Drosophila* are 41 % identical over almost their full lengths (Figure 3a). The Sws protein is involved in the regulation of interactions between neurons and glia in the developing fly brain; mutations in the *sws* gene lead to glial hyperwrapping around neurons and axons followed by apoptosis of both cell types and age-dependent neurodegeneration [21]. NTE is expressed in neurons [6], and possibly may regulate their interaction with glia during development of the vertebrate nervous system. In turn, this suggests embryonic exposure to neuropathic OPs might cause a quite different form of human neurotoxicity to that seen in adults. It seems probable that NTE, a protein with serine esterase activity and a long-recognized association with a form of neurodegeneration, is the human homologue of the *sws* gene product. From the above, it is possible that genetically altered NTE is involved in human neurodegenerative disease.

Within its common C-terminal domain, human NTE is 29 % identical with the predicted YCHK protein of *Escherichia coli* (Figure 3a). It is notable that the motif GXSXG around the putative active-site serine (Ser-966) of NTE is retained throughout the six representative members of this protein family (Figure 3b). This serine, plus a completely conserved His-21 residue upstream, and one of the several conserved aspartate or glutamate residues indicated in Figure 3(b), could comprise a catalytic triad (cf. [23]). Thus by analogy with NTE, other members of this protein family could have serine esterase activity and be potential targets for certain OPs.

It is striking that Ser-966 of NTE lies in the centre of a predicted transmembrane domain (see Figure 1); TMpred analysis suggests that the same is true for the serine residue in the analogous position of the other members of this family (Figure 3a). Such predictions of membrane-spanning protein segments should be treated with extreme caution, but are intriguing in terms of the two-stage mechanism for initiation of OP-induced neuropathy [4]. In the first stage, the OP reacts with NTE to leave a neutrally charged portion of the molecule bound covalently to the active-site serine. In the second stage, a rapid intramolecular change occurs, which converts the bound OP into a negatively charged form still attached covalently to the serine residue. This second step is essential for initiating the (unknown) events which lead, much later, to axonal degeneration. If the serine residue of NTE that acquires this negative charge is, at least, in the centre of a relatively hydrophobic helical section of polypeptide sequence, then this event might be expected to have a drastic effect on the protein's local structure.

We thank K. Lilley and the Protein and Nucleic Acid Chemistry Laboratory (Leicester University) for protein and DNA sequencing services, W. Schwaeble for gifts of cDNA libraries, and Toxicology Unit colleagues for valuable discussion.

REFERENCES

1 Powers, J. C. and Harper, J. W. (1986) in Proteinase Inhibitors (Barrett, A. J. and Salvesen, G., eds.), pp. 55–151, Elsevier Science Publishers BV, Amsterdam
2 Aldridge, W. N. (1989) in Design of Enzyme Inhibitors as Drugs (Sandler, M. and Smith, H. J., eds.), pp. 294–313, Oxford University Press, Oxford
3 Moretto, A. and Johnson, M. K. (1987) in Toxicology of Pesticides: Experimental, Clinical and Regulatory Perspectives (Costa, L. G., Galli, C. L. and Murphy, S. D., eds.), pp. 33–48, Springer Verlag, New York
4 Johnson, M. K. (1987) Trends Pharmacol. Sci. **8**, 174–179
5 Lotti, M. (1992) Crit. Rev. Toxicol. **21**, 465–487
6 Glynn, P., Holton, J. L., Nolan, C. C., Read, D. J., Brown, L., Hubbard, A. and Cavanagh, J. B. (1998) Neuroscience (Oxford) **83**, 295–302
7 Johnson, M. K. (1969) Biochem. J. **111**, 487–495
8 Johnson, M. K. (1969) Biochem. J. **114**, 711–717
9 Johnson, M. K. (1974) J. Neurochem. **23**, 785–789
10 Glynn, P., Read, D. J., Guo, R., Wylie, S. and Johnson, M. K. (1994) Biochem. J. **301**, 551–556
11 Schagger, H. and Jagow, G. V. (1987) Anal. Biochem. **166**, 368–379
12 Fernandez, J., DeMott, M., Atherton, D. and Mische, S. M. (1992) Anal. Biochem. **201**, 255–264
13 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
14 Atschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Nucleic Acids Res. **25**, 3389–3402
15 Hansen, J. E., Lund, O., Rapacki, K. and Brunak, S. (1997) Nucleic Acids Res. **25**, 278–282
16 Hofmann, K. and Stofel, W. (1993) Biol. Chem. Hoppe-Seyler **34**, 166
17 Williams, D. G. and Johnson, M. K. (1981) Biochem. J. **199**, 323–333
18 Davis, C. S. and Richardson, R. J. (1987) Biochem. Pharmacol. **36**, 1393–1399
19 Ruffer-Turner, M. E., Read, D. J. and Johnson, M. K. (1992) J. Neurochem. **58**, 135–141
20 Brenner, S. (1988) Nature (London) **334**, 528–530
21 Kretzschmar, D., Hasan, G., Sharma, S., Heisenberg, M. and Benzer, S. (1997) J. Neurosci. **17**, 7425–7432
22 Philipp, W. J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., Bergh, S., Bloom, B. R., Jacobs, Jr., W. R. and Cole, S. T. (1996) Proc. Natl. Acad. Sci. U.S.A. **93**, 3132–3137
23 Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. and Silman, I. (1991) Science **253**, 872–879
24 Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E. et al. (1996) Science **274**, 540–546

Received 2 March 1998; accepted 18 March 1998