Characterization of the Human Dihydropyrimidinase-Related Protein 2 (DRP-2) Gene

Kenzo KITAMURA, Muneyuki TAKAYAMA, Naoki HAMAJIMA, Makoto NAKANISHI, Makoto SASAKI, Yuichi ENDO, Tadashi TAKEMOTO, Hiroshi KIMURA, Masayoshi IWAKI, and Masaru NONAKA

Departments of Biochemistry, Surgery I and Pediatrics, Nagoya City University Medical School, Mizuho-ku, Nagoya 467-8601, Japan, Department of Biochemistry, Fukushima Medical College, Fukushima-shi, Fukushima 960-1295, Japan, Department of Experimental Radiology, Shiga University of Medical Science, Otsu-shi, Shiga 520-2192, Japan, Department of Ophthalmology, Aichi Medical University, Nagakute-cho, Aichi 480-1195, Japan, and Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113-0033, Japan

(Received 21 July 1999; revised 1 September 1999)

Abstract

The genes within the dihydropyrimidinase-related protein (DRP) family were originally identified in humans by their homology to dihydropyrimidinase (DHP). Four members of this gene family, DRP-1, -2, -3 and -4, are expressed mainly in the fetal and neonatal brains of mammals and chickens, and have been implicated as intracellular signal transducers in the development of the nervous system. We isolated the human DRP-2 gene, and determined its transcriptional start site and exon/intron organization. The gene spanned more than 62 kb, and contained 14 exons with lengths ranging from 62 bp to 2606 bp. The transcriptional start site was determined by an RNase protection assay and 5' rapid amplification of cDNA ends (RACE), and a highly GC-rich promoter was identified that contained possible regulatory elements such as a TATA box, CAAT box and three GC boxes. Comparison of the phase and position of intron insertions within the human DRP-2 gene with those within DRP-1, DHP and two Caenorhabditis elegans DRP/DHP homologs, indicated that DRPs are more conserved in their exon/intron organization than DHP.

Key words: DRP-2; dihydropyrimidinase; neuronal development; exon/intron organization

1. Introduction

The DRP (dihydropyrimidinase-related protein) gene family is composed of four members, DRP-1, -2, -3 and -4. Although we identified these genes as homologs of human DHP (dihydropyrimidinase), rodent and chicken counterparts of human DRPs have been identified using various approaches and named accordingly. Chicken CRMP-62 (collapsin response mediator protein of relative molecular mass 62 kDa) was functionally cloned as a factor required for collapsin-induced growth cone collapse. Rat TOAD-64 (turned on after division, 64 kDa) was identified as a protein exclusively expressed by post-mitotic neurons. Mouse Ulip (unc-33-like phosphoprotein) was identified as a neuronal phosphoprotein whose expression is regulated during development. This gene shows sequence similarity to C. elegans unc-33, which is potentially involved in the control of neuritic outgrowth and axonal guidance. These rodent and chicken genes correspond with the human DRP genes as follows: DRP-1/CRMP-1/Ulip3, DRP-2/CRMP-2/Ulip2, DRP-3/CRMP-4/Ulip1, and DRP-4/CRMP-3/Ulip4. The overall amino acid sequence identity of the DRP family members ranges from 66% to 80%, suggesting that they have arisen by gene duplication from a common ancestor. DRPs are implicated in the developmental process of the nervous system of mammals, chicken and amphibia. However, extra-neuronal functions have also been predicted for DRPs since they are differentially expressed in various adult tissues, although at a lower level than observed in the developing brain. The restricted expression of DRP-3 in post-meiotic round spermatids has provided a very clear example of extra-neuronal expression of DRPs. Clarification of the biochemical function of DRPs may help in elucidating the physiological roles of these proteins.
these proteins undertake in neuronal and extra-neuronal cells.

The DRPs show 57% to 59% amino acid identity with DHP throughout their length, with the exception of about 90 residues at the C-terminus. DHP (EC 3.5.2.2) is the second enzyme in the uracil and thymine degradation pathway, and catalyzes the amide-hydrolysis of dihydrouracil and dihydrothymine.10 DHP is a homotetrameric metalloenzyme with a molecular mass of 215 kDa, a subunit mass of 54 kDa, and a zinc ion within each subunit.11 The complete primary structures of human1 and rat12 DHP have been determined from their cDNA sequences.

In this paper, we describe the structural analysis of the human DRP-2 gene, with the aim of gaining a better understanding of the evolutionary relationship between DRPs and DHP. We found that the exon/intron organization of the DRP-2 gene was identical to the human DRP-1 gene,13 but was significantly different from the human DHP gene.14

2. Materials and Methods

2.1. Isolation and characterization of human DRP-2 gene

Two human genomic phage libraries (Health Science Research Resources Bank, Osaka, Japan) were screened for clones containing exons from the human DRP-2 gene using cDNA probes 1 to 4 (Fig. 1). Walking probes were then prepared from the terminal regions of the isolated clones. The probes were labeled with [α-32P]dCTP (3000 Ci/mmol) using the Rediprime DNA labeling system (Amersham, Buckinghamshire, UK), and used to screen approximately 1 x 10⁶ plaques from each human genomic DNA library. Exon-containing DNA fragments were subcloned into the pGEM-3Zf (+) vector (Promega Corporation, Madison, WI), and at least five independent clones were sequenced to confirm the exon sequences. The sequence data was analyzed using the GeneExpress Mac 9.0 program (Software Development Co. LTD. Tokyo, Japan).

2.2. RNase protection analysis

Preliminary sequence analysis indicated that the 1.5-kb BamHI fragment of clone 1-7 (Fig. 2A) contained an exon encoding the 5' end of the human DRP-2 cDNA.1 PCR primers, RPA s-1 (5'-TTTGGATCCCGTACGCGATCCGCT-3') and RPA as-1 (5'-TTTGGATCCCGTACGCGATCCGCT-3') (Fig. 4B) were designed to amplify the 344-bp fragment that spanned the possible transcriptional initiation site. PCR amplification was carried out using the 1.5-kb BamHI fragment as a template, and the PCR product was subcloned into the BamHI site of the pGEM-3zf (+) vector. After linearization with HindIII, an RNA probe was synthesized using T7 RNA polymerase and [α-32P]UTP.

An RNase protection assay was performed using human fetal brain total RNA (Clontech Laboratories Inc., Palo Alto, CA) and a High-Speed Hybridization Ribonuclease Protection Assay kit (Ambion Inc., Austin, TX), following the protocol outlined by the manufacturer. Briefly, the probe was hybridized with 20 μg of...
Figure 2. Schematic representation of the genomic and cDNA structure of human DRP-2. A. Restriction map of the phage clones containing exons of the human DRP-2 gene (B, BamHI; H, HindIII; E, EcoRI). B. Exon/intron organization of the human DRP-2 gene. The exons are indicated by boxes. The distance between exons 3 and 4 is not known (> 20 kb). C. Structure of the human DRP-2 cDNA. The diagonally hatched region represents the protein-coding region.

2.3. Rapid amplification of 5’-end of human DRP-2 cDNA

Rapid amplification of the cDNA ends of DRP-2 was carried out using the 5’ RACE system (Gibco BRL Life Technologies Inc., Rockville, MD). Total RNA from human fetal brain (0.5 μg) was reverse-transcribed for 50 min at 42°C using Super Script II (Gibco BRL Life Technologies Inc., Rockville, MD) and the primer complementary to the +310 to +330 nucleotides of human DRP-2 cDNA (Genbank accession number, D78013; 5’-ATCAGAAGACGATCGCTCGTG-3’). After removing the excess primer with a spin column, the first strand reaction products were tailed with 2 mM dCTP and terminal deoxynucleotidyl transferase. Amplification was performed using an abridged anchor primer (5’-GGCCACGGCTCGACTAGTAC-3’) and a nested primer complementary to the +152 to +172 nucleotides of the sequence shown in Fig. 4B (5’-TGAAAAGGTGCAACCGCTTC-3’). The PCR product was subcloned into the pGEM-T Easy vector and sequenced using the -21M13 forward primer.

3. Results

3.1. Isolation of human DRP-2 gene

Six phage clones (1-7, 1-5, 1-10, 1-3, 1-6 and 2-1) were isolated from the human genomic library in the λEMBL3 arm using probes 1 to 4 from human DRP-2 cDNA (Fig. 1). These clones were aligned from the results of restriction mapping and Southern hybridization analyses (Fig. 2). Two gaps, between clones 1-10 and 1-3, and between clones 1-6 and 2-1, were identified. To fill the latter gap, a human genomic library in the λDASH2 arm was screened using the 4.3-kb BamHI arm were aligned from the results of restriction mapping and Southern hybridization analyses (Fig. 2). Two gaps, between clones 1-10 and 1-3, and between clones 1-6 and 2-1, were identified. To fill the latter gap, a human genomic library in the λDASH2 arm was screened using the 4.3-kb BamHI fragment of the 5’ end of the insert in clone 2-1 as a probe. This resulted in the isolation of clone 3-1. Walking was repeated once more, resulting in the isolation of a clone that overlapped with clones 1-6 and 3-1 (clone 4-1). Although we could not fill the gap between clones 1-10 and 1-3, struc-
Table 1. Exon/intron boundaries of the human DRP-2 gene.

| Exon number and (size) | cDNA position of exon$^a$ | 5’ splicing site$^b$ | Intron number and (size) | 3’ splicing site$^a$ |
|------------------------|---------------------------|---------------------|--------------------------|---------------------|
| 1 (320bp)              | 1-320                     | AGGTACACAG GTCTAGTCC | 1 (4kb)                  | AGGTTTACAG |
| 2 (89bp)               | 321-409                   | GTTGGACCCG GTAAAGGGAA | 2 (1.9kb)                | GCTATATTAG GCAATAGGAA |
| 3 (185bp)              | 410-594                   | ACTAGATATG GTAAAGAGCT | 3 (20kb)                 | TTCTGTGTGATG |
| 4 (165bp)              | 595-759                   | AAGTTTACAG GTAAAGGGCA | 4 (2.5kb)                | CGGCTACTAC GGTTAAAAGTC |
| 5 (812bp)              | 760-821                   | GAATTCCAG GTAAAGAATTT | 5 (0.5kb)                | TTCTGTGTGATG |
| 6 (812bp)              | 822-902                   | GATTCGGAGG GTAAAGGGCT | 6 (0.6kb)                | AACTATGGCA GATGGGCAAG |
| 7 (693bp)              | 903-971                   | ACCTGGACGG GTAAAGTTTC | 7 (7kb)                  | TGTTTGGGAG GCTGACCCAG |
| 8 (1212bp)             | 972-1092                  | GGAAAGGAGG GTAAAGGGCTG | 8 (9kb)                  | GCTGTGTGATG |
| 9 (157bp)              | 1093-1245                 | TGCTCTCTCT GTAAAGGCTG | 9 (0.4kb)                | TTCTGTGTGATG |
| 10 (1422bp)            | 1250-1391                 | CAGTGGTCTG GTAAAGGGCTG | 10 (3.2kb)               | TTCTGTGTGATG |
| 11 (1572bp)            | 1392-1562                 | ACACACACAC GTAAAGGCTG | 11 (4.6kb)               | TTCTGTGTGATG |
| 12 (1802bp)            | 1563-1742                 | AAGTGGACAG GTAAAGGTTT | 12 (0.8kb)               | TTCTGTGTGATG |
| 13 (1646bp)            | 1743-1908                 | ATTTTTTTCT GTAAAGGTTT | 13 (2.2kb)               | TTCTGTGTGATG |
| 14 (2603bp)            | 1909-4514                 | AGTTTTTCTG GTAAAGGTTT |                         | TTCTGTGTGATG |

$^a$ Nucleotides are numbered according to the cDNA sequence, starting with +1 for the first nucleotide of the transcriptional initiation site.

$^b$ Exon and intron sequences are shown in upper-case (10 bp of exon sequence) and lower-case (10 bp of intron sequence), respectively, and the intron residues adjoining the splice junction are shown in boldface.

The exon/intron organization of the human DRP-2 gene was compared with that of human DRP-1, human DHP, and two C. elegans homologs, R06C7.3 and C47E12.8 (EMBL accession numbers Z71266 and Z68882, respectively). As shown in Fig. 3, the position and phase of intron insertion were completely conserved between the DRP-2 and DRP-1 genes. Only one intron was common to all these five genes, three introns were shared by DRP-2, DRP-1, R06C7.3 and C47E12.8, and four introns were shared by DRP-2, DRP-1 and DHP.

3.3. Determination of transcriptional initiation site of human DRP-2 gene

We used two different approaches to determine the 5’ end of the human DRP-2 mRNA. In the first approach, we performed 5’ RACE using human fetal brain mRNA and an internal primer specific to the nucleotide sequence near the 5’-end of the cDNA clone. Three positive clones were identified using this method, and the longest clone extended 7 bp further upstream than the 5’-end of the cDNA clone. This nucleotide was tentatively assigned to the +1 position (Fig. 4B). In the second approach, we performed an RNase protection assay using the same human fetal brain mRNA. An RNA probe was synthesized by PCR amplifying the genomic DNA fragment using primers that supposedly spanned the transcriptional...
The clone that contained exon 1 (Clone 1-7; Fig. 4B). The 5' upstream region of the human DRP-2 gene was very GC rich, and one TATA-box, one CAAT-box and three GC-boxes were identified.

The dots indicate identity with the DRP-2 residues. The dashes indicate the gaps introduced to increase the identity. The position corrected sequence is shown here. The alignment of these amino acid sequences was performed using the CLUSTAL W program. 23

start site (−205 to +121, Fig. 4B). After hybridization with the human fetal brain total RNA, free RNA was digested with RNase T1, and the size of the protected RNA was determined after electrophoresis through a denatured polyacrylamide gel. As shown in Fig. 4A, two major bands of 131 bp and 127 bp were detected. These bands were estimated to have the 1.5-kb BamHI fragment of the phage clone that contained exon 1 (Clone 1-7; Fig. 4B). The 5' upstream region of the human DRP-2 gene was very GC rich, and one TATA-box, one CAAT-box and three GC-boxes were identified.

The nucleotide sequences of the protected RNA bands of 131 bp and 127 bp were determined using the 1.5-kb BamHI fragment. As shown in Fig. 4A, two major bands of 131 bp and 127 bp were detected. These bands were estimated to have the 1.5-kb BamHI fragment of the phage clone that contained exon 1 (Clone 1-7; Fig. 4B). The 5' upstream region of the human DRP-2 gene was very GC rich, and one TATA-box, one CAAT-box and three GC-boxes were identified.

The dots indicate identity with the DRP-2 residues. The dashes indicate the gaps introduced to increase the identity. The position corrected sequence is shown here. The alignment of these amino acid sequences was performed using the CLUSTAL W program. 23

4. Discussion

From an evolutionary viewpoint, it is interesting that the recent completion of C. elegans genome sequencing has indicated that C. elegans has only three genes (unc-33, R06C7.3 and C47E12.8) with significant sequence similarity to vertebrate DRP/DHP (four DRPs and DHP). Since the amino acid sequence identities of R06C7.3 and C47E12.8 to vertebrate DRP/DHP are between 46% and 59%, whereas the amino acid identities of unc-33 to DRP/DHP are between 35% to 37%, vertebrate DRP/DHP are considered to have diverged from a common ancestor similar to R06C7.3 and C47E12.8. The R06C7.3 and C47E12.8 share three positions and phases of intron insertion with the DRP-2 and DRP-1 genes, but none with the DHP gene except for the one position common to all these five genes (Fig. 3). Therefore, with respect to exon/intron organization, DRPs seem to be more conserved than DHP. However, with respect to conservation of amino acid residues critical for protein function, DHP seems to be more conserved than DRPs.

Figure 3. Comparison of the positions and phases of intron insertion in the DRP-2 gene, with those of the human DRP-1 and DHP genes, and C. elegans R06C7.3 and C47E12.8 genes (EMBL accession numbers Z71266 and Z68882, respectively). By cDNA cloning, we found a mistake in the assignment of exons in the deposited data of R06C7.3 (Takemoto et al. unpublished data), and the corrected sequence was used for the CLUSTAL W program. 23 The dots indicate identity with the DRP-2 residues. The dashes indicate the gaps introduced to increase the identity. The position and phase of the intron insertions are shown by triangles and numbers, respectively. The DHP residues involved in metal binding are indicated by the asterisks.

By guest
Downloaded from https://academic.oup.com/dnaresearch/article-abstract/6/5/291/353934 on 30 July 2018
Figure 4. The transcriptional initiation site of the human DRP-2 gene. A. RNA protection analysis. The transcriptional start site of the human DRP-2 gene was determined by RNase protection analysis using fetal brain RNA. Two major protected bands of 131 bp and 127 bp were detected. The four lanes on the left contain the sequence ladders used as size markers. The undigested probe is shown on the right. B. The nucleotide sequence around the transcriptional start site of the DRP-2 gene. The 5' end residue of the 5'RACE product is shown as +1. The two initiation sites determined by RNase protection analysis are indicated by the triangles. The residues in the first exon are shown in boldface, and the initiation codon (Met) is underlined. The TATA box, CAAT box and three GC boxes are indicated. The PCR primers used for RNase protection and 5'RACE analyses are indicated by the arrows.

Recently, it has been proposed that the amidohydrolases that are related to urease share the same active site architecture, including four histidine residues and one aspartic acid residue involved in metal-binding.\(^{15}\) Actual involvement of these residues in metal binding and enzyme activity has been shown in \textit{Bacillus stearothermophilus} D-hydantoinase by site-directed mutagenesis.\(^{16}\) Human\(^1\) and rat\(^2\) DHP, as well as R06C7.3 and C47E12.8, have retained all five of these metal-binding residues, whereas all reported DRP sequences to date have substitutions at one or more of these sites.\(^{1,2,7}\) Therefore, it is unlikely that DRPs have amidohydrolase activity, although four members of the DRP family form a heterotetramer that is similar to the homotetramer structure of DHP.\(^{17}\) The biochemical function of DRPs remains to be clarified.

Inconsistencies in the chromosomal localization of the human DRP-2 gene have been reported, since two different expressed sequence tags corresponding to DRP-2 have been assigned to two different chromosomes, \(^{8,18}\) and 21 (EMBL accession number, Z47338). Recently, the mouse DRP-2 (CRMP-2) gene has been mapped to the central region of chromosome 14 by interspecific backcross analysis, and by fluorescence in situ hybridization (FISH) analysis, and the corresponding human region (8p21) has been shown to harbor the human DRP-2 gene.\(^{19}\) We also performed FISH analysis for the DRP-2 gene, and confirmed that this gene resides at 8p21 (data not shown). The human DRP-1 gene has been located at 4p15–16.1, near the Huntington disease loci.\(^{13,20}\) and we have assigned the human DHP gene to 8q22.\(^{14}\) Overall, this data indicates that the DRP/DHP loci are widely dispersed throughout the human genome.

In the mouse, the wobbler lethal (\textit{wl}) mutation that causes various neurological and neuromuscular disorders has been mapped in the vicinity of the DRP-2 gene.\(^{21}\) In humans, DRP-2 has been associated with the neurofibrillary tangles observed in Alzheimer's patients,\(^{22}\) suggesting that DRP-2 may be involved in the neurofibrillary degeneration observed in this disease. The structural information on the human DRP-2 gene obtained in the present study may facilitate further analysis of a possible relationship between the DRP-2 gene and the \textit{wl} mutation or Alzheimer's disease.

\textbf{Acknowledgments:} We thank Drs. Chisato Namikawa-Yamada and Noriyuki Kuroda for their help-
ful comments and suggestions. This work is supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan (10558089) to M.N.

References

1. Hamajima, N., Matuda, K., Sakata, S., Tamaki, N., Sasaki, M., and Nonaka, M. 1996. A novel gene family defined by human dihydropyrimidinase and three related proteins with differential tissue distribution. *Gene*, **180**, 157–163.

2. Kato, Y., Hamajima, N., Inagaki, H. et al. 1998. Postmeiotic expression of the mouse dihydropyrimidinase-related protein 3 (DRP-3) gene during spermiogenesis, *Mol. Reprod. Dev.*, **51**, 105–111.

3. Goshima, Y., Nakamura, F., Strittmatter, P., and Strittmatter, S. M. 1995. Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33, *Nature*, **376**, 509–514.

4. Minturn, J. E., Geschwind, D. H., Fryer, H. J., and Hockfield, S. J. 1995. Early postmitotic neurons transiently express TOAD-64, a neural specific protein, *Comp. Neurol.*, **355**, 369–379.

5. Byk, T., Dobransky, T., Cifuentes-Diaz, C., and Sobel, A. 1996. Identification and molecular characterization of Unc-33-like phosphoprotein (Ulip), a putative mammalian homolog of the axonal guidance-associated unc-33 gene product, *J. Neurosci.*, **16**, 688–701.

6. Li, W., Herman, R. K., and Shaw, J. E. 1992. Analysis of the *Caenorhabditis elegans* axonal guidance and outgrowth gene unc-33, *Genetics*, **132**, 675–689.

7. Wang, L. H. and Strittmatter, S. M. 1996. A family of rat CRMP genes is differentially expressed in the nervous system, *J. Neurosci.*, **16**, 6197–6207.

8. Byk, T., Ozon, S., and Sobel, A. 1998. The Ulip family phosphoproteins—common and specific properties, *Eur. J. Biochem.*, **254**, 14–24.

9. Kamata, T., Daar, I. O., Subleski, M., Copeland, T., Kung, H. F., and Xu, R. H. 1998. Xenopus CRMP-2 is an early response gene to neural induction, *Brain Res. Mol. Brain Res.*, **57**, 201–210.

10. Fritzson, P. 1957. The catabolism of C14-labeled uracil, dihydouracil, and 3-uredopropionic acid in rat liver slices, *J. Biol. Chem.*, **226**, 223–228.

11. Kikugawa, M., Kaneko, M., Fujimoto-Sakata, S. et al. 1994. Purification, characterization and inhibition of dihydropyrimidinase from rat liver, *Eur. J. Biochem.*, **219**, 393–399.

12. Matsuda, K., Sakata, S., Kaneko, M. et al. 1996. Molecular cloning and sequencing of a cDNA encoding dihydropyrimidinase from the rat liver, *Biochim. Biophys. Acta*, **1307**, 140–144.

13. Torres, R. and Polymeropoulos, M. H. 1998. Genomic organization and localization of the human CRMP-1 gene, *DNA Res.*, **5**, 393–395.

14. Hamajima, N., Kowaki, M., Vreken, P. et al. 1998. Dihydropyrimidinase deficiency: Structural organization, chromosomal localization, and mutation analysis of the human dihydropyrimidinase gene, *Am. J. Hum. Genet.*, **63**, 717–726.

15. Holm, L. and Sander, C. 1997. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease, *Proteins*, **28**, 72–82.

16. Kim, G. J. and Kim, H. S. 1998. Identification of the structural similarity in the functionally related amidohydrolases acting on the cyclic amide ring, *Biochem. J.*, **330**, 295–302.

17. Wang, L. H. and Strittmatter, S. M. 1997. Brain CRMP forms heterotetramers similar to liver dihydropyrimidinase, *J. Neurochem.*, **69**, 2261–2269.

18. Koyama, K., Sudo, K., and Nakamura, Y. 1995. Isolation of 115 human chromosome 8-specific expressed-sequence tags by exon amplification, *Genomics*, **26**, 245–253.

19. Kamata, T., Subleski, M., Hara, Y. et al. 1998. Isolation and characterization of a bovine neural specific protein (CRMP-2) cDNA homologous to unc-33, a *C. elegans* gene implicated in axonal outgrowth and guidance, *Brain Res. Mol. Brain Res.*, **54**, 219–236.

20. Ishida, Y., Hadano, S., Nagayama, T., Tomiyasu, H., Wakasa, K., and Ikeda, J. 1994. Isolation and characterization of 21 novel expressed DNA sequences from the distal region of human chromosome 4p, *Genomics*, **22**, 302–312.

21. Bronson, R. T., Sweet, H. O., Spencer, C. A., and Davison, M. T. 1992. Genetic and age related models of neurodegeneration in mice: dystrophic axons, *J. Neurogenet.*, **8**, 71–83.

22. Yoshida, H., Watanabe, A., and Ihara, Y. 1998, Collapsin response mediator protein-2 is associated with neurofibrillary tangles in Alzheimer’s disease, *J. Biol. Chem.*, **273**, 9761–9768.

23. Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, **22**, 4673–80.
