ELSEVIER

Gene 271 (2001) 93–98

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Human–mouse comparative sequence analysis of the *NEMO* gene reveals an alternative promoter within the neighboring *G6PD* gene

Petra Galgóczy[a], André Rosenthal[a,b,1], Matthias Platzer[a,*]

[a]*Institut für Molekulare Biotechnologie, Abt. Genomanalyse, Beutenbergstrasse 11, 07745, Jena, Germany*
[b]*Friedrich-Schiller-Universität, Schlossgasse 1, 07743, Jena, Germany*

Received 15 December 2000; received in revised form 14 March 2001; accepted 20 April 2001
Received by M. D'Urso

Abstract

NEMO (NFκB essential modulator) is a non-catalytic subunit of the cytokine-dependent IκB kinase complex that is involved in activation of the transcription factor NFκB. The human *NEMO* gene maps to Xq28 and is arranged head to head with the proximal *G6PD* gene. Mutations in *NEMO* have recently been associated with Incontinentia Pigmenti (Smahi et al., Nature 405 (2000) 466), an X-linked dominant disorder. Three alternative transcripts with different non-coding 5′ exons (1a, 1b and 1c) of *NEMO* have been described. In order to identify regulatory elements that control alternative transcription we have established the complete genomic sequence of the murine orthologs *Nemo* and *G6pdx*. Sequence comparison suggests the presence of two alternative promoters for *NEMO/Nemo*. First, a CpG island is shared by both genes driving expression of the *NEMO/Nemo* transcripts containing exons 1b and 1c in one direction and the housekeeping gene *G6PD/G6pdx* in the opposite direction. In contrast to human, an additional variant of exon 1c, named 1c+, was identified in several tissues of the mouse. This larger exon utilizes an alternative donor site located 1594 bp within intron 1c. The putative second promoter for *NEMO/Nemo* transcripts starting with exon 1a is unidirectional, and not associated with a CpG island. Surprisingly, this promoter is located in the second intron of *G6PD/G6pdx*. It shows very low basal activity and may be involved in stress/time- and/or tissue-dependent expression of NEMO. To our knowledge, an overlapping gene order similar to the *G6PD/NEMO* complex has not been described before.

Keywords: Overlapping genes; 5′ untranslated region; Shared CpG island; Binding motif

1. Introduction

NEMO is essential for activation of NFκB, an ubiquitously expressed transcription factor that plays a central role in activation of inflammatory and innate immune responses (summarized in Gosh et al., 1998). IKK-γ, synonymic to NEMO, acts as a non-catalytic subunit of the IκB kinase complex (IKK) and is involved in recruitment of upstream kinases to the IKK complex (Zhang et al., 2000). The *NEMO* gene maps to Xq28 in human and is arranged head to head with *G6PD* (Jin and Jeang, 1999). It is composed of three alternative 5′ non-coding exons followed by nine coding exons (Smahi et al., 2000). Mutations in the *NEMO* gene were identified as causative for Incontinentia Pigmenti (IP), an X-linked dominant disorder that causes severe abnormalities of the skin, hair, nails, teeth, eyes and nervous system in affected females and is usually lethal in males.

The identification of regulatory elements driving the expression of *NEMO* is of special interest. First, the existence of three different 5′ exons raises the question of alternative promoters and their biological function. Second, defects in regulatory elements might lead to alterations in expression and phenotypes different to IP. Whereas reliable programs are available for the detection of protein coding sequences (Guigo et al., 2000), the identification of non-translated and regulatory elements still remains difficult when analyzing the sequence of only one species (Fickett and Hatzigeorgiou, 1997). Cross-species sequence compar-

Abbreviations: AP-1, activator protein 1; bp, base pair(s); cDNA, DNA complementary to RNA; EST, expressed sequence tag; G6PD, glucose-6-phosphate-dehydrogenase; IκB, inhibitor of NFκB; IKK, IκB kinase; IP, Incontinentia Pigmenti; INR, initiator; kb, kilobase(s); LINE, long interspersed element; LTR, long terminal repeat(s); m, minimal match; NEMO, NFκB essential modulator; NFκB, Nuclear Factor of Kappa Light Chain Gene Enhancer in B-cells; PAC, P1-derived artificial chromosome; PCR, polymerase chain reaction; RT, reverse transcriptase; SINE, short interspersed element; UTR, untranslated region; Y, pyrimidine; w, window

* Corresponding author. Tel.: +49-3641-656241; fax: +49-3641-656255.
E-mail address: mplatzer@imb-jena.de (M. Platzer).
[1] Present address: MetaGen Gesellschaft für Genomforschung mbH, Ihnestrasse 63, 14195 Berlin, Germany.

0378-1119/01/$ - see front matter
PII: S0378-1119(01)00492-9

ison is an effective analytical tool for the identification of untranslated regions (UTRs), regulatory elements and promoters (reviewed in Hardison et al., 1997; Wasserman et al., 2000; Mallon et al., 2000).

In order to identify regulatory elements driving the expression of *NEMO* exons 1a, 1b and 1c, we have mapped and sequenced the murine *Nemo* gene and its surroundings on chromosome X, band B and compared it with the orthologous human sequence. In addition to a CpG island, which is shared by both genes, we identified a putative alternative promoter of the *NEMO/Nemo* gene within intron 2 of *G6PD/G6pdx*. The genetic architecture of the two genes is remarkable, as a similar organization of overlapping 5′ noncoding elements has not previously been described.

2. Materials and methods

2.1. Genomic mapping

High density arrayed grids of genomic P1-derived artificial chromosome (PAC) libraries (http://www.chori.org/bacpac/home.htm) of mouse (RP21; BACPAC Resources, Oakland, CA) and human (RP5; RZPD, Berlin) were screened by hybridization with radioactively end-labeled (T4-Polynucleotidekinase, Roche) oligonucleotides (mouse: mG6pdF-5′-GCC ATA AGT CAG ATC CAG, mG6pdR-5′-AAG CTG AAG TGA GAC TGC, mNemoF-5′-TCA GCT TGT TGA ACT CGC, mNemoR-5′-TCT ACA AGG CTG ACT TCC; human: hF-5′-CGG CGT GCT TAT CAT TAC, hR-5′-CCG AGA AAG TCC CAG TTT) according to the manufacturer's protocol (http://www.rzpd.de/general/html/glossary/protocols/hyb_protocol.shtml). Identified PACs were verified by PCR using the probe oligos as primers.

2.2. Genomic sequence analysis

PACs were sequenced using a combination of shotgun and directed approaches as described previously (Platzer et al., 1997). Homology searches were performed using BLAST (version 1.4; Altschul et al., 1990) and FASTA (version 2.0; Pearson and Lipman, 1988). Sequence alignments were performed using the GAP (Huang, 1994). Promoter prediction was carried out with: Transcription Start Site using both Ghosh/Prestridge (TSSG) and Wingender (TSSW) motif databases (http://dot.imgen.-bcm.tmc.edu:9331/seq-search/gene-search.htm); Promoter Scan II (PSII; Prestridge, 1995); Neural Network Promoter Prediction (NNPP; http://www-hgc.lbl.gov/projects/promoter.html); Transcription Factor Search 1.3 (TFS; http://www.genome.ad.jp/htbin/nph-tfsearch); and PatSearch1.1-TRANSFAC (http://transfac.gbf.de; Wingender et al., 2000).

Human and murine genomic sequences were aligned with a percentage identity plot (PIP) (Schwartz et al., 2000). Similarity of introns was calculated by aligning repeat masked sequences with FASTA, identifying gap-free alignments above thresholds of length and identity and summing up match percentages normalized for length using the longer intron of a given pair as a reference.

Potentially polymorphic positions were amplified by nested PCR (35 cycles: 1 min at 94°C, 30 s at 52°C and 1 min at 72°C) on genomic DNA from normal control individuals of Caucasian origin, using the outer primers: DoutF: 5′-GTG TGG TTA GAA GTG GC, DoutR: 5′-CTT CCT GGT AAT TTG GCT TG, EoutF: 5′-TAC CCA ATC GCT GTC TG, and EoutR: 5′-CAT TGA GAG CAT CCA AGC and the inner primers: DinF: 5′-GGA CAC GAC CTA CAT AC, DinR: 5′-GCC CTT TTT CTC ACT TC, EinF: 5′-ATC AGG CGT GGA AGA AG, and EinR: 5′-GCT TTT GAA ACG AGG GC. The inner primers were also used for sequencing of the PCR products.

2.3. RT-PCR analyses

cDNA library panels were obtained from Clontech Laboratories: Human MTC Panel I (#K1420-1): brain, heart, kidney, liver, lung, pancreas, placenta and skeletal muscle; Mouse MTC Panel I (#K1423-1): brain, heart, kidney, liver, lung, spleen, testis, skeletal muscle, 7-, 11-, 15- and 17-day-old embryo. Nested PCR (30 cycles: 30 s at 94°C, 30 s at 53°C and 2 min at 72°C) was performed using the primers: (i) mouse: Aout-5′-CAC TGA ATC TAT GGG TGT C, Ain-5′-TAT CTG CCC ATG TG GAC, Bout-5′-CTA GGA CTT TCT AGG AG, Bin-5′-TGC CCT CTT CAC TTC TG, Cout-5′-AAT CTG CCC TTC CCT TC, Cin-5′-CCC TAG TGC TAG AAC TC, 8out-5′-TGG CAG CCA ACT TTC AG, and 8in-5′-GCT GCT CCT GCA AAT AC; (ii) human: Aout-5′-CAG GTG GGG AAA CTA AG, Ain-5′-GCA GAC TAT CAA TCC CAG, Bout-5′-TAC CAG CGT TCA CAG TC, Bin-5′-TCT GAC GGA CTC TGC T, Cout-5′-TCT GAC CCT ACT CCT TG, Cin-5′-CTT CGG AAA TGC CTC AC, 8out-5′-CCT TCA GTT TGC TGT ACT C, and 8in-5′-TCT CAG CCT GGA AGT C.

3. Results and discussion

3.1. Sequencing of G6PD and NEMO in mouse and human

To generate hybridization probes we used cDNA sequences of the murine genes *G6pdx* (Accession number NM_008062) and *Nemo* (Accession number NM_010547) and isolated the PACs RP21-17J5 and RP21-412F21 that contain the genomic sequences of both genes in mouse. Human probes were derived from the database Accession number L44140 (Chen et al., 1996) and led to the identification of PAC clones RP5-865E18 and RP5-1087L19. Here we present two 40 kb genomic regions with a G + C content of 44.5% in mouse (Accession number AF326207) and 56.1% in human (Accession number AF277315).

We compared the 40 kb human region with the database entries L44140 and AJ271718 (Smahi et al., 2000). Our sequence shows discrepancies of 0.12% (38 bp/30,255 bp)

with L44140 and 0.29% (66 bp/23,106 bp) with AJ271718. By verification of the primary data, we have excluded sequencing errors at the divergent positions in our sequence. In order to determine whether these differences represent common polymorphisms, we re-sequenced 676 bp of G6PD intron 2 containing six of the divergent positions in six women and 11 men of Caucasian origin. In all checked chromosomes the sequences were identical to AF277315, indicating that L44140 and AJ271718 represent rare alleles in the Caucasian population or contain sequencing errors.

3.2. Genomic organization of the G6PD and NEMO genes

Alignment of the human and mouse *G6PD/G6pdx* cDNAs (Accession numbers X03674 and NM_008062, respectively) with the genomic sequences identified 13 exons, the first of which is non-coding.

Four alternative transcripts of human *NEMO* are represented in GenBank. Evaluation of these sequences results in the following gene model: *NEMO* is composed of three alternative non-coding first exons 1a (Accession number AF074382), 1b (Accession number AF062089) and 1c (Accession number AF091453), followed by nine coding exons. In a fourth transcript (Accession number AK000593), starting with exon 1b, exons 5 and 7 are skipped. The murine *Nemo* cDNA (Accession number AF069542) covers only the coding exons 2–10. Searches for ESTs in dbEST (GenBank release 25.06.00: version 118.0) revealed 90/11 and 46/1 matches for *G6PD/G6pdx* and *NEMO/Nemo*, respectively. Alignment of the human *NEMO*-specific ESTs with the genomic sequence confirms all exons. Alternative splice forms were not identified. The single murine *Nemo*-specific EST covers only exons 3–7 and failed to detect the murine 5′ exons. For this reason, we employed similarity searches with the corresponding human exons to identify the murine 5′ non-coding exons 1a, 1b and 1c of *Nemo* (Table 1).

The genomic organization and the relative arrangement of the *G6PD/G6pdx* and *NEMO/Nemo* genes is similar in human and mouse. In both species the genes are arranged head to head (Fig. 1). The transcription of *G6PD* is directed to the centromere whereas *NEMO* is transcribed in the opposite direction. Both human genes are shorter than their murine orthologs. The human *G6PD* spans 15.9 kb, while its murine counterpart is spread over 19.5 kb. The genomic sequences of *NEMO/Nemo* are 22.8 and 23.4 kb in human and mouse, respectively. Analysis of the local base content showed a CpG island at the 5′ end of *G6PD/G6pdx*, which

Table 1
Exon–intron organization of the *G6PD/G6pdx* and *NEMO/Nemo* genes in human and mouse[a]

	Number	Exon (bp)			Intron (bp)			
		Human	Mouse	Identity A (%)	Human	Mouse	Identity B (%)	Identity C (%)
G6PD	1	466	> 44	60	625	772	20	0
	2	8/**120**	8/**120**	89	9857	12911	9	3
	3	**38**	**38**	95	95	99	0	0
	4	**109**	**109**	87	551	826	16	0
	5	**218**	**218**	86	671	1387	0	0
	6	**159**	**159**	91	177	227	0	0
	7	**126**	**126**	88	365	307	0	0
	8	**94**	**94**	86	447	120	0	0
	9	**187**	**187**	82	139	265	0	0
	10	**236**	**236**	88	104	97	0	0
	11	**77**	**77**	88	105	93	0	0
	12	**93**	**93**	84	97	99	0	0
	13	**91**/609	**91**/721	74				
NEMO	1a	> 176*	> 195*	62	5252	4051	23	4
	1b	> 106*	> 106*	74	174	145	29	0
	1c	> 170*	> 79*	62	3899	2759	0	0
	1c+	–	> 1673	–	–	1165	–	–
	2	15/**187**	15/**187**	87	3977	796	0	0
	3	**212**	**212**	83	2156	3149	13	0
	4	**119**	**119**	87	1756	1624	6	0
	5	**153**	**153**	83	1131	815	10	0
	6	**97**	**76**	89	1025	3394	0	0
	7	**144**	**144**	88	605	283	15	0
	8	**143**	**143**	85	257	3944	11	0
	9	**62**	**62**	76	298	296	0	0
	10	**143**/585	**143**/592	73				

[a] Exon sizes shown in bold are coding regions. *Transcription start not experimentally determined. Identity A, GAP; identity B, FASTA ($m = 50$, $w = 60$); identity C, FASTA ($m = 70$, $w = 100$).

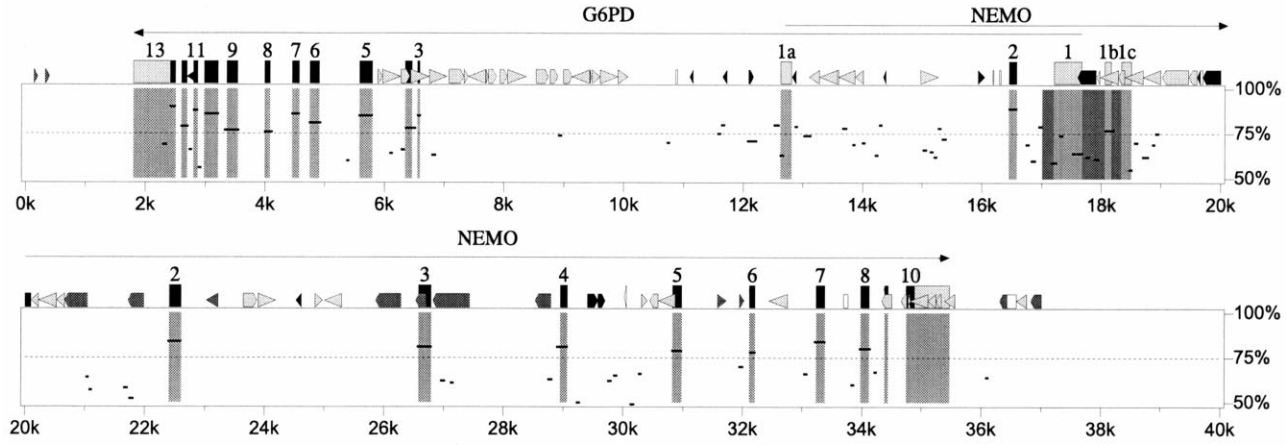

Fig. 1. PIP of human vs. mouse genomic sequence. Only segments with an identity of at least 50% over 60 bp or more were plotted. Features in the PIP are as follows: black boxes and light gray columns, coding exons; gray boxes and light gray columns, UTR; dark gray columns, CpG islands where the ratio CpG/GpC exceeds 0.75; white pointed boxes, L1 repeats; light gray triangles, SINEs other than MIR; black triangles, MIR; black pointed boxes, LINE2; dark gray triangles and pointed boxes, other kinds of interspersed repeats, such as LTR elements and DNA transposons. The degree of sequence identity is indicated as a percentage scale on the right of each row.

includes exon 1b of *NEMO* in human. The CpG island has been described as essential for the ubiquitous expression of the housekeeping gene *G6PD/G6pdx* in mouse and man (Toniolo et al., 1991; Philippe et al., 1994; Franzé et al., 1998). It is not surprising that the expression of the closely linked NEMO/Nemo gene is also driven by the same CpG island, since many CpG islands possess bidirectional transcription activity (Brenner et al., 1997; Platzer et al., 1997).

Remarkably, the genes overlap in their 5′ non-coding regions in both species: The most 5′ non-coding exon 1a of *NEMO/Nemo* is located in the second intron of *G6PD/G6pdx*. This arrangement is unusual. Overlaps that have been described before in eukaryotic systems typically result in the production of complementary transcripts, involved in antisense regulation (Shintani et al., 1999), or one small gene resides in an intron of a larger gene (Cawthon et al., 1991), or the exon of one gene coincides with promoter elements of the neighboring gene (van Bokhoven et al., 1996). To our knowledge this is the first time that an overlapping gene order, where the 5′ exons of two genes are nested in introns of the respective neighbor, has been observed. A driving force for this unusual gene organization may be the need to utilize additional regulatory elements for alternative transcripts especially in extremely gene dense loci such as Xq28 (Chen et al., 1996). In the case of *NEMO/Nemo* the overlapping gene architecture facilitates the usage of the CpG island for basal ubiquitous expression. On the other hand, an alternative unidirectional promoter may allow inducible expression of NEMO.

3.3. Expression pattern of the alternative NEMO/Nemo transcripts

In order to study the tissue distribution of the alternative *NEMO/Nemo* exons 1a, 1b and 1c as well as the skipping of exons 5 and 7, we performed nested RT-PCR analysis on human and murine MTC panels (Clontech) using forward primers that are specific for the three alternative 5′ exons and reverse primers located in exon 8. All RT-PCR reactions were repeated in four (human) to eight (mouse) independent experiments.

Our RT-PCR analysis confirms the presence of the three alternative transcripts of the human *NEMO* gene that differ in their 5′ UTR. PCR with exon 1b-specific primers generated the expected product in each of the tested human tissues on four separate occasions. In contrast, amplification with exon 1a-specific primers produced four times the expected product only in liver but failed twice with heart, kidney and lung and gave no results in the remaining tissues tested. Only one of four amplifications with exon 1c primers yielded products for heart, lung, liver and skeletal muscle (Fig. 2A).

Nested PCR with mouse-specific primers for exon 1b resulted in PCR products of 1025 bp in all tested murine tissues. Amplification with *Nemo* 1c-specific primers twice yielded PCR products of the expected size of 1041 bp in spleen and once in testis. By sequencing these PCR products, we could verify the predicted splice acceptor and donor sites of introns 1b and 1c of the murine *Nemo* gene. Interestingly, some PCR with 1c-specific primers amplified a product 1594 bp larger than expected: six times in heart and skeletal muscle, two times in embryos at 17 and once in spleen, kidney, embryos at E7 and E15. Sequencing revealed an alternative splice donor site in *Nemo* intron 1c not conserved in human. This novel 5′ exon overlaps in its 5′ part with exon 1c and therefore was named 1c+. The murine exon 1a could not be detected (Fig. 2B). Moreover, we did not observe transcripts skipping exons 5 and 7 in human or murine samples.

Our results indicate that in mouse and human the dominant transcript of the *NEMO/Nemo* gene starts with exon 1b.

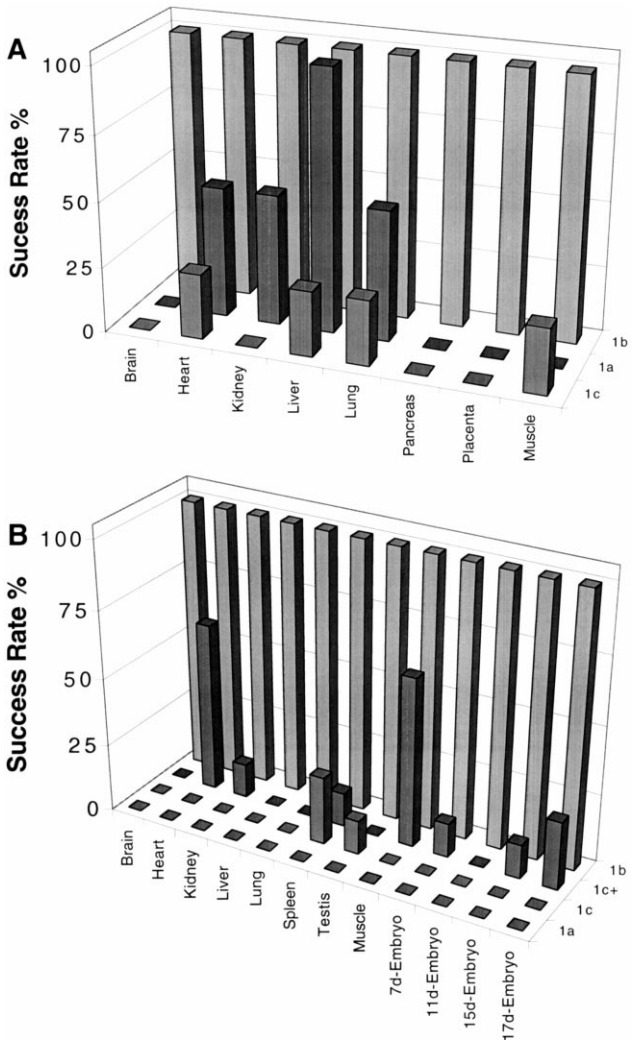

Fig. 2. Results of RT-PCR experiments using (A) human and (B) murine MTC panels.

The minor transcripts starting with exon 1a and 1c failed to be amplified reliably by RT-PCR, possibly because of low expression levels. This interpretation is supported by a lack of respective signals in Northern blot analysis (data not shown). In addition, the analysis of the EST database (see above) supports the idea that *Nemo* transcription in mouse is lower than in human. The dbEST version 118.0 contains 2,232,809 human and 1,604,115 murine ESTs, giving a human/mouse ratio of 1.4. The fact that *NEMO/Nemo* was represented by 46 human and only one murine EST gives a human/mouse expression ratio of about 30:1. Similarly, the level of *G6pdx* is substantially lower in mouse compared with human by a factor of 5.7 (90 human vs. 11 murine ESTs). In both species *NEMO/Nemo* seems to be expressed at lower levels than *G6PD/G6pdx* (by factors of two and 11, respectively). The lower expression levels in mouse can be connected with the fact that the CpG island driving the major transcripts is much less pronounced in comparison to the orthologous island in human (327 bp with one Sp1 site vs. 1327 bp with 12 Sp1 sites). This agrees with observations that the mouse genome underwent secondary changes leading to a narrower G + C distribution (Sabeur et al., 1993) and severe depletion of typical CpG island features (Matsuo et al., 1993).

3.4. Conserved regions in intron 2 of G6PD/G6pdx identify an alternative promoter

In order to identify sequences that might play a role in the regulation of expression, we searched for conserved regions in introns and the intergenic sequence. We first identified gap-free alignments with a minimal match of 50% over a window of at least 60 bp ($m = 50$, $w = 60$) in human and mouse. Several regions that meet these parameters were found in the genomic sequence of introns 1, 2 and 4 of *G6PD/G6pdx* and in introns 1a, 1b, 3, 4, 5, 7 and 8 of *NEMO/Nemo* (Table 1, Fig. 1). Second, using more stringent parameters ($m = 70$, $w = 100$) only three conserved *G6PD/G6pdx* regions, all of which are located in intron 2, fulfill the selection criteria. These regions are 168, 103 and 136 bp long and show an identity of 71, 80 and 74%, respectively, spread over 1081 bp in human, and 1111 bp in mouse. The promoter prediction program TSSW revealed a putative transcription start site within the human region 50 bp upstream of exon 1a of *NEMO*. This site is embedded in a pyrimidine-rich stretch that corresponds to an initiator (INR) element [Y-A-$(Y)_5$]. A CAAT box is positioned 60 bp further upstream. Although the software failed to predict a promoter for the corresponding mouse sequence, the human–mouse alignment shows conservation of the putative transcription start site.

In scrutinizing our sequence for the transcription factor binding motif, we focused on the three conserved regions. Whereas in the most 5′ region no conserved binding motifs were detectable, the 103 bp region shows five identical binding motifs in man and mouse, namely two sites for activator protein 1 (AP-1) and one site each for nuclear factors E2 (NF-E2), Lyf-1 and CdxA. AP-1 is involved in the transcriptional response to pathogens, inflammatory signals and stress, similar to NFκB (Wisdom, 1999). The 136 bp region that is positioned in the *NEMO/Nemo* intron 1a showed one conserved binding motif for RAR-related orphan receptor alpha1 (RORalp).

We assume that this is a unidirectional promoter driving the *NEMO/Nemo* transcript starting with exon 1a in human and mouse. The identification of conserved binding motifs for transcription factors may guide further studies of the stress/time- and/or tissue-dependent expression of NEMO.

Acknowledgements

We thank Beate Fischer, Hella Ludewig and Evelin Michaelis for expert technical assistance. This work was supported by grants from the German BMBF (BEO

0311108/0) and from the European Commission (BMH4-CT96-0338).

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Brenner, V., Nyakatura, G., Rosenthal, A., Platzer, M., 1997. Genomic organization of two novel genes on human Xq28: compact head to head arrangement of IDH gamma and TRAP delta is conserved in rat and mouse. Genomics 44, 8–14.

Cawthon, R.M., Andersen, L.B., Buchberg, A.M., Xu, G.F., O'Connell, P., Viskochil, D., Weiss, R.B., Wallace, M.R., Marchuk, D.A., Culver, M., 1991. cDNA sequence and genomic structure of EV12B, a gene lying within an intron of the neurofibromatosis type 1 gene. Genomics 9, 446–460.

Chen, E.Y., Zollo, M., Mazzarella, R., Ciccodicola, A., Chen, C.N., Zuo, L., Heiner, C., Burough, F., Repetto, M., Schlessinger, D., D'Urso, M., 1996. Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci. Hum. Mol. Genet. 5, 659–668.

Fickett, J.W., Hatzigeorgiou, A.G., 1997. Eukaryotic promoter recognition. Genome Res. 7, 861–878.

Franzé, A., Ferrante, M.I., Fusco, F., Santoro, A., Sanzari, E., Martini, G., Ursini, M.V., 1998. Molecular anatomy of the human glucose 6-phosphate dehydrogenase core promoter. FEBS Lett. 437, 313–318.

Gosh, S., May, M.J., Kopp, E.B., 1998. NFκB and Rel proteins: evolutionarily conserved mediators of immune responses. Annu. Rev. Immunol. 16, 225–260.

Guigo, R., Agarwal, P., Abril, J.F., Burset, M., Fickett, J.W., 2000. An assessment of gene prediction accuracy in large scale DNA sequences. Genome Res. 10, 1631–1642.

Hardison, R.C., Oeltjen, J., Miller, W., 1997. Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res. 7, 959–966.

Huang, X., 1994. On global sequence alignment. Comput. Appl. Biosci. 10, 227–235.

Jin, D.Y., Jeang, K.T., 1999. Isolation of full-length cDNA and chromosomal localization of human NF-kappaB modulator NEMO to Xq28. J. Biomed. Sci. 6, 115–120.

Mallon, A.M., et al., 2000. Comparative genome sequence analysis of the Bpa/Str region in mouse and man. Genome Res. 10, 758–775.

Matsuo, K., Clay, O., Takahashi, T., Silke, J., Schaffner, W., 1993. Evidence for erosion of mouse CpG islands during mammalian evolution. Somat. Cell Mol. Genet. 19, 543–555.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444–2448.

Philippe, M., Larondelle, Y., Lemaigre, F., Mariame, B., Delhez, H., Mason, P., Luzzatto, L., Rousseau, G.G., 1994. Promoter function of the human glucose-6-phosphate dehydrogenase gene depends on two GC boxes that are cell specifically controlled. Eur. J. Biochem. 226, 377–384.

Platzer, M., Rotman, G., Bauer, D., Uziel, T., Savitsky, K., Bar-Shira, A., Gilad, S., Shiloh, Y., Rosenthal, A., 1997. Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. Genome Res. 7, 592–605.

Prestridge, D.S., 1995. Predicting Pol II promoter sequences using transcription factor binding sites. J. Mol. Biol. 249, 923–932.

Sabeur, G., Macaya, G., Kadi, F., Bernardi, G., 1993. The isochore patterns of mammalian genomes and their phylogenetic implications. J. Mol. Evol. 37, 93–108.

Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., Miller, W., 2000. PipMaker – a web server for aligning two genomic DNA sequences. Genome Res. 10, 577–586.

Shintani, S., O'hUigin, C., Toyosawa, S., Michalova, V., Klein, J., 1999. Origin of gene overlap: the case of TCP1 and ACAT2. Genetics 152, 743–754.

Smahi, A., et al., 2000. Genomic rearrangement in NEMO impairs NF-kappaB activation and is a cause of incontinentia pigmenti. The International Incontinentia Pigmenti (IP) Consortium. Nature 405, 466–472.

Toniolo, D., Filippi, M., Dono, R., Lettieri, T., Martini, G., 1991. The CpG island in the 5′ region of the G6PD gene of man and mouse. Gene 102, 197–203.

van Bokhoven, H., Rawson, R.B., Merkx, G.F., Cremers, F.P., Seabra, M.C., 1996. cDNA cloning and chromosomal localization of the genes encoding the alpha- and beta-subunits of human Rab geranylgeranyl transferase: the 3′ end of the alpha-subunit gene overlaps with the transglutaminase 1 gene promoter. Genomics 38, 133–140.

Wasserman, W.W., Palumbo, M., Thompson, W., Fickett, J.W., Lawrence, C.E., 2000. Human-mouse genome comparisons to locate regulatory sites. Nat. Genet. 26, 225–228.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., Schacherer, F., 2000. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. 28, 316–319.

Wisdom, R., 1999. AP-1: one switch for many signals. Exp. Cell Res. 253, 180–185.

Zhang, S.Q., Kovalenko, A., Cantarella, G., Wallach, D., 2000. Recruitment of the IKK signalosome to the p55 TNF receptor: RIP and A20 bind to NEMO (IKKgamma) upon receptor stimulation. Immunity 12, 301–311.