Vol. 205, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1994 Pages 612-617

CHARACTERIZATION OF THE GENE FOR HUMAN NEUTROPHIL-ACTIVATING PEPTIDE 78 (ENA-78)

Max S. Corbett, Ina Schmitt[+], Olaf Riess[+], and Alfred Walz*

Theodor Kocher Institute, University of Bern, CH-3000 Bern 9, Switzerland

[+]Department of Molecular Human Genetics, Ruhr-University, Bochum, Germany

Received October 11, 1994

The genomic DNA for ENA-78 has been obtained from a human chromosome 4 flow-sorted cosmid library. Three out of 25'000 screened single colonies yielded the same 2.2-kB EcoRI ENA-78 gene fragment. A similar size fragment was observed on genomic southern blots, suggesting the presence of a single ENA-78 gene. The transcriptional start site was localized using a 5' RACE protocol on first strand cDNA prepared from stimulated alveolar type-II epithelial cell (A549) poly(A) mRNA. The ENA-78 gene contains four exons and three introns and the open reading frame of 342 nucleotides encodes for a protein of total 114 amino acids. The 5' flanking region contains potential binding sites for several nuclear factors such as AP-2, NF-κB, and interferon regulatory factor -1.

Epithelial-derived neutrophil-activating peptide 78 (ENA-78) was originally isolated and cloned from an IL-1β stimulated human pulmonary type-II-like epithelial cell line, A549 (1). ENA-78 is a secreted protein of 78 amino acids and belongs to the C-X-C family of chemotactic cytokines. Members of this family include IL-8 (2), GRO(α, β, γ) (3), NAP-2 (4), GCP-2 (5), platelet factor 4 (PF4) (6) and γIP-10 (7). With the exception of PF4 and γIP-10, these 8-10 kD proteins are potent activators of neutrophil functions, inducing chemotaxis, enzyme release, receptor upregulation and adhesion (8). The peptides differ markedly in their cellular origin and/or kinetic of induction suggesting different roles in pathophysiology. IL-8 and GROα are produced upon stimulation with IL-1 and TNFα by almost any type of cell. In contrast, ENA-78 shows a much more restricted appearance. It is predominantly produced by epithelial cells and to a lesser degree by monocytes and macrophages (9,10). The fact that high levels of ENA-78 were detected in patients with rheumatoid arthritis (11) and that bovine ENA was immunohistochemically identified in the hyperplastic type-II alveolar epithelial cells of lung sections (12) suggests a role for ENA-78 in the genesis of pulmonary and other inflammations. Therefore, it is of interest to elucidate the regulatory events leading to ENA-78 gene expression.

*To whom correspondence should be addressed. Fax: (31)-631-4145.

0006-291X/94 $5.00

In this paper we report the isolation and characterization of ENA-78 genomic DNA and the 5' cDNA, and present the complete protein sequence. Putative regulatory binding sites for nuclear factors are discussed.

METHODS

Isolation of cosmids containing human ENA-78 genomic DNA. The construction of a cosmid library from flow-sorted human chromosome 4 has been described recently (13). The filters containing approximately 25'000 single colonies were screened using a random primed 230 bp BamHI/EcoRI cDNA fragment encoding exon II, III and IV of ENA-78 (1). The filters were hybridized over night at 65°C in Church buffer (14) and then washed in 6x SSC for 30 min at room temperature, followed by two washes in 0.1x SSC, 0.1% SDS at 65°C for 30 min each. Five positive clones were analyzed by southern hybridization and positive EcoRI fragments recloned into plasmid pT7T3 18U (Pharmacia Biotech AG, Dübendorf, Switzerland). Inserts were analyzed by oligonucleotide primed DNA sequencing with Sequenase (United States Biochemical Corp., Cleveland, OH).

Cloning of the 5' cDNA. Total cellular RNA from A549 cells was isolated using a modification of the method of Chirgwin et al. (15). Briefly, IL-1 stimulated A549 monolayers were scraped into a solution of 25 mM Tris/HCl, pH 8.0, containing 4.2 M guanidine isothiocyanate, 0.5% sarcosyl and 100 mM 2-mercaptoethanol. After homogenization, an equal volume of 100 mM Tris/HCl, pH 8.0, containing 10 mM EDTA and 1% sodium dodecyl sulfate was added, and the mixture was extracted twice with chloroform-phenol and chloroform-isoamyl alcohol. The RNA was precipitated with isopropyl alcohol, washed with 80% ethanol. Poly(A) mRNA was isolated by using Oligo-dT_{18} magnetisable particles (kit from Serotec, Oxford, England). First strand 5' cDNA was then synthesized using 360 ng oligonucleotide AS30 (5' TAGGATCCAATTTCCTTCCCGTTCTTCAGG 3'; indicated in Fig.1 with a dotted line), 2 µg of poly(A) mRNA, 400 units of MMLV reverse transcriptase under conditions described earlier (1). RNA was then hydrolyzed by the addition of 0.4 M NaOH for 30 min at 65°C. After neutralization the cDNA was purified on a Microcon 30 spin column (Grace AG, Wallisellen, Switzerland). The 5' rapid amplification of cDNA ends (5' RACE) was then performed by ligating an anchor oligo (5' p-CACGAATTCACTATCGATTCTGGAACCTTCAGAGG-NH_3 3') to the 3' end of the first-strand cDNA with T4 RNA ligase (16). PCR amplification of 5' cDNA was then carried out with an anchor primer (5' CTGGTTCGGCCCACCTCTGAAG GTTCCAGAATCGATAG 3') and oligonucleotide AS36C (5' CCTCTAGAAGCTTATGG CGAACACTTGCAGATTACT 3'; indicated in Fig.1 with a dotted line). PCR conditions were as follows: 1 min at 94°C; 2 min 55°C, 2 min at 72°C for 34 cycles. The 370 bp PCR product was cut with EcoRI and HindIII, purified by agarose gel electrophoresis and cloned into plasmid pT7T3 18U for DNA sequence analysis.

RESULTS AND DISCUSSION

The screening of 25'000 single colonies of a chromosome 4-specific cosmid library with a 230 bp cDNA probe yielded 5 positive clones, containing each about 40 kB of human DNA. Upon southern hybridization three of the clones were shown to contain an identical 2.2 kB genomic ENA-78 EcoRI fragment. Genomic southern blot analysis of epithelial cell line A549 or leukocyte DNA revealed one EcoRI fragment of similar size (not shown), indicating the presence of a single copy of the ENA-78 gene. The genomic 2.2 kB EcoRI ENA-78 fragment obtained from the cosmids was recloned in the pT7T3 U18 plasmid and completely sequenced (Fig. 1). In our earlier studies on the protein structure of the secreted ENA-78 we had determined

```
-706  GAATTCTCAGTAAGCGGACTTACCAAAGTAGGTGATCTGTAGGGGAGTTAACAAAATTCA
-646  GTGGTCCTTTCAGGCCACTGACTTCAAGTGGCAAGAGACAAGGGTCTCTTGTTATCATGT
-586  TATCTTGGCTTCCAAAGCTGGTTGAAGTCCAGAGATTCATAAAGTCATTCAAGAAACCTA
-526  GAATGACCTGCCTGCAAGAAGACAGGAAGGACTTTCAGTTTATAGCAATTCAAACATGAA
-466  TAACATTTCCTGATTAATAGTAATAATAATTAGAAAGGATTGACTTTCAGAAATTTTTCT
-406  CAAATCAAGGCTCCTGTTACTTTGGTTCCACCTTTTCTCTCTAGAAGGAGAGGAGGAGCA
-346  TCTCCCAGATGCTGCGTGCTCCAGAAAAGCCGGCATCCCTAGCCCGCTCTGGCACAGGCC
-286  ATGAGGCGCTGCTGAATCCTGCTGAATAGCTACTCCCTTCTAGCTGGAGCCACAGCTCCC
-226  TCCACCGCGGAACAGGGTTACAACGTCCCTCTCGGTAGAGGTGCACGCAGCTCCTCCTGG
-166  CCACCCTCCCCACCAGTTCCCATTGTCTGGCCCCCCTCCCCCAACCTCTTCTTTCCACAC
-106  TGCCCCATGAGTTCAGGGAATTTCCCCAGCATCCCAAAGCTTGAGTTTCCTGTCAGTGGG

 -46  GAGAGATGAGTGTAGATAAAAGGAGTGCAGAAGGAACGAGGAAGCCACAGTGCTCCGGAT
                                                          ↑
  15  CCTCCAATCTTCGCTCCTCCAATCTCCGCTCCTCCACCCAGTTCAGGAACCCGCGACCGC

  75  TCGCAGCGCTCTCTTGACCACTATGAGCCTCCTGTCCAGCCGCGCGGCCCGTGTCCCCGG
                              M  S  L  L  S  S  R  A  A  R  V  P  G
 135  TCCTTCGAGCTCCTTGTGCGCGCTGTTGGTGCTGCTGCTGCTGCTGACGCAGCCAGGGCC
        P  S  S  S  L  C  A  L  L  V  L  L  L  L  L  T  Q  P  G  P
 195  CATCGCCAGCGGTGAGAGCGCATGGCGCGCGGGACGCACTCGCACTCGGGCACAGAGGTG
        I  A  S
 255  CATCCCAGCCTCTGCGGGGTCGCTGCGTTCCAGGGAACTCTCCCAGCAACCTGCCCTATA

 315  AAGGGTGTCTCTCTTTCTTCCCCAGCTGGTCCTGCCGCTGCTGTGTTGAGAGAGCTGCGT
                                  A  G  P  A  A  A  V  L  R  E  L  R
 375  TGCGTTTGTTTACAGACCACGCAAGGAGTTCATCCCAAAATGATCAGTAATCTGCAAGTG
       C  V  C  L  Q  T  T  Q  G  V  H  P  K  M  I  S  N  L  Q  V
 435  TTCGCCATAGGCCCACAGTGCTCCAAGGTGGAAGTGGTGTAAGTTCTGTGCTGCTGTGTC
       F  A  I  G  P  Q  C  S  K  V  E  V  V
 495  CGCTGTGACCTTGGCAAGAGAGAAATCCCGCAGCCTGGGTCTTCAACCTTGGTATCTCAT

 555  GAGTGTATCTTCTTTTTCTTTCCTTCAGAGCCTCCCTGAAGAACGGGAAGGAAATTTGTC
                                      A  S  L  K  N  G  K  E  I  C
 615  TTGATCCAGAAGCCCCTTTTCTAAAGAAAGTCATCCAGAAAATTTTGGACGGGTACTTGT
      L  D  P  E  A  P  F  L  K  K  V  I  Q  K  I  L  D  G
 675  CACTTTGATCTTTGTGGTTTCTAAATCTGATCTAGGGAGACCATAGACTTCACAAGGTCT

 735  TTATTCTCTGTACGATTTAAGTAACACTTTTCATGTTTAGAATTAAAAGGTTGTTGAATT

 795  GGGAAAGTTTTTCTGGATTGTCCTGGGAAAATATACCAATCTTACATGTAATTACTTGAG

 855  CAATTACACACAGCTTGTCACTAAGTTATGTTTTTTGTTTACCCATTGCTTTTATTGATT

 915  TTTGTATTCTCCTTTTTTACCAAACATCATAAACGCTGAGTTTTGACAAGGGTGGAGTAG

 975  AAAGGAGTGTGAAAAATGGTTAAACTAATATAACATTTTTCTCAACAGTGGAAACAAGGA
                                                        G  N  K  E
1035  AAACTGATTAAGAGAAATGAGCACGCATGGAAAAGTTTCCCAGTCTTCAGCAGAGAAGTT
       N ------
1095  TTCTGGAGGTCTCTGAACCCAGGGAAGACAAGAAGGAAAGATTTTGTTGTTGTTTGTTTA
1155  TTTGTTTTTCCAGTAGTTAGCTTTCTTCCTGGATTCCTCACTTTGAAGAGTGTGAGGAAA
1215  ACCTATGTTTGCCGCTTAAGCTTTCAGCTCAGCTAATGAAGTGTTTAGCATAGTACCTCT
1275  GCTATTTGCTGTTATTTTATCTGCTATGCTATTGAAGTTTTGGCAATTGACTATAGTGTG
1335  AGCCAGGAATCACTGGCTGTTAATCTTTCAAAGTGTCTTGAATTGTAGGTGACTATTATA
1395  TTTCCAAGAAATATTCCTTAAGATATTAACTGAGAAGGCTGTGGATTTAATGTGGAAATG
1455  ATGTTTCATAAGAATTC
```

Fig. 1. Nucleotide sequence of the genomic DNA encoding ENA-78. The translated amino acid sequence is shown below the appropriate DNA sequence. Numbers on the left indicate nucleotide positions in respect to the transcription initiation site (marked with arrow). Acceptor and donor sequences at splice junctions are underlined. Putative binding sites for transcription factors and enhancers are underlined or double underlined (for details see table 1). The oligonucleotides used for the 5' RACE protocol are indicated by dotted lines.

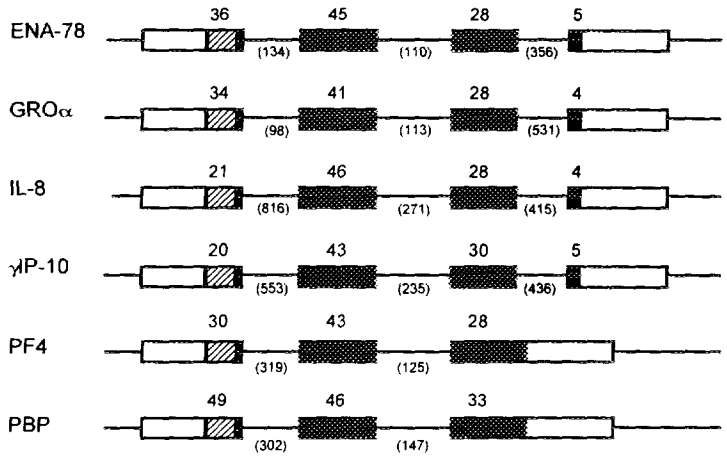

Fig. 2. Structures of the genes encoding ENA-78, IL-8 (19), GROα (3), γIP-10 (20), PF4 (21) and PBP/NAP-2 (22). Introns are shown as lines with the numbers of nucleotides in parentheses. Boxes indicate exons, open spaces represent the 5' and 3' noncoding regions, hatched areas represent leader sequences and the filled areas indicate coding sequences for the mature protein. The numbers of amino acids in each exon are indicated above each sequence.

only a partial cDNA sequence that was missing the 5' flanking region (1). Using a 5'RACE protocol we have now cloned and sequenced the ENA-78 cDNA from the transcription start site to the end of the coding region, thus allowing to resolve the genomic structure. The gene for ENA-78 consists of four exons and three introns (Fig.1 and Fig. 2). Acceptor (AG) and donor (GT) sequences (17) at splice junctions were well conserved at all the intron boundaries. The ENA-78 gene codes for a protein of 114 amino acids. The cleavage site of the signal peptide was determined according Von Heijne (18). Based on this method, the site with the highest probability to be cleaved was between amino acid 31 and 32 (Proline-Glycine). The mature peptide purified from human alveolar epithelial type-II A549 cells was 78 amino acids in length (1). This suggests that after secretion ENA-78 is proteolytically processed by 5 amino acids.

The overall structure of the ENA-78 gene resembles the genes of other C-X-C chemokines, such as IL-8 (19), GROα (3) and γIP-10 (20). In contrast, the genes of the platelet-derived cytokines platelet factor 4 (21) and platelet basic protein (22) share intron 1 and 2 with the other cytokines but lack intron 3 (Fig. 2).

The ENA-78 gene contains a number of putative regulatory sequences. The sequence GATAAAA corresponding to the TATA box was located at position - 32 to -26 upstream from the transcription initiation site (Fig. 1) . A CAAT box, another element believed to be important for transcription initiation was located at -168 to -160 (GGCCACCCT). The CAAT box is normally located closer to the initiation site (around -70) and the site observed here might therefore be inactive. In addition to constitutive promoter elements, a number of putative binding sites for regulatory proteins have been observed in the upstream region of the ENA-78 gene. A κB module (GGGAATTTCC) has been located at position -91 to -82. This structure might function as binding site for NF-κB, a factor shown to play an important role in transcriptional activation of the IL-8 gene by cytokines IL-1 and TNFα (23,24). In the IL-8 gene the κB

Table 1. Putative binding sites for transcription factors and regulatory proteins

Module	Consensus sequence	ENA-78 sequence		
TATA box	TATAAAA	-32	GATAAAA	-26
CAAT box	GGCCAATCT	-168	GGCCACCCT	-160
κB	GGGACTTTCC	-91	GGGAATTTCC	-82
AP-2	CCCCAGGC	-174	CCTCCTGGC	-166
		-129	CCCCCAACC	-121
AP-2-like	GCTGGAGGG	-144	GCTGGAGCC	-136
IRF-1	AAAGGA	-433	AAAGGA	-428
		-28	AAAGGA	-23
cAMP RE	TGACGTCA	-628	TGACTTCA	-621
AUA box	ATTTA	1439	ATTTA	1443

module is flanked on both sides by binding sites for nuclear factor NF-IL-6 (C/EBP) and it was demonstrated that maximal transcriptional activation of the IL-8 gene in Jurkat T lymphocytes requires intact binding sites for both NF-κB and NF-IL-6 (24). Similarly, strong synergistic activation of the IL-6 promoter was observed by transcription factors NF-IL-6 and NF-κB (23). In the ENA-78 gene we do not observe flanking NF-IL-6 binding sites, suggesting that induction of ENA-78 transcription by IL-1 or TNFα may be differently regulated.

Transcription factor AP-2 has been shown to play a role in epidermal-specific gene expression (25). The region upstream from the ENA-78 promoter (-121 to -174) contains three AP-2 or AP-2-like binding sites (see Table 1), which might account for the cell type specific expression of ENA-78. Upon induction with IL-1 or TNFα ENA–78 is predominantly expressed in epithelial cells and to a lesser degree in monocytes and macrophages (10). IL-8, which is produced by almost all cell types in the body, does not have AP-2-like binding sites in the same upstream region as observed for ENA-78. The 5' flanking region of the ENA-78 gene also contains motifs that could serve as binding sites for interferon regulatory factor 1 (IRF-1) (26) and motifs resembling cAMP responsive elements (19). The 3' untranslated region contains the AUUUA motif which was demonstrated to contribute to mRNA instability (27).

Functional analysis using 5' flanking deletions in the promoter upstream region will help to elucidate the significance of the binding sequences observed and to understand the mechanisms involved in ENA-78 gene regulation.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation, grant 31-36162.92 (A.W.) and by a grant of the DFG (Ri682/3-1) (O.R.).The chromosome-specific library used in this study was constructed at the Human Genome Center, Los Alamos National Laboratory, Los Alamos, NM (USA), under the auspices of the US Department of Energy and was kindly provided by L. Deaven.

REFERENCES

1. Walz, A., Burgener, R., Car, B., Baggiolini, M., Kunkel, S.L., and Strieter, R.M. (1991) J. Exp. Med. 174, 1355-1362.
2. Baggiolini, M., Walz, A., and Kunkel, S.L. (1989) J. Clin. Invest. 84, 1045-1049.
3. Haskill, S., Peace, A., Morris, J., Sporn, S.A., Anisowicz, A., Lee, S.W., Smith, T., Martin, G., Ralph, P., and Sager, R. (1990) Proc. Natl. Acad. Sci. USA 87, 7732-7736.
4. Walz, A. and Baggiolini, M. (1989) Biochem. Biophys. Res. Commun. 159, 969-975.
5. Proost, P., De Wolf-Peeters, C., Conings, R., Opdenakker, G., Billiau, A., and Van Damme, J. (1993) J. Immunol. 150, 1000-1010.
6. Deuel, T.F., Keim, P.S., Farmer, M., and Heinrikson, R.L. (1977) Proc. Natl. Acad. Sci. USA 74, 2256-2258.
7. Luster, A.D., Unkeless, J.C., and Ravetch, J.V. (1985) Nature 315, 672-676.
8. Baggiolini, M., Dewald, B., and Moser, B. (1994) Adv. Immunol. 55, 97-179.
9. Schmouder, R.L., Strieter, R.M., Walz, A., and Kunkel, S.L. (1994) Transplantation (in press).
10. Walz, A., Strieter, R.M., and Schnyder, S. (1994) Adv. Exp. Med. Biol. 351, 129-137.
11. Koch, A.E., Kunkel, S.L., Harlow, L.A., Mazarakis, D.D., Haines, G.K., Burdick, M.D., Pope, R.M., Walz, A., and Strieter, R.M. (1994) J. Clin. Invest. (in press).
12. Nishihira, J., McPhail, L.C., and O Flaherty, J.T. (1986) Biochem. Biophys. Res. Commun. 134, 587-594.
13. Riess, O., Siedlaczck, I., Kredtke, S., Melmer, G., Epplen, J.T., and Deaven, L.L. (1994) Cytogenet. Cell Genet. 65, 238-242.
14. Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991-1995.
15. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
16. Edwards, J.B.M.M., Delort, J., and Mallet, J. (1991) Nucleic Acids Res. 19, 5227-5232.
17. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
18. Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690.
19. Mukaida, N., Shiroo, M., and Matsushima, K. (1989) J. Immunol. 143, 1366-1371.
20. Luster, A.D. and Ravetch, J.V. (1987) Mol. Cell. Biol. 7, 3723-3731.
21. Eisman, R., Surrey, S., Ramachandran, B., Schwartz, E., and Poncz, M. (1990) Blood 76, 336-344.
22. Majumdar, S., Gonder, D., Koutsis, B., and Poncz, M. (1991) J. Biol. Chem. 266, 5785-5789.
23. Matsusaka, T., Fujikawa, K., Nishio, Y., Mukaida, N., Matsushima, K., Kishimoto, T., and Akira, S. (1993) Proc. Natl. Acad. Sci. USA 90, 10193-10197.
24. Kunsch, C., Lang, R.K., Rosen, C.A., and Shannon, M.F. (1994) J. Immunol. 153, 153-164.
25. Leask, A., Byrne, C., and Fuchs, E. (1991) Proc. Natl. Acad. Sci. USA 88, 7948-7952.
26. Fujita, T., Reis, L.F.L., Watanabe, N., Kimura, Y., and Taniguchi, T. (1989) Proc. Natl. Acad. Sci. USA 86, 9936-9940.
27. Shaw, G. and Kamen, R. (1986) Cell 46, 659-667.