Model for Alternative Splicing of Insulin Receptor in Myotonic Dystrophy Type 1

E. A. Tonevitsky and E. V. Trushkin

Translated from *Byulleten' Eksperimental'noi Biologii i Meditsiny*, Vol. 147, No. 6, pp. 716-720, June, 2009
Original article submitted April 8, 2009

Muscular dystrophy is a common multisystem disease, which results from the impairment of alternative splicing. An increase in the number of unstable CTG and CCTG repeats in untranslated regions of the DMPK and ZNF9 genes is followed by dysregulation of RNA-binding proteins. Further changes are followed by dysfunction of insulin receptors, membrane Cl^- channels, and other proteins. We developed a new mathematical model for the regulation of splicing of exon 11 in the IR gene, which describes the effect of various factors on alternative splicing.

Key Words: *myotonic dystrophy; model of splicing regulation*

Encoding gene regions alternate with noncoding DNA, which is not involved in protein expression. The majority of eukaryotic genes are transcribed in the form of pre-mRNA, which undergoes splicing into mRNA. Splicing is a two-step process. Noncoding sequences (introns) are cut off during sequential transesterification. This process is accompanied by the fusion of coding sequences (exons).

The selection of exons is often an alternative process. A fragment of pre-mRNA is removed (intron) or enters the mature mRNA (alternative exon). Alternative splicing is the major regulator of gene expression. This process results in the formation of several transcripts of one protein-coding gene, which improves the use of genetic information. More than 85% human protein-coding genes are characterized by alternative splicing.

The impairment of alternative splicing can cause developmental abnormalities and dysfunction of the organism. Published data show that 50% diseases due to gene mutations are associated with changes in alternative splicing [2].

The diseases (*e.g.*, myotonic dystrophy types 1 and 2) can be related to mutations in untranslated regions. Myotonic dystrophy (MD) is the most common type of muscular dystrophy in adults. This disorder is observed in 1 of 8000 people. The major clinical manifestations are myotonia, muscle weakness, cardiac arrhythmia, cardiomyopathy, hypersomnia, insulin resistance, testicular atrophy, cataract, forehead alopecia, and progressive mental deficiency [1]. Both types of dystrophy have the same symptoms, but are associated with various mutations in the separate and independent genes. MD type 1 is more frequently observed. This disease is associated with increased number of unstable CTG repeats in the 3'-untranslated region of the DMPK gene in chromosome 19q13.3. Under normal conditions, the number of CTG repeats varies from 5 to 37. However, the number of CTG repeats significantly increases in MD patients (50-5000). The number of repeats is unstable and can increase annually by 50-80 repeats [10]. A correlation exists between the number of repeats and onset of the disease (when the number of repeats does not exceed 400).

The MD type 2-inducing mutation is expansion of tetranucleotide (CCTG) repeat in intron 1 of the zinc finger protein 9 (ZNF9) gene on chromosome 3q. DNA diagnostic procedure shows typical expansion of tetranucleotide repeat CCTG in patients with MD type 2 (from 75 to 11,000 repeats; 5000 repeats on average). Complete complex configuration of the muta-

Russian Research Institute of Sport and Physical Education, Moscow, Russia. ***Address for correspondence:*** eugene.trushkin@vniifk.ru. E. V. Trushkin

0007-4888/09/1476-0772

tion in MD type 2 appears as follows: $(TG)_n(TCTG)_n(CCTG)_n$. Several regions with expanded repeat CCTG are interrupted by 2 motifs (GCTG and TCTG): $(TG)_n(TCTG)_n(CCTG)_n$ GCTG CCTG TCTG $(CCTG)_n$. The region of CCTG repeat expansion is increased in patients with MD type 2. This region increases by 2000 b.p. over 3 years of disease progression (Fig. 1) [9].

The majority of dominant diseases are associated with dysfunction of mutant proteins. The relationship between various symptoms of MD type 1 and existence of repeats in the noncoding region remained unknown. It was hypothesized that expression of mRNA and DMPK decreases under these conditions. However, DMPK-knockout mice were characterized by mild myopathy [13]. Now studies showed that transcripts containing CUG and CCUG repeats have a modulatory effect on the regulation and location of several RNA-binding proteins, including CUG-BP, MBNL1, MBNL2, MBNL3 [4], and hnRNP H [11].

CUG-BP protein expression was shown to increase in muscles of MD patients [12]. The authors hypothesized that CUG-BP binds to CUG repeats. However, this hypothesis was erroneous. Increased concentration and activity of CUG-BP is observed in patients with moderate and severe MD. The number of repeats in these patients is above 250 [3]. CUG repeats probably affect kinase C (PKC), which hyperphosphorylates and increases stability of CUG-BP [7]. Increased activity of CUG-BP has a modulatory effect on splicing of several proteins [11].

MBNL proteins (*e.g.*, MBNL1, MBNL2, and MBNL3) bind to abnormal hairpins that are formed by long CUG and CCUG repeats. This leads to a decrease in protein concentration and appearance of typical inclusions in the cytoplasm and nucleus [11,15]. The cells with cytoplasmic clusters are characterized by insignificant changes in alternative splicing of MD-associated genes [3]. Considerable amounts of MBNL1 and MBNL2 are expressed in the muscle tissue. MBNL3 was detected only in the placenta [11]. Model experiments showed that MBNL plays an important role in cell differentiation. Transgenic mice expressing increased number of CUG repeats have a malformation, which is symptomatically similar to MD. These abnormalities are associated with no changes in the mode of alternative splicing in gene exons, which normally occurs on days 2-20 of postnatal development. Changes in alternative splicing of MBNL during this period are accompanied by the decrease in cytoplasmic MBNL concentration and increase in nuclear MBNL concentration. MBNL-knockout mice have symptoms similar to MD. The results of model experiments on abnormal splicing are consistent with published data on aberrant splicing in MD patients [8]. Similarly to CUG-BP, MBNL is a RNA-binding protein that contributes to exon inclusion into mature RNA of some genes or exclusion of the exon in other genes.

Recent experiments revealed the existence of another regulatory mechanism for splicing of MD-associated genes. The protein hnRNP H can bind to mRNA. The binding site for hnRNP H is similar to that for CUG-BP. Under these conditions, hnRNP H and CUG-BP perform the same functions. It should be emphasized that binding of hnRNP H to MBNL is a RNA-independent process [11]. Model experiments on MBNL-knockout mice showed that the increase in MBNL expression by adeno-associated viruses is followed by a rapid and significant reduction of MD symptoms [6]. The observed changes are probably associated with binding of hnRNP H to MBNL in clusters of CUG repeats [11].

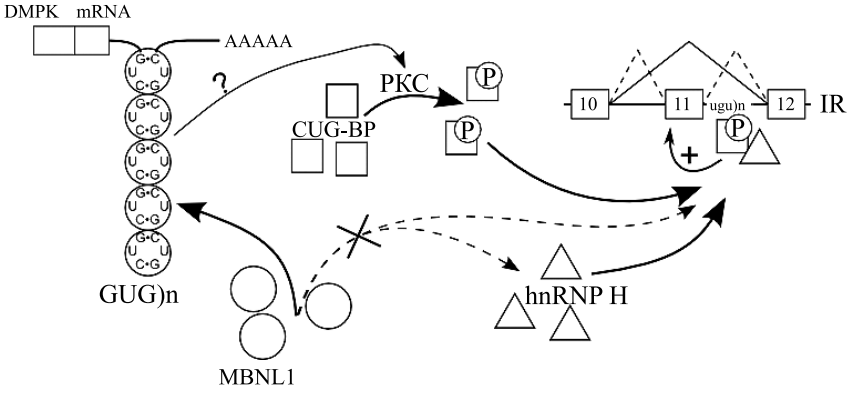

Fig. 1. Scheme for the regulation of splicing. RNA with CUG repeats is arranged in nuclear clusters due to attachment of double-strand CUG-binding proteins (*e.g.*, MBNL1). The increase in the concentration of RNA with CUG repeats is mediated by an unknown mechanism. These changes contribute to phosphorylation of CUG-BP by protein kinase, which increases protein activity and lifetime. The decrease in the concentration of MBNL is accompanied by an increase in the level of free hnRNP H. Binding of these proteins to U/G motifs of the regulatory sequences is followed by a shift in splicing to the embryonic tissue-like process. Normal, dotted line; pathology, solid line.

The main manifestation of type 1 and type 2 MD is clinical and electrical myopathy. Model experiments on transgenic mice showed that MD type 1 is accompanied by a decrease in transmembrane Cl^- conductance, impairment of CIC-1 protein splicing (responsible for chloride channel formation in muscle cells), and significant reduction of CIC-1 on the membrane surface.

Another manifestation of MD types 1 and 2 is insulin resistance, which determines the development of diabetes in patients. Due to dysregulation of pre-mRNA splicing for insulin receptor (IR) in skeletal muscle [12], exon 11 is excluded from mature RNA. It results in the prevalence of insulin-insensitive IR-A in cells. Similar changes were revealed in patients with MD type 2.

These data suggest that the major pathogenic effect of MD is related to impairment of alternative splicing. The mechanism of these disturbances appears as follows. During embryogenesis and early neonatal development, CUG-BP activates splicing of "fetal" exons that are erroneously preserved in adult patients with MD. CUG-BP concentration progressively decreases during maturation of healthy people. MBNL1 is located in the nucleus and affects exon splicing (shift to "adult-type" process). This mechanism is impaired in MD patients. MBNL1 is engulfed by CUG repeats and does not perform its specific function. These repeats activate kinase, which results in hyperphosphorylation and stabilization of CUG-BP (Fig. 2).

Our work was designed to develop a new mathematical model for the regulation of exon 1 splicing in the IR gene. This model evaluates the effect of various factors on alternative splicing, which holds much promise for high-efficacy therapy of MD.

MATERIALS AND METHODS

The ratios of IR gene pre-mRNA bound to the splicing enhancer and suppressor are designated as q_E and q_S, respectively. MBNL proteins are required for expression of the IR-B isoform (independently on the concentrations of CUG-BP and hnRNP H). The increase in the content of CUG-BP or hnRNP H in normal cells is followed by a decrease in IR-B expression. The binding site for MBNL proteins differs from that for CUG-BP and hnRNP H [14]. Therefore, suppressors and enhancers have an independent effect on the isoform ratio. q_E and q_S may be considered as the probabilities of independent events for "prohibition" and "permission" of exon 11 inclusion. Under these conditions the probability of isoform IR-B formation will appear as the overall probability of events with probabilities q_E and $(1-q_S)$: $p=q_E\times(1-q_S)$.

The interaction of a MBNL enhancer and CUG-BP and hnRNP H suppressor with IR gene pre-mRNA can be described by the kinetic model (**1**).

$$\begin{array}{l} S_1+M \underset{k_{-1}}{\overset{k_{+1}}{\Leftrightarrow}} S_1M \\ S_2+\mathrm{H} \underset{k_{-2}}{\overset{k_{+2}}{\Leftrightarrow}} S_2M \\ H+M \underset{k_{+3}}{\overset{k_{-3}}{\Leftrightarrow}} HM \end{array} \qquad (1)$$

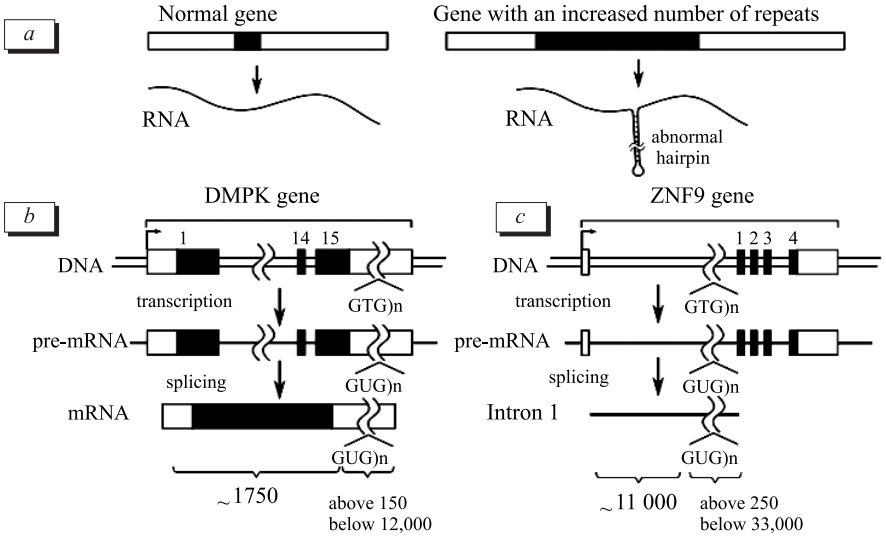

Fig. 2. Formation of abnormal hairpins during MD types 1 and 2. (*a*) The increased number of repeats is followed by the formation of abnormal hairpin on pre-mRNA. The length of this hairpin can significantly exceed the length of the remaining part in the pre-mRNA sequence. (*b*) Hairpin formation in the noncoding region of DMPK protein pre-mRNA. (*c*) Hairpin formation in intron 1 of the ZNF9 gene. Lines on the schemes of transcription and splicing designate introns. Black rectangles, exons; white rectangles, noncoding regions.

$$S_2+C \underset{k_{-4}}{\overset{k_{+4}}{\Leftrightarrow}} S_2C$$

where S_1 is the concentration of mRNA not containing MBNL; S2 is the concentration of mRNA not containing hnRNP H and CUG-BP; M is the concentration of free MBNL; S_1M is the concentration of a pre-mRNA—MBNL complex; C is the concentration of free CUG-BP; S_2C is the concentration of a pre-mRNA—CUG-BP complex; S_2H is the concentration of a pre-mRNA—hnRNP H complex; HM is the concentration of a hnRNP H—MBNL complex; and k_{+N} and k_{-N} are the constants of direct and reverse reactions, respectively.

This kinetic model is described by the differential equation system (**2**).

$$\begin{aligned}
\frac{d[S_1]}{dt} &= -[S_1][M]k_{+1}+[S_1M]k_{-1},\\
\frac{d[S_1]}{dt} &= -[S_2][H]k_{+2}+[S_2H]k_{-2}-[S_2][C]k_{+4}+[S_2C]k_{-4},\\
\frac{d[H]}{dt} &= -[S_2][H]k_{+1}+[S_2H]k_{-1}-[H][M]k_{+3}+[HM]k_{-3},\\
\frac{d[M]}{dt} &= -[S_1][M]k_{+1}+[S_1M]k_{-1}-[H][M]k_{+3}+[HM]k_{-3},\\
\frac{d[HM]}{dt} &= -[H][M]k_{+3}+[HM]k_{-3}, \qquad (\mathbf{2})\\
\frac{d[S_1M]}{dt} &= -[S_1][M]k_{+1}-[S_1M]k_{-1},\\
\frac{d[S_2C]}{dt} &= -[S_2][C]k_{+4}-[S_2C]k_{-4},\\
\frac{d[S_2H]}{dt} &= -[S_2H]k_{-2}+[S_2][H]k_{+2},\\
\frac{d[C]}{dt} &= -[S_2C]k_{-4}+[S_2][C]k_{+4}.
\end{aligned}$$

Under steady-state conditions, the system (**2**) appears as (**3**) in moving from the reaction rate to the dissociation constant.

$$\begin{aligned}
[S_1] &= S_0-[S_1M],\\
[S_2] &= S_0-[S_2H]-[S_2C],\\
[M] &= M_0-[S_1M]-[HM],\\
[H] &= H_0-[S_2H]-[HM],\\
[C] &= C_0-[S_2C], \qquad (\mathbf{3})\\
0 &= -[S_1M]+[S_1M]Kd_1,\\
0 &= -[S_2][H]+[S_2H]Kd_2,\\
0 &= -[H][M]+[HM]Kd_3,\\
0 &= -[S_2][C]+[S_2C]Kd_4,
\end{aligned}$$

where S_0, M_0, C_0, and H_0 are the total concentrations of pre-mRNA, MBNL, CUG-BP, and hnRNP H, respectively; and Kd_1, Kd_2, Kd_3, and Kd_4 are the dissociation constants, respectively.

After substitution of the first five equations into the remaining four equations and solution of the system (**3**) relative to $[S_1M]$, $[S_2C]$, [HM], and [S2H], the probability of exon inclusion appears as follows: $p=[S_1M]\times([S_0]-[S_2C])/[S_0]^2$.

During the development of this model, the following assumptions were made.

First, allosteric binding of MBNL to the IR mRNA site is not taken into account (Hill constant 1.4) [15]. It is assumed that all substances have one binding site.

Second, various proteins of the MBNL group are considered as one protein.

RESULTS

Published data [1] were used to verify the proposed model. The IR-A/IR-B isoform ratio was compared in normal cells and MD type 1 myoblasts and fibroblasts that had various effects on the expression of RNA-binding proteins by small interfering RNA and recombinant adenovirus vectors (Table 1, Fig. 3).

The dissociation constants appeared as follows: Kd_1=0.02 μM; Kd_2=0.6 μM; Kd_3=0.02 μM; and Kd_4=

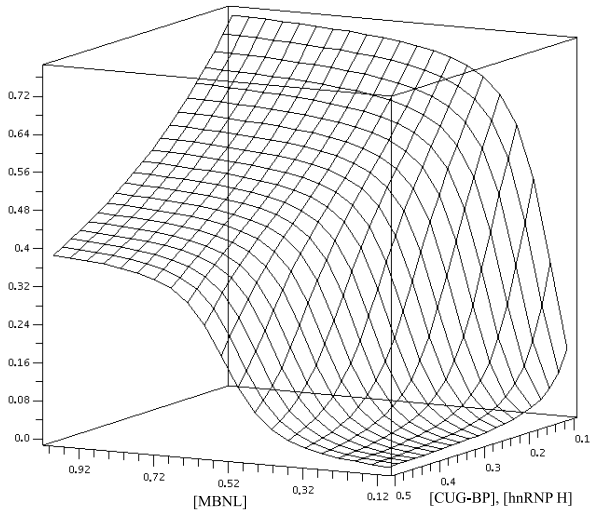

Fig. 3. Dependence of the IR-B isoform ratio on the concentration of RNA-binding proteins.

TABLE 1. Comparative Analysis of Published Data and Results of Modeling

Treatment	Normal		MD type 1	
	published data	model	published data	model
Without treatment	0.45	0.41	0.06	0.03
CUG-BP+	0.10	0.27	—	—
hnRNP H+	0.15	0.18	—	—
MBNL+	0.45	0.52	0.25	0.06
CUG-BP-	0.45	0.68	0.10	0.04
hnRNP H-	0.45	0.52	0.05	0.16
MBNL-	0.05	0.08	—	—
CUG-BP- hnRNP H+	0.45	0.26	—	—
CUG-BP+ hnRNP H-	0.25	0.33	—	—
CUG-BP- hnRNP H-	—	—	0.10	0.40
MBNL+ hnRNP H+	0.20	0.46	—	—
MBNL+ hnRNP H-	—	—	0.30	0.28

Note. Signs "+" and "-" in the first column designate increase (by 2 times) and decrease (up to 5-10%) in the concentration of RNA-binding proteins.

0.6 μM. The total concentration of pre-mRNA was taken as S_0=0.4 μM. The total concentration of MBNL was considered in the M_0 range of (0; 1) μM. The total concentrations of CUG-BP and MBNL were assumed to be equal and analyzed in the range H_0=C_0 (0; 0.5) μM. The constants were selected to provide the consistency of this model to published data. The value p was calculated in these ranges of RNA-binding protein concentration.

The results of modeling showed that this model is not completely consistent with experimental data. These findings suggest the cooperative binding of CUG-BP to pre-mRNA. A sigmoid dependence of the concentration of a CUG-BP—pre-mRNA complex on the concentration of CUG-BP is probably provided by another regulatory mechanism.

REFERENCES

1. J. D. Brook, M. E. McCurrach, H. G. Harley, *et al*., *Cell*, **68**, No. 4, 799-808 (1992).
2. L. Cartegni, S. L. Chew, and A. R. Kreiner, *Nat. Rev. Genet*., **3**, No. 4, 285-298 (2002).
3. W. Dansithong, C. M. Wolf, P. Sarkar, *et al*., *PLoS ONE*, **3**, No. 12 (2008). [www.plosone.org].
4. M. Fardaei, M. T. Rogers, H. M. Thorpe, *et al*., *Hum. Mol. Genet*., **11**, No. 7, 805-814 (2002).
5. T. H. Ho, N. Charlet-B, M. G. Poulos, *et al*., *EMBO J*., **23**, No. 15, 3103-3112 (2004).
6. R. N. Kanadia, J. Shin, Y. Yuan, *et al*., *Nat. Acad. Sci*., **103**, No. 31, 11,748-11,753 (2006).
7. N. M. Kuyumcu-Martinez, G. Wang, and T. A. Cooper, *Mol. Cell*, **28**, No. 1, 68-78 (2007).
8. X. Lin, J. W. Miller, A. Mankodi, *et al*., *Hum. Mol. Genet*., **15**, No. 13, 2087-2097 (2006).
9. C. L. Liquori, Y. Ikeda, M. Weatherspoon, *et al*., *Am. J. Hum. Genet*., **73**, No. 4, 849-862 (2003).
10. L. Martorell, D. G. Monckton, J. Gamez, *et al*., *Hum. Mol. Genet*., **7**, No. 2, 307-312 (1998).
11. S. Paul, W. Dansithong, D. Kim, *et al*., *EMBO J*., **25**, No. 18, 4271-4283 (2006).
12. A. V. Philips, L. T. Timchenko, and T. A. Cooper, *Sci*., **280**, 737-741 (1998).
13. S. Reddy, D. B. Smith, M. M. Rich, *et al*., *Nat. Genet*., **13**, No. 3, 325-335 (1996).
14. R. S. Savkur, A. V. Philips, and T. A. Cooper, *Nat. Genet*., **29**, No. 1, 40-47 (2001).
15. M. B. Warf and J. A. Berglund, *RNA*, **13**, No. 12, 2238-2251 (2007).