

Research article

Complex splicing pattern generates great diversity in human *NFI* transcripts

Ina Vandembroucke, Tom Callens, Anne De Paepe and Ludwine Messiaen*

Address: Centre for Medical Genetics, Ghent University Hospital, De Pintelaan 185, 9000 Ghent, Belgium

E-mail: Ina Vandembroucke - Ina.Vandembroucke@rug.ac.be; Tom Callens - Tom.Callens@rug.ac.be; Anne De Paepe - Anne.DePaepe@rug.ac.be; Ludwine Messiaen* - Ludwine.Messiaen@rug.ac.be

*Corresponding author

Published: 24 May 2002

BMC Genomics 2002, **3**:13

Received: 28 March 2002

Accepted: 24 May 2002

This article is available from: <http://www.biomedcentral.com/1471-2164/3/13>

© 2002 Vandembroucke et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Mutation analysis of the neurofibromatosis type I (*NFI*) gene has shown that about 30% of *NFI* patients carry a splice mutation resulting in the production of one or several shortened transcripts. Some of these transcripts were also found in fresh lymphocytes of healthy individuals, albeit typically at a very low level. Starting from this initial observation, we were interested to gain further insight into the complex nature of *NFI* mRNA processing.

Results: We have used a RT-PCR plasmid library based method to identify novel *NFI* splice variants. Several transcripts were observed with specific insertions/deletions and a survey was made. This large group of variants detected in one single gene allows to perform a comparative analysis of the factors involved in splice regulation. Exons that are prone to skipping were systematically analysed for 5' and 3' splice site strength, branch point strength and secondary structure.

Conclusion: Our study revealed a complex splicing pattern, generating a great diversity in *NFI* transcripts. We found that, on average, exons that are spliced out in part of the mRNA have significantly weaker acceptor sites. Some variants identified in this study could have distinct roles and might expand our knowledge of neurofibromin.

Background

The genome of higher eukaryotes codes for most of its proteins through short coding sequences (exons) interrupted by non-coding sequences (introns). These intervening sequences must be removed precisely after transcription of the gene, a process called splicing. Splicing is carried out by the spliceosome, a large complex consisting of proteins and small nuclear RNAs (snRNAs) [1–3]. RNA splicing is dependent upon the identity of nucleotide sequences at the exon/intron boundaries. The 5' boundary or donor site of introns in most eukaryotes con-

tains the dinucleotide GU, while the 3' boundary or acceptor site contains the dinucleotide AG. These GU-AG motifs are merely parts of longer consensus sequences that span the 5' and 3' splice sites. With the exception of the conserved canonical GU and AG at the 5' and 3' splice sites, splice site sequences may contain several deviations from the consensus. The match of a splice site to this consensus reflects the strength of the site. In addition, a third conserved intronic sequence that is known to be functionally important in splicing is the branch point, located 10–50 nucleotides upstream from the AG dinucleotide. The

branch point is characterised by a rather weak consensus sequence YURAY [4]. In the first step of splicing the 2' hydroxyl of an adenosine of the branch point attacks the phosphate at the 5' splice site, producing a free 5' exon and a lariat intermediate. During the second step, the 3' hydroxyl of the 5' exon attacks the phosphate at the 3' splice site, yielding a ligated mRNA and a lariat intron.

Now that the total human genome has been sequenced, the surprising outcome is that differences between humans and other organisms are not reflected by the genome complexity. Alternative splicing might play a role in this apparent discrepancy. Alternative splicing is an important mechanism to create protein diversity and to regulate gene expression in a tissue- or developmental-specific manner. To understand the mechanism of alternative splice site selection the identification of all *cis*-acting elements that contribute to the efficiency with which splice sites are recognised must be taken into account. These include 1) the strength of the 3' and 5' splice sites, which can be calculated using the consensus values that were determined by Shapiro and Senepathy [5] or by the Splice Site Prediction program by Neural Networks (SSPNN, [http://www.fruitfly.org/seq_tools/splice.html]) which takes more known sequences into account, 2) the branch point sequences and 3) certain sequences such as exonic splicing enhancers (ESE) that can activate the use of a weak 3' splice site or elements such as exonic splicing silencer elements (ESS) that inhibit splicing [1]. In addition to these *cis*-acting elements, exon length [6] and secondary structure have also been mentioned to play a role in splicing [7].

Neurofibromatosis type 1 (NF1) is an autosomal dominantly inherited disorder affecting one in 3500 individuals [8]. The gene spans 350 kb of genomic DNA and encodes approximately 13 kb mRNA containing 60 exons. Three exons (9br, 23a and 48a) are known to be subjected to alternative splicing and these isoforms were also detected at the protein level with isoform-specific antibodies [9,10]. Four additional alternative transcripts were described (ex29-, ex30-, ex29/30- and the N-isoform) but no further analysis at the protein level was reported [11,12]. During mutation analysis in the *NF1*-gene we disclosed several additional splice variants in which specific exons were skipped and that were also present in fresh lymphocytes of non-affected persons, albeit typically at a low level [13]. Several other studies also report the existence of lowly expressed splice *NF1* variants. Some of these variants are more abundant when RNA is analysed from aged blood or from blood that was kept at a non-physiological temperature [13–16]. Starting from this initial observation, we were interested to determine the extent of aberrant and/ or alternative splicing in *NF1*. A RT-PCR plasmid library was made to identify new *NF1* splice variants. We have made a survey of these variants and searched systematically for a number of factors involved in splice site selection aiming to explain the observed variants.

Results and discussion

Identification of novel *NF1* splice variants in normal cells

The entire *NF1*-coding region from two normal non-affected individuals was amplified in 5 overlapping segments by RT-PCR according to Heim [17] starting from fresh blood leukocytes. As expression of specific *NF1* transcripts can change due to environmental changes [13–16,18], we extracted the RNA immediately after blood drawing, allowing to identify splice variants that are not induced by stress factors. Plasmid libraries were created and 150 colonies per fragment were lysed and reamplified with the same primers in order to search for the presence of minor splice variants. The PCR products were digested with a restriction enzyme and separated by electrophoresis on an agarose gel (Fig. 1). For the five fragments, a total of 114 clones out of 750 colonies in total showed a different restriction pattern. Sequence analysis of the inserts revealed that these represented *NF1* transcripts with specific insertions/deletions. Some of these resulted from cryptic splice site usage, others were found to have deletions precisely at the exon boundaries ('exon skipping'), very few resulted from intron retention. All variants that were found are summarised in Fig. 2. Transcripts containing more than one change per single transcript were also identified. Some changes were in frame (14 variants, green in Fig. 2) others were out of frame (32 variants, red in Fig. 2) and hence would lead to a premature stop codon. All together, 46 different transcripts were detected. The number

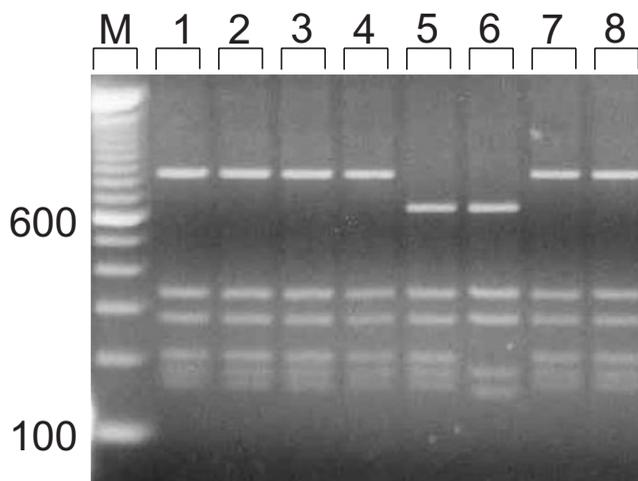


Figure 1
Representative restriction pattern of eight colonies of *NF1*-F4 digested with *Avall*. (1)-(4), (7), (8): full length *NF1*-F4, (5): *NF1*- Δ E30, (6): *NF1*- Δ [E30, E36/E37]

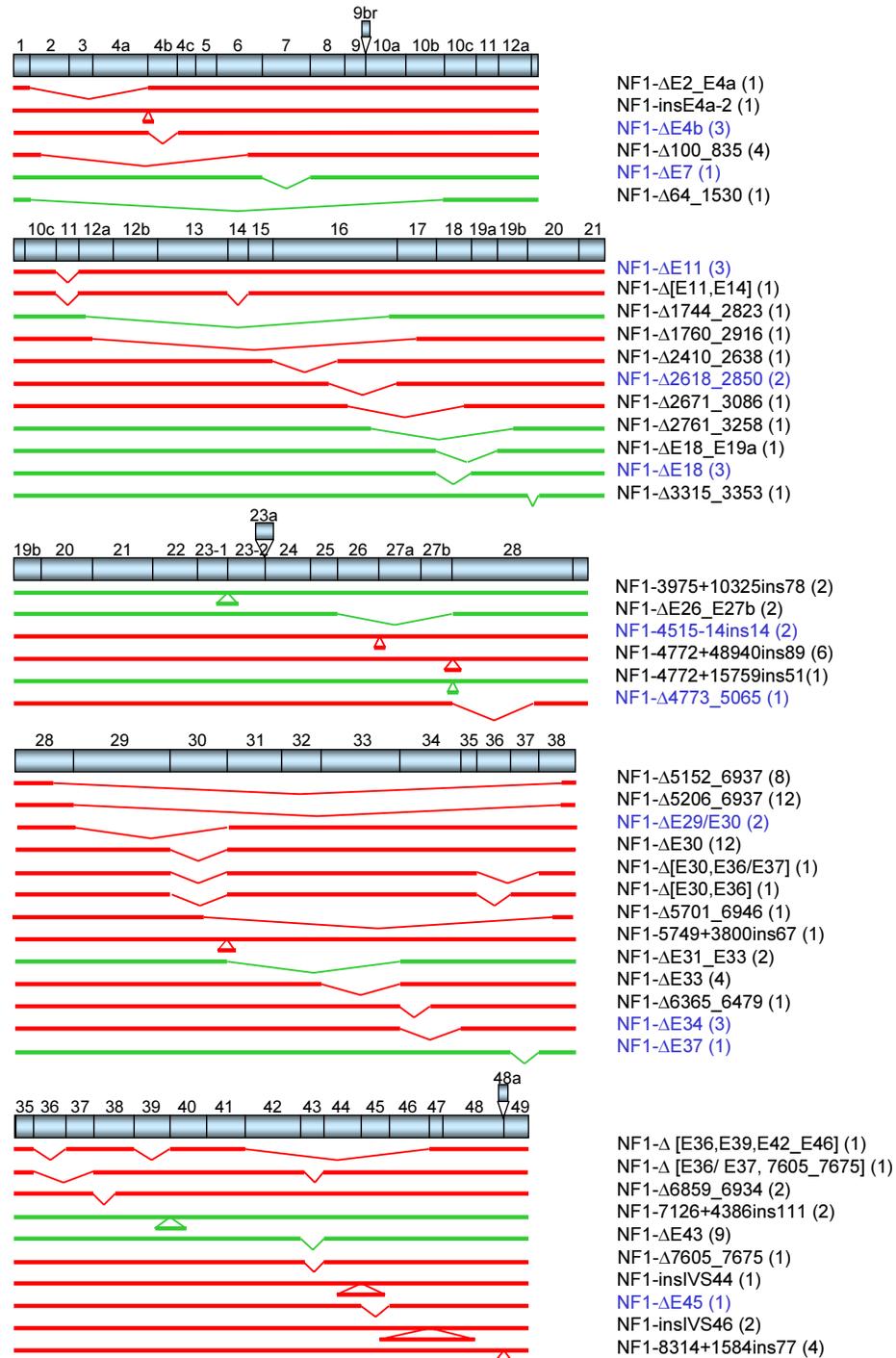


Figure 2
Schematic representation of all splice variants detected in five overlapping NFI fragments. In frame events are indicated in green whereas out of frame events are red. Transcript variants described in blue are identical to misspliced transcripts, described previously for NFI patients, caused by a genomic mutation. The number of colonies detected containing a specific variant is indicated between brackets.

of abnormal sized transcripts is not exhaustive as only 150 colonies per fragment were analysed (750 colonies in total). The frequency of each clone captured is indicated between brackets in Fig. 2. This frequency might indicate that some NF1 transcripts are more abundant in blood, as compared to rarer NF1 transcript variants. Some of these variants have been independently observed by other groups: NF1ins4a-2 [14-16], NF1ΔE4b [14-16], NF1ΔE7 [15,19], NF1ΔE29/E30 [11], NF1ΔE30 [11], NF1ΔE33 [16], NF1ΔE34 [16], NF1ΔE37 [15,16], NF1ΔE43 [16], NF1-Δ7605-7675 [16] and NF1ΔE45 [16]. Some of these transcripts increased with time or cold shock [14,16]. It has been shown that ~30 % of NF1 germline mutations result in splicing errors [13,20]. The vast majority of them are due to alterations outside the canonical AG/GT acceptor and donor splice sites. Some of the splice variants identified in this study are identical to misspliced transcripts formed by the mutant allele carrying a specific mutation in NF1 patients (11 variants, blue in Fig. 2). For example, the NF1-Δ2618_2850 transcript was found in a NF1 patient carrying an IVS16+3del6 [13] mutation. In this patient mutation at the splice donor of exon 16 leads to the constitutively activation of the cryptic donor present in exon 16 at position 2618 that normally is used in only a minor proportion of the transcripts.

Role of splice site strength in splice site selection

Intensive effort has already been made to understand the mechanisms involved in splice site selection. Because we detected such a large group of variants in one single gene, a detailed examination of the different parameters that play a role in alternative splice regulation could be performed and compared statistically. Early splice site choice in mammals involves protein-protein interactions across the exon, as is proposed in the exon definition model of Berget [21]. Exon definition is mediated in part by SR proteins bridging between the small nuclear ribonucleoprotein (snRNP) U1 at a downstream 5' splice site and the U2AF heterodimer at the upstream 3' splice site. First we calculated the strength of the 5' and 3' splice sites of all constitutive and alternative exons and of all deletions and insertions reported in this study. Calculations were performed using the Splice Site Prediction by Neural Network (SSPNN, [http://www.fruitfly.org/seq_tools/splice.html]) program (Table 1). In general, sequences that show a better match to the consensus sequences, indicated by a higher calculated score, are considered to be 'strong', as specificity of the splicing reaction is mediated in part by RNA-RNA interactions between these sequences and snRNA molecules. The calculated scores demonstrate that deletions or insertions sometimes occur at strong cryptic splice sites (e.g. the NF1-Δ2618_2850 variant), whereas in a few cases no or only a weak consensus splice site was found (score of < 0.7: underscored in Table 1, e.g.

NF1-Δ5701_6946). The latter might result from aberrant splicing or PCR artefacts.

We wanted to examine whether an association exists between the observed splicing pattern and splice site strength. To this end, a comparison of the scores was made between those single exons that were spliced out in a fraction of the mRNA (bold in Table 1), including the alternative exons 9br, 23a and 48a, and the mean values for all NF1 exons (Table 2). Skipping of more than one exon probably involves other factors and mechanisms than those based on splice parameters alone. The exon definition model is no more applicable as the distance would be too long to accommodate exon-bridging interactions. Therefore, only exons that were spliced out *individually* were compared to the mean values. The SSPNN program gave the exons that can become skipped a significantly lower score (0.63, $p = 0.03$) at the acceptor site, compared to the mean NF1 acceptor score for all NF1 exons (0.81). No such correlation could be found at the donor site. Exon bridging interactions may occur less efficiently as a result of weak binding of splice factors at the acceptor site, thus modulating the splicing efficiency. This finding underlines the importance of the acceptor site in splice site selection. This is in agreement with a recent study [22] in which it was shown that mutating the acceptor site to a stronger site was more effective for full intron splicing of an alternative intron, than strengthening the donor site. Likewise, there was an inverse relationship found between the length of the polythymidine tract, which influences the acceptor strength, at the exon 9 acceptor site and the proportion of exon 9 deleted CFTR mRNA transcripts [23]. Our results indicate that acceptor splice site strength is an important parameter in the regulation of NF1 splicing.

Role of branch point position and strength

Some studies have pointed to the importance of the branch point in alternative splicing [24,25]. The distance between the branch point and the 3' splice site and the base pairing of the branch point with the U2RNA is important for splicing. The majority of branch points used in vertebrate splicing are between 10 and 50 nucleotides upstream from the acceptor splice site. We have analysed 1) the position of the putative branch point sequence and 2) the match of the putative branch point sequence to the loosely defined consensus YURAY. We developed a program to calculate the strength of a branch point. The region between -50 and -10 upstream from the acceptor site was examined for a branch point signal and the corresponding branch point score was calculated (Table 1). A relatively low score points to a suboptimal branch point or to a branch point situated outside the chosen window (-50, -10). Branch points located at long distances are also suboptimal. However, mean branch point strength (Table

Table 1: Neurofibromatosis Type I splicing parameters.

<i>Exon nr.</i>	<i>Frame</i>	<i>Score acceptor</i>	<i>Score donor</i>	<i>Branch point score</i>	<i>Branch point position</i>
Exon 1	IF		0.91		
Exon 2	IF	1.00	0.99	0.84	38
Exon 3	IF	0.79	1.00	0.99	31
Exon 4a	OF	0.81	0.99	0.77	16
Exon 4b	OF	1.00	1.00	0.97	41
Exon 4c	OF	<u>0.66</u>	1.00	0.87	38
Exon 5	OF	0.99	0.99	0.74	49
Exon 6	OF	<u>0.55</u>	1.00	0.97	30
Exon 7	IF	<u>0.64</u>	0.93	0.87	23
Exon 8	IF	<u>0.68</u>	0.99	0.74	33
Exon 9	IF	1.00	0.99	0.90	37
Exon 9br	IF	<u>0.05</u>	0.96	0.82	36
Exon 10a	IF	0.98	0.99	0.92	47
Exon 10b	IF	0.98	0.86	0.98	26
Exon 10c	IF	0.99	0.99	0.82	27
Exon 11	OF	<u>0.01</u>	0.79	0.73	42
Exon 12a	OF	<u>0.23</u>	1.00	0.82	36
Exon 12b	IF	0.98	<u>0.19</u>	0.87	20
Exon 13	OF	0.72	0.99	0.71	28
Exon 14	OF	<u>0.65</u>	0.99	0.88	36
Exon 15	IF	0.97	1.00	0.88	34
Exon 16	IF	<u>0.46</u>	0.92	0.97	43
Exon 17	OF	0.99	1.00	0.82	35
Exon 18	IF	0.98	0.99	0.98	32
Exon 19a	IF	0.98	0.94	<u>0.67</u>	43
Exon 19b	IF	0.99	<u>0.13</u>	0.90	31
Exon 20	OF	0.96	0.79	0.82	43
Exon 21	OF	0.90	0.99	<u>0.69</u>	26
Exon 22	IF	0.99	0.92	0.99	23
Exon 23-1	OF	0.99	0.75	0.76	28
Exon 23-2	OF	1.00	0.98	0.74	38
Exon 23-a	IF	0.95	0.98	0.90	15
Exon 24	IF	0.99	<u>0.41</u>	0.87	29
Exon 25	OF	0.91	0.88	0.98	29
Exon 26	IF	0.79	1.00	0.97	20
Exon 27a	IF	<u>0.49</u>	<u>0.53</u>	0.97	31
Exon 27b	IF	0.99	0.99	0.90	27
Exon 28	OF	1.00	0.98	0.97	32
Exon 29	OF	0.81	0.97	0.82	38
Exon 30	OF	0.98	0.96	0.75	18
Exon 31	OF	0.98	<u>0.58</u>	0.97	19
Exon 32	IF	0.78	0.99	0.77	33
Exon 33	OF	0.99	1.00	0.74	27
Exon 34	OF	0.93	<u>0.41</u>	0.76	24
Exon 35	OF	0.98	1.00	0.90	46
Exon 36	OF	<u>0.56</u>	0.89	0.72	28
Exon 37	IF	<u>0.09</u>	0.99	0.82	42
Exon 38	IF	0.93	1.00	0.89	29
Exon 39	OF	<u>0.47</u>	0.82	0.90	29
Exon 40	IF	1.00	0.70	0.89	35
Exon 41	OF	0.98	1.00	0.84	23
Exon 42	OF	0.98	0.73	0.82	27
Exon 43	IF	0.99	0.77	0.84	14
Exon 44	OF	0.91	0.71	0.82	28
Exon 45	OF	<u>0.52</u>	0.96	0.74	20
Exon 46	OF	1.00	<u>0.63</u>	0.90	34
Exon 47	OF	0.80	1.00	0.82	11
Exon 48	OF	0.82	0.98	0.90	16

Table 1: Neurofibromatosis Type I splicing parameters. (Continued)

Exon 48 a	IF	0.16	0.90	0.89	32
Exon 49	OF	1.00		0.78	24
NFI-insE4a-2	OF	<u>0.51</u>	0.99	0.90	10
NFI-Δ100_835	OF	0.81	<u>0.48</u>	0.87	13
NFI-Δ64_1530	IF	/	/	0.82	30
NFI-Δ1744_2823	IF	/	/	0.84	20
NFI-Δ1760_2916	OF	/	/	0.83	41
NFI-Δ2410_2638	OF	<u>0.03</u>	1.00	<u>0.69</u>	24
NFI-Δ2618_2850	OF	0.99	0.95	0.82	35
NFI-Δ2671_3086	OF	/	/	0.77	20
NFI-Δ2761_3258	IF	0.81	<u>0.07</u>	0.77	25
NFI-Δ3315_3353	IF	/	<u>0.13</u>	0.80	41
NFI-3975+10325ins78	IF	0.76	0.99	0.92	32
NFI-4515-14ins14	OF	0.92	1.00	0.97	17
NFI-4772+48940ins89	OF	<u>0.01</u>	0.93	0.77	22
NFI-4772+15759ins51	IF	<u>0.14</u>	0.98	0.90	30
NFI-Δ4773_5065	OF	0.84	0.99	1.00	37
NFI-Δ5152_6937	OF	<u>0.47</u>	0.72	0.82	24
NFI-Δ5206_6937	OF	<u>0.47</u>	0.98	0.82	24
NFI-Δ5701_6946	OF	/	/	0.82	33
NFI-5749+3800ins67	OF	<u>0.28</u>	<u>0.35</u>	0.89	28
NFI-Δ6365_6479	OF	0.82	1.00	<u>0.69</u>	17
NFI-Δ6859_6934	OF	/	0.99	0.82	21
NFI-7126+4386ins111	IF	<u>0.52</u>	0.95	0.78	44
NFI-Δ7605_7675	OF	0.91	/	0.82	22
NFI-8314+1584ins77	OF	0.99	0.92	0.80	20

‘/’: No score can be obtained by SSPNN, not even at the lowest threshold. Exons in bold were detected to skip as individual exon in some transcripts. Underscored scores have values of <0.70.

Table 2: Mean splicing parameters for exons that splice out in part of the mRNA (A) versus the mean of all NFI exons (B).

	acceptor score		donor score		branch point score	
	A	B	A	B	A	B
mean	0.63	0.81	0.90	0.88	0.83	0.85
p-value	0.030		0.704		0.411	

2) was not significantly lower (p = 0.411) for the exons associated with skipping.

Role of secondary structure in splice site selection

Recently, it was suggested that NFI exon skipping events could be explained by structural alterations in possible higher free-energy structures of the pre-mRNA at the donor site [19]. Here in, it was hypothesised that when an al-

ternative structure is too different from the lowest free-energy structure (ground state) this sequence is not recognised by the splicing machinery and that, as a consequence, the corresponding exon becomes skipped. We further elaborated this hypothesis by comparing the lowest free energy structures of the donor- and acceptor sites of all exons that were prone to skipping (bold in Table 1) to their first higher free-energy structures using the mfold program [26,27]. Some structures showed remarkable alterations between both energy states, but many others did not at all. A representative example is given in figure 3 where a striking structural change was found between the energy states of the donor of exon 7 (Fig. 3A,3B). On the other hand, very similar structures were obtained for the donor of exon 4b (Fig. 3C,3D) although this exon was shown to be skipped in a significant part of the mRNA in several tissues (unpublished results) using the highly accurate real-time PCR technique [28]. Moreover, when the window is displaced over a few nucleotides, as exemplified here for exon 8, the ground state alters considerably, (Fig. 3E,3F), demonstrating the difficulty of assigning sig-

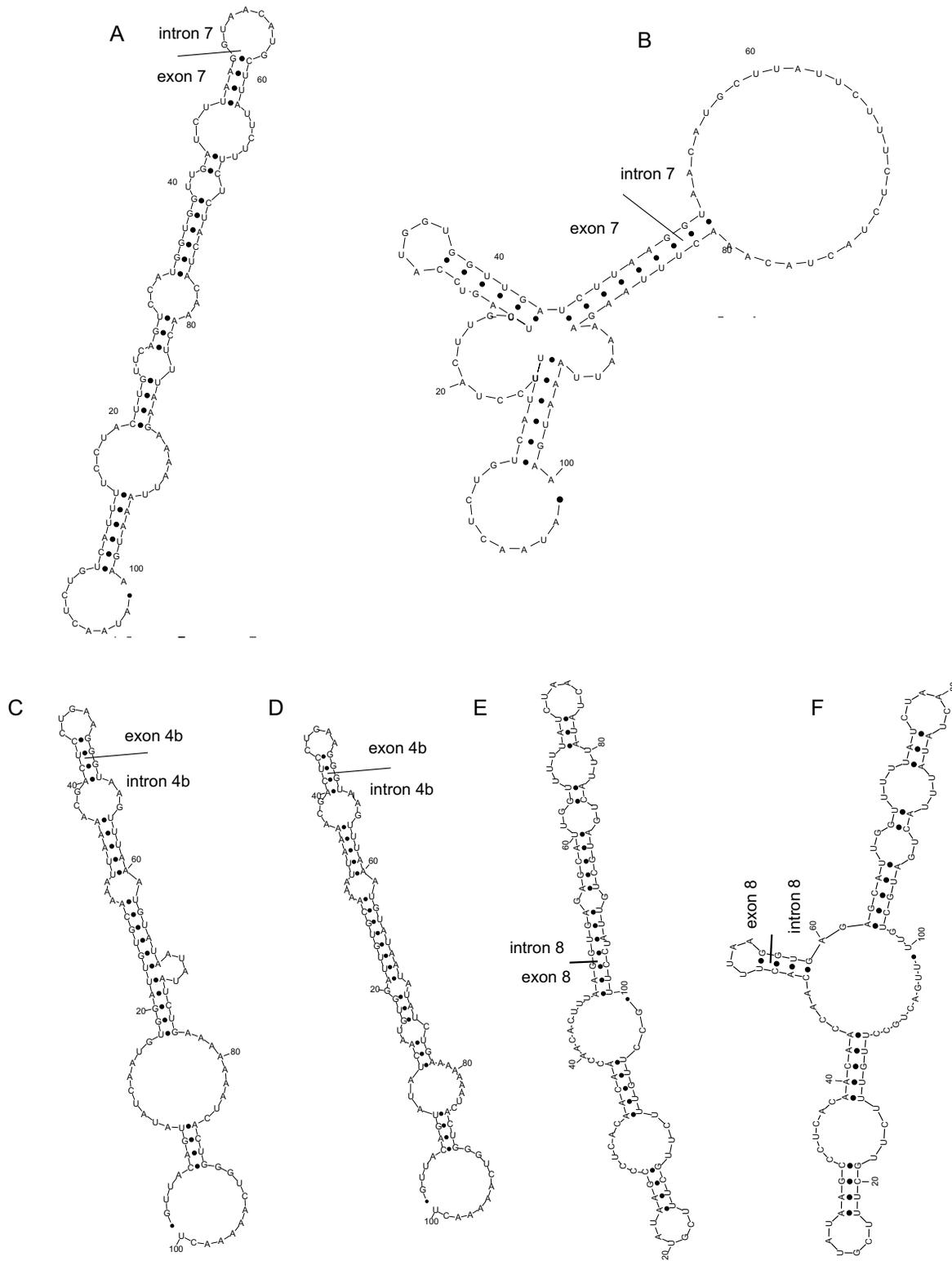


Figure 3
Simulations of secondary structure. Secondary structure of the sequence surrounding the donor site of exon 7 (A, B), exon 4b (C, D) and exon 8 (E-F). A and C: lowest free-energy structures; B and D: first higher energy-structure; E and F: prediction using a window of (-50, +50) (E) or (-56, +44) (F) with respect to the splice site.

nificance to these *in silico* results. These data indicate that one has to be extremely careful when correlating exon skipping events to secondary structure predictions. Limitations of the mathematical model, uncertainties in the thermodynamic parameters and influences of the secondary structure *in vivo* by specific interactions of RNA-protein complexes limit the prediction capacities of secondary structure by these programs.

On the other hand, comparison of the ground state of a mutant donor site, as was done for G1185+1U and G1185+1A in [19], to the higher energy structure of the normal donor is superfluous to explain skipping. It is well known that splice sites mutated at their highly conserved AG-GT bases are not recognised anymore by the splicing machinery since recognition e.g. of the splice donor by the U1snRNP is mainly directed by RNA-RNA base pairing [29,30].

Conclusions

The data presented in this study indicate that splicing in *NF1* is extremely complex. Although several alternative transcripts have been found previously for the *NF1* gene [9–12], we identified multiple *novel NF1* splice variants and made a survey of them. This large catalogue of variants detected in one single gene allowed to perform a comparative analysis of the factors involved in splice regulation in order to explain the observed variants. We performed an extensive analysis of different splicing parameters. These include the strength of the 5' and 3' splice site, the branch point strength and secondary structure predictions at the donor and acceptor site. On the basis of the analysed features we can conclude that an interplay of various factors is involved in regulation of *NF1* splicing. Competition between the alternative splice sites depends on the relative quality of the different constitutive splice signals and we found that the acceptor strength probably plays a major role in this regulation. The functional significance of the transcripts identified in this study still remains unclear. What fraction of the splicing represents 'noise', caused by the relaxation of the RNA splicing, is currently unknown. Although the RNA was extracted immediately after blood drawing, excluding as far as possible influences due to stress factors, some splice events occurred at non-consensus splice sites and probably represent 'aberrant' splicing. However, some of these novel identified transcripts could potentially encode proteins of different sizes and may have distinct roles. Further research was started to analyse tissue-specific expression of several transcripts. Preliminary results show that some transcripts are highly expressed in specific tissues, demonstrating that splicing in *NF1* may be regulated in a tissue specific manner. Further analysis of these transcripts is needed and it will remain a challenging task to elucidate the biological significance of all of them.

Materials and methods

RNA isolation and cDNA preparation

RNA was isolated from fresh blood leukocytes from two normal individuals using TRIzol LS Reagent (Invitrogen) according to the manufacturer's instructions. RNA extraction was performed immediately after blood drawing, excluding as far as possible influences due to stress factors. Total RNA (2 µg) was reverse transcribed using SuperScript II Reverse transcriptase (Invitrogen) and random hexamers (Amersham Biosciences).

RT-PCR library formation and PCR analysis

Primers used for the amplification of the total *NF1* cDNA in 5 overlapping fragments were as described [17]. Reactions were performed as described [13]. The PCR products were cloned in the pCR2.1-TOPO-vector (Invitrogen) and transformed into TOP-10F' cells. The colonies were lysed and the subcloned fragments were reamplified with the same primers. The five *NF1* fragments (NF1-F1 to NF1-F5) were digested subsequently with different restriction enzymes: *ScrFI*, *StyI*, and *BsrI* for NF1-F1, *HaeIII* and *AvaII* for NF2-F2, *HaeII* for NF1-F3 and NF1-F5, and *AvaII* for NF1-F4. All transcripts with a different restriction pattern were sequenced (primer sequences available upon request) using the Thermo Sequenase fluorescent labelled primer cycle sequencing kit (Amersham Biosciences) and analysed on an ALF-express automated DNA sequencer (Amersham Biosciences).

Nomenclature for the description of the different *NF1* transcripts

In order to clearly describing the identity (ID) of the transcripts, following nomenclature was used:

- The ID of all transcript variants start with 'NF1' followed by '-'.
 - Sequence changes are all described at the RNA level, with 1 = first base of methionine (ATG) at the start. Numbering system for insertions are based on the *NF1* genomic sequence (Genbank Accession Nr. AC004526).
 - Deletions are designated with the symbol 'Δ', insertions with 'ins'.
 - Deletions are described with a '_' separating the first and the last deleted exon.
 - Skipping of a complete exon is designated by an 'E' preceding the exon number. Two consecutive deleted exons are separated by '/'. When more than two consecutive exons were deleted, the first and the last deleted exon are separated by a '_' character, non-consecutive exons are separated by a ',' character to indicate that intermediate exons are not deleted.

- Insertions are described by a '+' or a '-' after the nucleotide flanking the insertion site, followed by the position of the first inserted nucleotide of the intron (beginning of intron = +1, for insertions flanking the acceptor site: end of preceding intron = -1), followed by 'ins' and the length of the inserted sequence.
- Insertions of a complete intron are designated by 'IVS' preceding the intron number.
- When more than one change was detected, these changes are described between square brackets, separated by a ',' character.

Splice site scores

The sequence environment of all donor- and acceptor sites was analysed using Splice Site Prediction by Neural Network (SSPNN, URL address: [http://www.fruitfly.org/seq_tools/splice.html]).

Branch point scores

We searched for branch point sequences in a window of -50, -10 with respect to the acceptor site. A program was written in DELPHI to calculate the strength of a branch point. Every 5-nucleotide sequence within the window was evaluated as a potential branch point. The 'standard' branch point sequence weight-table [4] was used to assign a score at every 5-nucleotide sequence within the window. The computed score reflects how closely it resembles known branch point sequences. The sequence with the highest score was selected as the 'putative' branch point.

Secondary structures

The mfold computer program version 3.0 [http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi] was used to predict the RNA secondary structure and to calculate the folding free energy [26,27]. We have computed the minimum-free-energy structures from all the donor- and acceptor sites of the 60 exons of *NF1* within a window of 100 bp. It has been proposed [31] that there is a window of about 100 nucleotides after transcription where the pre-mRNA is free to fold. The size of this window would be defined by the time taken for the protein complexes to bind as the nascent RNA emerges.

Statistical analysis

The one-way ANOVA test was used to assess significance in the comparison of mean values.

Acknowledgements

Ina Vandenbroucke was supported by a grant of the Ghent University (BOF 01107799 and 011D3801). We thank Katharina Wimmer from the University of Vienna for helpful discussions.

References

1. Lopez AJ: **Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation.** *Annu Rev Genet* 1998, **32**:279-305
2. Smith CW, Valcarcel J: **Alternative pre-mRNA splicing: the logic of combinatorial control.** *Trends Biochem Sci* 2000, **25**:381-388
3. Staley JP, Guthrie C: **Mechanical devices of the spliceosome: motors, clocks, springs, and things.** *Cell* 1998, **92**:315-326
4. Harris NL, Senapathy P: **Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis.** *Nucleic Acids Res* 1990, **18**:3015-3019
5. Shapiro MB, Senapathy P: **RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression.** *Nucleic Acids Res* 1987, **15**:7155-7174
6. Dominski Z, Kole R: **Selection of splice sites in pre-mRNAs with short internal exons.** *Mol Cell Biol* 1991, **11**:6075-6083
7. Balvay L, Libri D, Fiszman MY: **Pre-mRNA secondary structure and the regulation of splicing.** *Bioessays* 1993, **15**:165-169
8. Riccardi VM: **Von Recklinghausen neurofibromatosis.** *N Engl J Med* 1981, **305**:1617-1627
9. Gutmann DH, Geist RT, Rose K, Wright DE: **Expression of two new protein isoforms of the neurofibromatosis type I gene product, neurofibromin, in muscle tissues.** *Dev Dyn* 1995, **202**:302-311
10. Gutmann DH, Zhang Y, Hirbe A: **Developmental regulation of a neuron-specific neurofibromatosis I isoform.** *Ann Neurol* 1999, **46**:777-782
11. Park VM, Kenwright KA, Sturtevant DB, Pivnick EK: **Alternative splicing of exons 29 and 30 in the neurofibromatosis type I gene.** *Hum Genet* 1998, **103**:382-385
12. Suzuki H, Takahashi K, Kubota Y, Shibahara S: **Molecular cloning of a cDNA coding for neurofibromatosis type I protein isoform lacking the domain related to ras GTPase-activating protein.** *Biochem Biophys Res Commun* 1992, **187**:984-990
13. Messiaen LM, Callens T, Mortier G, Beysens D, Vandenbroucke I, Van Roy N, Speleman F, De Paepe A: **Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects.** *Hum Mutat* 2000, **15**:541-555
14. Ars E, Serra E, de la Luna S, Estivill X, Lazaro C: **Cold shock induces the insertion of a cryptic exon in the neurofibromatosis type I (NF1) mRNA.** *Nucleic Acids Res* 2000, **28**:1307-1312
15. Wimmer K, Eckart M, Rehder H, Fonatsch C: **Illegitimate splicing of the NF1 gene in healthy individuals mimics mutation-induced splicing alterations in NF1 patients.** *Hum Genet* 2000, **106**:311-313
16. Thomson SAM, Wallace MR: **RT-PCR splicing analysis of the NF1 open reading frame.** *Hum Genet* 2002, **111**:495-502
17. Heim RA, Kam-Morgan LN, Binnie CG, Corns DD, Cayouette MC, Farber RA, Aylsworth AS, Silverman LM, Luce MC: **Distribution of 13 truncating mutations in the neurofibromatosis I gene.** *Hum Mol Genet* 1995, **4**:975-981
18. Messiaen LM, Callens T, Roux KJ, Mortier GR, De Paepe A, Abramowicz M, Pericak-Vance MA, Vance JM, Wallace MR: **Exon 10b of the NF1 gene represents a mutational hotspot and harbors a recurrent missense mutation Y489C associated with aberrant splicing.** *Genet Med* 1999, **1**:248-253
19. Kaufmann D, Leistner W, Kruse P, Kenner O, Hoffmeyer S, Hein C, Vogel W, Messiaen L, Bartelt B: **Aberrant splicing in several human tumors in the tumor suppressor genes neurofibromatosis type 1, neurofibromatosis type 2, and tuberous sclerosis 2.** *Cancer Res* 2002, **62**:1503-1509
20. Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X: **Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type I.** *Hum Mol Genet* 2000, **9**:237-247
21. Berget SM: **Exon recognition in vertebrate splicing.** *J Biol Chem* 1995, **270**:2411-2414
22. Romano M, Marcucci R, Baralle FE: **Splicing of constitutive upstream introns is essential for the recognition of intra-exonic suboptimal splice sites in the thrombopoietin gene.** *Nucleic Acids Res* 2001, **29**:886-894
23. Chu CS, Trapnell BC, Currustin S, Cutting GR, Crystal RG: **Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA.** *Nat Genet* 1993, **3**:151-156

24. Libri D, Balvay L, Fiszman MY: **In vivo splicing of the beta tropomyosin pre-mRNA: a role for branch point and donor site competition.** *Mol Cell Biol* 1992, **12**:3204-3215
25. Goux-Pelletan M, Libri D, d'Aubenton-Carafa Y, Fiszman M, Brody E, Marie J: **In vitro splicing of mutually exclusive exons from the chicken beta-tropomyosin gene: role of the branch point location and very long pyrimidine stretch.** *Embo J* 1990, **9**:241-249
26. Zuker M, Mathews DH, Turner DH: **Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology.** Dordrecht, Kluwer Academic Publishers 1999
27. Mathews DH, Sabina J, Zuker M, Turner DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.** *J Mol Biol* 1999, **288**:911-940
28. Vandenbroucke II, Vandesompele J, De Paepe A, Messiaen L: **Quantification of splice variants using real-time PCR.** *Nucleic Acids Res* 2001, **29**:E68-68
29. Mount SM, Pettersson I, Hinterberger M, Karmas A, Steitz JA: **The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro.** *Cell* 1983, **33**:509-518
30. Zhuang Y, Weiner AM: **A compensatory base change in U1 snRNA suppresses a 5' splice site mutation.** *Cell* 1986, **46**:827-835
31. Eperon LP, Graham IR, Griffiths AD, Eperon IC: **Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase?** *Cell* 1988, **54**:393-401

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



BioMedcentral.com

Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com