Haemophilia (2006), **12,** 258–262 DOI: 10.1111/j.1365-2516.2006.01216.x

Predicting severity of haemophilia A and B splicing mutations by information analysis

Y. VON KODOLITSCH*, J. BERGER† and P. K. ROGAN‡§

**Department of Cardiology, University Hospital Eppendorf; †Institute of Mathematics and Computer Science in Medicine, University Hospital Eppendorf, Hamburg, Germany; ‡Laboratory of Human Molecular Genetics, Children's Mercy Hospital; and §Computing and Engineering, University of Missouri-Kansas City, Schools of Medicine and Kansas City, MO, USA*

Summary. Bleeding symptoms and clotting activity vary among mutations that alter mRNA splicing of either the factor VIII or factor IX genes. We analyzed splicing mutations in both genes for changes in individual information (R_i, in bits) involving both donor or acceptor sites. Mutations with low or negative R_i values (<2.4 bits) or significant changes in R_i ($\Delta R_i \geq 7$ bits) exhibited either reduced protein activity, increased clotting time and bleeding frequency and were predictive of severe disease. Thus, information analysis of splicing mutations may be useful in predicting phenotypes in hemophilia.

Keywords: *F8c*, *F9*, information theory, mRNA splicing, Mutation analysis, phenotype

Molecular diagnostic analysis of factors VIII (*F8C*) and IX (*F9*) genes identifies mutations that affect clotting activities, factor levels, bleeding severity, and the risk of inhibitor development. Characterization of individual genotypes may be useful in guiding treatment of asymptomatic patients to prevent arthropathy or inhibitor development [1–4]. Of the approximately 16% of mutations in genes which affect mRNA splicing [5], some abolish normal mRNA, and others reduce transcript levels active abnormal splice isoforms, or have little or no effect on expression. In some instances, effects on splicing are not inherently evident from nucleotide sequence changes. It is generally not feasible to assess the effect of every variant on mRNA levels, because haemophilic genes are poorly expressed in leucocytes and because of the logistics of undertaking such comprehensive analyses.

We examined the possibility that splicing mutations in *F8C* and *F9* and the corresponding clinical phenotypes could be predicted *in silico* with information theory-based models which measure changes in donor and acceptor splice site affinity. The individual information content (R_i) measures the strength (in bits) of the splice site relative to other functional sites and is related to binding affinity. Splicing mutations reduce R_i values of natural splice sites and/or significantly increase the values to create or strengthen cryptic sites [6,7]. Previous studies of other disorders have shown that individual information analysis can accurately predict the severity of splicing mutations based on the calculated level of residual normal mRNA and corresponding phenotypes [6–9]. Complete splice site inactivation ($R_i < 2.4$ bits) results in either exon skipping or cryptic splicing and severe disease. Conversely, mutant splice sites with $R_i \geq 2.4$ bits may be incompletely inactivated, reducing levels of mRNA, resulting in milder disease [7]. Most substitutions with little or no change in R_i values have little effect on splicing [6]. The reduction in correctly spliced mRNA can also be estimated from the minimum fold change in splice site strength as $1/(2^{\Delta R_i})$ [7–9].

We analysed 94 variants predicted to alter splicing from the Human Mutation (http://www.hgmd.org), Hemophila A (http://europium.csc.mrc.ac.uk) and B (http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html) databases, among 584 single nucleotide substitutions in the *F8C* or *F9* genes [5]. Clotting activities and clinical severity were derived from databases and the original reports [5]. The information contents of 71 splicing variants (including seven in *F8C* and 64 in

Correspondence: Peter K. Rogan, Children's Mercy Hospital and Clinics, 2401 Gillham Road, Kansas City, MO 64108, USA.
Tel.: 816 983 6511; fax: 816 983 6515;
e-mail: progan@cmh.edu

Accepted after revision 23 January 2006

Table 1. Usefulness of splice-site information content (R_i) measurements for predicting severity of haemophilia A and B.

	Sensitivity		Specificity	
	Per cent of patients	True positives/total with severe phenotype*	Per cent of patients	True negatives/total with non-severe phenotype*
Severe reduction of clotting activity				
R_i < 2.4 bits	86	32/37	61	19/31
ΔR_i > 7 bits (0.7%)	77	41/53	80	12/15
Severe bleeding (Eyster)				
$R_i \leq 2.4$ bits	100	9/9	46	6/13
ΔR_i > 7 bits (0.7%)	100	19/19	83	5/6

*Number of mutations.

F9) were determined with the Automated Splice Site Analysis system (https://splice.cmh.edu) and compared with their corresponding clinical phenotypes (refer to Appendix, which comprehensively cites all the mutations), after excluding those with incomplete or missing phenotypes ($n = 13$) or those whose effects could not be dissociated from either linked missense changes ($n = 8$) and compound heterozygous mutations ($n = 2$). Reduction in activity was graded severe for levels ≤1% of normal, moderate for values 1–5%, and mild for values >5% of normal [10,11]. Clinical severity of 28 splice site variants was graded mild, moderate or severe according to Eyster bleeding symptom criteria [12].

Information analysis predicted the complete loss of splice site recognition for 37 mutations ($R_{i,variant} < 2.4$ bits), residual or leaky splicing for 31 mutations ($2.4 \leq R_{i,variant} << R_{i,normal}$) and essentially no effect for three variants ($R_{i,variant} \sim R_{i,normal}$). However, concomitant activation of cryptic splice sites was predicted for 17 mutations. In the two patients in which mRNA had been analysed, predicted null alleles were confirmed. Information contents of splice mutations were related to corresponding phenotypes by determining thresholds for discriminating mild or moderately reduced protein activity from severe phenotypes with a receiver operating characteristic curve (ROC; 13). Most predicted non-functional splice sites with R_i values <2.4 bits were clinically severe, correctly predicting both a severe reduction of clotting activity with 86% sensitivity and 61% specificity and severe bleeding symptoms with 100% sensitivity and 46% specificity (Table 1). Similarly, we found that changes in splice site strength ($\Delta R_i \geq 7$ bits) predicted to reduce mRNA levels to < 0.7% of normal corresponded well with severe phenotypes ($P = .001$; area under ROC curve was 0.74 for both R_i and ΔR_i). Using this ΔR_i threshold severely reduced clotting activity was predicted with 77% sensitivity and 80% specificity and severe bleeding symptoms with 100% sensitivity and 83% specificity. Cryptic splice site activation was predicted for 13 mutations with severely reduced protein expression, whereas four cryptic mutations exhibited mildly or moderately reduced protein levels, presumably due to residual splicing at the corresponding natural splice sites [7–9].

While mRNA analysis remains the gold standard for splicing mutation analysis [14], it is nevertheless not feasible to evaluate all potential F8C and F9 variants in each patient for possible effects on splicing. Information theory-based models can prioritize among variants for subsequent characterization as potential splicing mutations. These models, which have proven reliable for assessing splice site mutations of other genes [6,7,9], can often predict phenotypic severity of mutations responsible for the major forms of haemophilia.

Acknowledgements

We wish to thank Dr. Brian Wicklund for his valuable comments on this manuscript.

References

1 Carlsson M, Berntorp E, Bjorkman S, Lindvall K. Pharmakokinetik dosing in prophylactic treatment of hemophilia A. *Eur J Haematol* 1993; **51**: 247–52.
2 Fakharzadeh SS, Kazazian HHJ. Correlation between factor VIII genotype and inhibitor development in hemophilia A. *Semin Thromb Hemost* 2000; **26**: 167–71.
3 Lusher JM. Hemophilia treatment. Factor VIII inhibitors with recombinant products: prospective clinical trials. *Haematologica* 2000; **85**: 2–5.
4 van den Berg HM, Fischer K, Mauser-Bunschoten EP *et al*. Long-term outcome of individualized prophylactic treatment of children with severe hemophilia. *Br J Haematol* 2001; **112**: 561–5.
5 Krawczak M, Cooper DN. The human gene mutation database. *Trends Genet* 1997; **13**: 121–2.

6 Rogan PK, Schneider TD. Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Hum Mutat* 1995; **6**: 74–6.
7 Rogan PK, Faux BM, Schneider TD. Information analysis of human splice site mutations. *Hum Mutat* 1998; **12**: 153–71.
8 Allikmets R, Wassermann WW, Hutchinson A *et al*. Organization of the ABCR gene: analysis of promotor and splice junction sequences. *Gene* 1998; **215**: 111–22.
9 von Kodolitsch Y, Pyeritz RE, Rogan PK. Splice site mutations in atherosclerosis candidate genes: relating individual information to phenotype. *Circulation* 1999; **100**: 693–9.
10 David D, Moreira I, Morais S, de Deus G. Five novel factor IX mutations in unrelated hemophilia B patients. *Hum Mutat* 1998; **1**(Suppl.): S301–3.
11 Ketterling RP, Drost JB, Scaringe WA *et al*. Reported in vivo splice site mutations in the factor IX gene: severity of splicing defects and a hypothesis for predicting deleterious splice donor mutations. *Hum Mutat* 1999; **13**: 221–31.
12 Eyster ME, Lewis JH, Shapiro SS *et al*. The Pennsylvania hemophilia program 1973–1978. *Am J Hematol* 1980; **9**: 277–86.
13 Jenicek M. *Foundations of Evidence-Based Medicine*. New York, NY: Taylor and Francis, 2003; 125–6.
14 Maddalena A, Bale S, Das S, Grody W, Richards SP. Technical standards and guidelines: molecular genetic testing for ultra-rare disorders. *Genet Med* 2005; **7**: 571–83.

Appendix. Information analysis of splicing mutations in hemophilia A and B

No.	Patient database ID*	GenBank Accession number, Mutation, coordinate	Information-theory based prediction: $R_{i(nat)} \rightarrow R_{i(mut)}$ (in bits) Natural splice site	Cryptic splice site; coordinate	Reduction of mRNA-levels	Phenotypic characteristics: Protein mass	Protein activity	Clinical severity (Eyster)
F8C gene								
6	JH107	M88638, ds+5, G → A, 246	7.1 → 3.6	–	+	…	…	+
5	GLA8	M88633, ds+3, A → G, 506	9.6 → 9.2	–	+	…	++	+
2	Lisboa1	M88632, as-2, A → G, 102	12.7 → 4.5	–	+++	++	++	…
1	HP114	M88629, as-1, G → A, 220	13.1 → 5.5	5.8 → 5.8;216	+++	…	+++	+++
3	Pat 2	M88633, as-2, A → G, 385	13.8 → 5.6	4.7 → 6.9;385	+++	…	+++	+++
4	JH105	M88634, as-1, G → C, 50	11.4 → 4.1	–	+++	…	…	+++
7	Pat 2	M88642, ds+5, G → A, 4025	7.0 → 3.5	–	+	…	+++	+++
F9 gene								
12	UK206	K02402, ds+5, G → C, 3087	4.6 → 0.7	–	+++	…	+	…
46	HB782	K02402, ds+5, G → A, 13475	11.1 → 7.6	–	+	…	…	+
47	no ID	K02402, ds+2, T → C, 13472	11.1 → 3.7	4.1 → 5.7;13475	+++	…	+	…
48	HB44/52	K02402, ds+13, A → G, 20775	2.4 → 2.4	−2.4 → 3.7;20775	+	…	+	…
49	HB595/596/598/599/601	K02402, ds+13, A → G, 20775	2.4 → 2.4	−2.4 → 3.7;20775	+	…	+	+
50	HB686/692	K02402, ds+13, A → G, 20775	2.4 → 2.4	−2.4 → 3.7;20775	+	…	+	+
14	UK258	K02402, ds+1, G → T, 3	4.6 → −3.2	–	+++	…	++	…
18	UK785	K02402, as-3, T → G, 9640	12.5 → 8.2	4.4 → 13.1;9640	++	…	++	++
24	DK13	K02402, ds+3, G → C, 9457	7.1 → 2.7	–	++	…	++	…
30	HB53	K02402, as-1, G → A, 13356	5.3 → −2.3	–	+++	…	++	…
32	UK128	K02402, ds+5, G → A, 9672	7.6 → 4.1	–	+	…	++	…
36	Cardiff4 UK329	K02402, as-2, A → C, 20632	11.2 → 3.8	–	+++	++	++	…
37	Toronto1, HB48	K02402, as-2, A → G, 20632	11.2 → 3.0	−1.2 → 4.5;20632	+++	++	++	…
44	UK38	K02402, ds+7, A → G, 13477	11.2 → 11.5	–	+	++	…	…
45	UK218	K02402, ds+5, G → A, 13475	11.1 → 7.6	–	+	…	++	…
13	HB801	K02402, ds+5, G → C, 3087	4.6 → 0.7	–	+++	…	++	+++
20	HB216	K02402, as-1, G → C, 9642	12.5 → 5.2	–	+++	…	++	+++
26	UK74, UK77	K02402, ds+6, T → C, 9460	7.1 → 5.7	–	+	…	++	+++
33	HB20, HB364	K02402, ds+5, G → C, 9672	7.6 → 3.6	–	+	…	++	+++
41	HB675	K02402, as-1, G → C, 20633	11.2 → 3.9	0.2 → 6.2;20635	+++	…	++	+++
8	UK70	K02402, as-1, G → A, 9290	8.3 → 0.7	–	+++	…	+++	…
9	Spain	K02402, as-1, G → T, 9290	8.3 → −0.5	–	+++	…	+++	…
10	Malmö33	K02402, ds+5, G → A, 3087	4.6 → 1.1	–	+++	+++	++	…
11	H293	K02402, ds+1, G → C, 3083	4.6 → −5.2	–	+++	+++	+++	…
15	no ID	K02402, ds+1, G → T, 3083	4.6 → −3.2	–	+++	…	+++	…
16	HB304	K02402, ds+2, T → G, 3084	4.6 → 1.9	–	+++	…	+++	+++
17	Lisboa 1	K02402, ds+5, G → T, 3087	4.6 → 0.7	–	+++	+++	+++	…
19	UK242	K02402, as-1, G → A, 9642	12.5 → 4.9	–	+++	…	+++	…
21	Ribeirao Preto6	K02402, ds+1, G → A, 9455	7.1 → −5.7	–	+++	…	+++	…

Appendix. (continued)

No.	Patient database ID*	GenBank Accession number, Mutation, coordinate	Information-theory based prediction			Phenotypic characteristics		
			$R_{i(nat)} \rightarrow R_{i(mut)}$ (in bits)					
			Natural splice site	Cryptic splice site; coordinate	Reduction of mRNA-levels	Protein mass	Protein activity	Clinical severity (Eyster)
22	HB788	K02402, ds+1, G → C, 9455	7.1 → −5.5	–	+++	…	+++	+++
23	no ID	K02402, ds+5, G → A, 9459	7.1 → 3.6	–	+	…	+++	…
25	UK231	K02402, ds+6, T → C, 9460	7.1 → 5.7	–	+	+++	+++	…
27	HB197	K02402, as-3, A → G, 13354	5.3 → 3.5	−1.5 → 7.0;13354	+++	+	+++	+++
28	HB634	K02402, as-3, A → G, 13354	5.3 → 3.5	−1.5 → 7.0;13354	+++	…	+++	+++
29	UK289	K02402, as-2, A → T, 13355	5.3 → −2.2	–	+++	…	+++	…
31	HB680	K02402, as-1, G → T, 13356	5.3 → −3.1	–	+++	…	+++	+++
34	no ID	K02402, ds+2, T → C, 9669	7.6 → 0.1	2.9 → 2.9;9660	+++	+++	+++	…
35	Oxford 2	K02402, ds+2, T → G, 9669	7.6 → −0.6	2.9 → 2.9;9660	+++	+++	+++	…
38	HB540	K02402, as-2, A → G, 20632	11.2 → 3.0	−1.2 → 4.5;20632	+++	…	+++	+++
39	HB677	K02402, as-2, A → T, 20632	11.2 → 3.8	–	+++	…	+++	+++
40	Malmö11	K02402, as-1, G → C, 20633	11.2 → 3.9	0.2 → 6.2;20635	+++	+++	+++	…
42	H235	K02402, as-1, G → T, 20633	11.2 → 2.3	0.2 → 6.2;20635	+++	+++	+++	…
43	GER2489G	K02402, as-4, T → C, 20630	11.2 → 11.5	–	+	…	+++	…
51	UK290	K02402, ds+1, G → A, 20763	2.4 → −10.4	–	+++	…	+++	…
52	no ID	K02402, ds+1, G → T, 20763	2.4 → −5.4	–	+++	…	+++	…
53	HB214	K02402, ds+2, T → A, 20764	2.4 → −5.4	–	+++	…	+++	+++
54	no ID	K02402, ds+2, T → C, 20764	2.4 → −5.0	–	+++	…	+++	…
55	UK102	K02402, ds+2, T → G, 20764	2.4 → −5.8	–	+++	…	+++	…
56	HB259	K02402, as-2, A → G, 23326	11.7 → 1.1	0.5 → 8.1;23326	+++	…	+++	+++
57	UK171	K02402, as-1, G → A, 33003	11.2 → 3.6	–	+++	…	+++	…
58	HB102	K02402, as-1, G → C, 33003	11.2 → 3.9	–	+++	…	++	+++
59	DK40	K02402, ds+1, G → A, 23531	5.2 → −7.6	–	+++	+++	+++	…
60	no ID	K02402, ds+1, G → A, 23531	5.2 → −7.6	–	+++	+++	+++	…
61	HB642	K02402, ds+1, G → A, 23531	5.2 → −7.6	–	+++	…	+++	+++
62	no ID	K02402, ds-1, G → A, 23530	5.2 → 2.2	–	+++	+++	+++	…
63	2352(G)	K02402, ds-1, G → A, 23530	5.2 → 2.2	–	+++	…	+++	…
64	HB783	K02402, ds-1, G → A, 23530	5.2 → 2.2	–	+++	…	+++	+++
65	UK39	K02402, ds+1, G → T, 23531	5.2 → −2.6	–	+++	+++	+++	…
66	Oxford1	K02402, ds+1, G → T, 23531	5.2 → −2.6	–	+++	+++	+++	…
67	UK272	K02402, ds+2, T → C, 23532	5.2 → −2.3	–	+++	…	+++	…
68	Patient 2	K02402, as-1, G → A, 33786	4.7 → −2.9	–	+++	+++	+++	…
69	HB114	K02402, as-1, G → A, 33786	4.7 → −2.9	–	+++	…	+++	…
70	HB676	K02402, ds+1, G → C, 33119	9.2 → −2.5	–	+++	…	+++	+++
71	Caen3	K02402, ds+1, G → A, 33119	9.2 → −3.6	–	+++	+++	+++	…

*References to original reports for mutation designations are catalogued in the Human Mutation Database [5]. Symbols: as, indicates acceptor splice-site; clinical severity, clinical severity according to Eyster and colleagues [12]; coordinate, coordinate of cryptic splice site; ds, donor splice-site; patient-ID, identification number, cases with two or more ID-numbers include different individuals of the same family or identical individuals with two different ID numbers; $R_{i(nat)} \rightarrow R_{i(mut)}$, respective information contents of natural and mutant splice site in bits; +++, ++, + correspond to severe, moderate or mild, predicted reduction of functional mRNA or phenotypic characteristic; …, indicates that data were not reported in the study describing the mutation; - indicates mutations in which cryptic splice sites were not predicted.