Analysis of highly conserved regions of the 3’UTR of MECP2 gene in patients with clinical diagnosis of Rett syndrome and other disorders associated with mental retardation

Mónica Santos\(^a,b\), Jin Yan\(^c\), Teresa Temudo\(^d\), Guiomar Oliveira\(^e\), José Pedro Vieira\(^f\), Jinong Fen\(^c\), Steve Sommer\(^c\) and Patrícia Maciel\(^a,\ast\)

\(^a\)Life and Health Sciences Research Institute (ICVS), School of Health Sciences School, University of Minho, Braga, Portugal
\(^b\)ICBAS, University of Porto, Porto, Portugal
\(^c\)Centre for Molecular Diagnosis, City of Hope Medical Centre, Duarte, CA, USA
\(^d\)HGSA, Porto, Portugal
\(^e\)Hospital Pediátrico de Coimbra, Centro Hospitalar de Coimbra, Coimbra, Portugal
\(^f\)Hospital Fernando Fonseca, Amadora, Portugal

Abstract. In this work we explored the role of the 3’UTR of the MECP2 gene in patients with clinical diagnosis of RTT and mental retardation; focusing on regions of the 3’UTR with almost 100% conservation at the nucleotide level among mouse and human. By mutation scanning (DOVAM-S technique) the MECP2 3’UTR of a total of 66 affected females were studied. Five 3’UTR variants in the MECP2 were found (c.1461+9G>A, c.1461+98insA, c.2595G>A, c.9961C>G and c.9964delC) in our group of patients. None of the variants found is located in putative protein-binding sites nor predicted to have a pathogenic role. Our data suggest that mutations in this region do not account for a large proportion of the RTT cases without a genetic explanation.

Keywords: Autism, neurodevelopment, 3’untranslated region

1. Introduction

Mutations in the MECP2 gene are associated with Rett syndrome (RTT, OMIM # 312750) [1], a disorder affecting mainly females, presenting with mental retardation, autistic features and a movement disorder. The disease is characterized by an apparently initial normal developmental period, followed by a reduction in growth and loss of motor and cognitive skills around the age of 6 to 18 months [15,16]. Mutations in the coding region of MECP2 are identified in around 80–90% of classical RTT patients (presenting the above described phenotype), and 30% of variant forms of the disease, which includes more severe presentations such as neonatal encephalopathy and milder presentations such as the preserved speech variant or forme fruste [16,19]. Thus, the cause of disease in a significant proportion of patients with RTT remains unknown. It has been proposed that mutations in non-coding regions of MECP2, untranslated regions (5’ and 3’UTR) and introns, or in other genes may be the unidentified cause of the disorder in these cases.

The role of the 3’UTR of a gene might be the regulation of its expression at different levels, such as the
“translatability” of an mRNA, its stability, nuclear export or the sub-cellular localization of the translation event, thus affecting the levels of the resulting protein (for a review see [7]). The latter mechanisms have been demonstrated in neurons for the immediate-early gene Arc (apoptosis repressor with CARD) and the protein αCamKII (calcium/calmodulin-dependent protein kinase II-alpha) which are translated locally at/near the synapse [5].

Additionally, the 3′UTR of the CamKII gene has been linked to the regulation of activity-dependent CamKII protein expression, via glutamate N-methyl-D-aspartate glutamate receptor (NMDAR) activation. This regulation is on the basis of synaptic plasticity, learning and memory formation [28], all of which are disturbed in RTT.

A study based on the human gene mutation database (HGMD) estimated that around 0.2% of the disease-associated mutations reside in the regulatory regions of the 3′UTR of genes [6]. Mutations in the 3′UTR have been identified as the genetic cause of a number of diseases with a neurological component: (1) myotonic dystrophy, a neuromuscular disorder characterized by hypotonia, mental retardation and muscle development defects, due to a CTG repeat expansion in the 3′UTR of the dystrophia myotonica protein kinase gene (DMPK) [12]; (2) IPEX syndrome (immune dysfunction, polyendocrinopathy, enteropathy, X-linked), caused by a mutation within the first polyadenylation signal of the forkhead box P3 gene (FOXP3) [3]; (3) Fukuyama-type congenital muscular dystrophy, presenting with mental retardation and brain defects, caused by a defect in the Fukutin gene [17] and (4) familial Danish dementia, caused by a decamer duplication in the integral membrane protein 2B gene (BRI) [25]; finally, (5) a nucleotide change found in the 3′ regulatory region of the cyclin-dependent kinase 5, regulatory subunit 1 gene (CDK5R1), was also suggested to be a cause of non-syndromic mental retardation [24]. The mechanisms involved are variable and these mutations cause pathology either by affecting (1) the splicing of other genes as in the case of myotonic dystrophy [20], (2) the mRNA maturation as is the case in IPEX syndrome [2], (3) the mRNA stability as in the case of Fukuyama-type congenital dystrophy [17] or the expression levels as it happens in familiar Danish dementia [25]. However, other variants at the 3′UTR of other genes were not found to be associated with disease, as is the case of the CAA insertion polymorphism in the NOGO gene in schizophrenia and bipolar disorder [14] and the B7-1 gene in patients with multiple sclerosis [27] among many others.

The MECP2 gene has eight different transcripts that result from alternative splicing and different sites of polyadenylation and one of the longest known 3′UTR tails, with 8.5-kb [9]. The longest transcript is more than 10 kb long and has several blocks of highly conserved residues between the human and mouse genomes (Fig. 1). This argues in favour of a potential regulatory role of this 3′UTR in the function of the MeCP2 protein, in different cell types and at different developmental times. The longest transcript is also generally described as the predominant form in the brain [9,10,21].

In this work we explored the potential role of the extremely large and conserved 3′UTR of the MECP2 gene in patients with clinical diagnosis of RTT and mental retardation: we performed a bioinformatics analysis in order to identify putative protein-binding sites and we searched for mutations in selected regions of this 3′UTR, highly conserved among mouse and human, in an attempt to identify sequence variations leading to disease.

2. Subjects and methods

2.1. Bioinformatics

We used the ESResearch program (http://ast.bioinfo.tau.ac.il/ESR.htm) to search the MECP2 3′UTR for potential binding sites of RNA-binding proteins [11,13,26,29] and the BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/) to identify regions at high conservation between human and mouse MECP2 3′UTR.

2.2. Subjects

We have included in the study a total of 66 affected females who had already been studied for MECP2 mutations in its coding region and exon-intron boundaries, including detection of large rearrangements by robust dosage-PCR method (33 were positive for MECP2 mutations in the coding region of the gene), 51 with a clinical diagnosis of RTT, 35 classical and 16 atypical, according to the reviewed criteria of Hagberg and colleagues [15,16], 13 with mental retardation with autism and 2 with an Angelman syndrome-like clinical presentation. Whenever possible, both parents were also included in the study. A control population of 40 males and 83 females was used in order to characterize the frequency of the new variants found. All participants or their legal representatives gave informed consent.
for this study. The study conforms with The Code of Ethics of the World Medical Association (Declaration of Helsinki) and was reviewed and approved by the research ethics committee of the Life and Health Sciences Research Institute.

2.3. Molecular analysis of the MECP2 3'UTR

Genomic DNA was extracted from the peripheral blood of the patients and available parents using the Puregene DNA extraction system (Gentra).

Selection of the non-coding regions of the MECP2 3'UTR to be analysed was performed based on conservation among species (86%–98% conservation at the nucleotide level), in a total of 9 blocks (NM_004992: c.1607–c.1956, c.2561–c.2891, c.3551–c.3805, c.3768–c.4128, c.6851–c.7029, c.7116–c.7436, c.8372–c.8645, c.8607–c.8872 and c.9844–c.10182). These 9 blocks of conservation include three of the four polyadenylation signals and their nearby regulatory regions of the MECP2 gene (∼1.8, ∼7 and ∼10) (Fig. 1).

The 3'UTR blocks were scanned for mutations with the Detection Of Virtually All Mutations-SSCP (DOVAM-S) technique [18,23]. The different DNA segments were first amplified robotically, pooled, denatured and electrophoresed under five nondenaturing conditions varying in gel matrix, buffer, temperature, and additive: (1) 10% PAGE+/30 mM Tricine/Triethanolamine at 20°C, (2) 10% HR1000/30 mM Tricine/Triethanolamine at 4°C, (3) 10% PAGE+/TBE/5% glycerol at 20°C, (4) 10% HR1000/TBE/2.5% glycerol at 4°C, (5) 10% PAGE+/30 mM Capso at 4°C. PCR products with mobility shifts were sequenced with the ABI model 377 (Perkin-Elmer Model 377, Norwalk, CT) and nucleotide alterations were analyzed. Sequence changes were confirmed by re-amplification of genomic DNA and sequencing in the opposite direction. The primers used for amplification of the 3'UTR of MECP2 as well as the PCR conditions used, are listed in Table 1.

For the DNA amplification of the different 3'UTR blocks, a final volume of 25 µl PCR mixture (2.5 mM MgCl₂, 0.2 mM dNTP, 2.5 µM of each primer and 2U of AmpliTaq Gold (Applied Biosystems)) was used. The thermal cycling profile consisted of an initial denaturation for 10 min at 94°C, 35 cycles of a denaturation for 15 sec at 94°C, annealing for 30 sec at 55 °C, extension for 1 min at 72°C, and a final extension for 10 min at 72°C.

2.4. Allele specific-PCR (AS-PCR)

The presence of the MECP2 3'UTR variants c.2595G>A and c.9961C>G was tested in a control population by AS-PCR. The primers used for allele specific amplification of the MECP2 3'UTR variants, as well as the PCR conditions used, are listed in Table 1.

Two PCR mixtures for each variant were prepared in order to amplify either the normal or the mutated allele, in a final volume of 25 µl PCR mixture that consisted of 0.5 mM MgCl₂, 0.2 mM dNTP, 0.8 µM of each primer pair (either for the normal allele or for the mutated allele) and 1.5 U of Taq DNA polymerase (Fermentas). The thermal cycling profile consisted of an initial denaturation for 5 min at 95°C, 35 cycles of a denaturation for 1 min at 95°C, annealing for 1 min at Ta °C (specific for each variant, see Table 1), extension for 1 min at 72°C, and a final extension for 5 min at 72°C. PCR products were electrophoresed in a 2% agarose gel, and visualized under UV light.
Table 1

| Gene    | Region       | Primer | Sequence (5’>3’) | Size (bp) | Tm (°C) | Ta (°C) |
|---------|--------------|--------|------------------|-----------|---------|---------|
| MECP2   | 3' UTR       | MD14   | CGTGACCGAGAGAATGIC | 350       | 60      | 55      |
|         |              | MU14   | CGGAGGGAGGAGTGCC  |           | 58      |         |
|         |              | MD15   | CCAGTTATCTTCAATCTCC | 331       | 58      | 55      |
|         |              | MU15   | AGAATGAAAGAGAATGAA |           | 58      |         |
|         |              | MD16   | GCTTAAAGGCTTGCTTG  | 255       | 60      | 55      |
|         |              | MU16   | CAGCAGCTCACATGGGACA | 60        |         |         |
|         |              | MD17   | CCAGAAAACCCACAGGCA | 361       | 60      | 55      |
|         |              | MU17   | TTGGGCTTGAGGGATTTTG | 58        |         |         |
|         |              | MD18   | GCAGATGAAAGGCTTGCA | 274       | 58      | 55      |
|         |              | MU18   | GAAAAAAGGGAGAAGAAG | 60        |         |         |
|         |              | MD19   | CTGGCATCCTGACACTAAAC | 321       | 58      | 55      |
|         |              | MU19   | ACCAGAACCCTGGGACC  | 58        |         |         |
|         |              | MD20   | GGAGTGCAGAAGGCTTGCA | 274       | 58      | 55      |
|         |              | MU20   | GAAAACCCCGAGAAGAAG | 60        |         |         |
|         |              | MD21   | TTCCCTTTTGGCCCTTACTGTC | 266       | 58      | 55      |
|         |              | MU21   | GAAAGAAGAAGCTCTAGAAAT |         | 58      |         |
|         |              | MD22   | GCCGGGACACACTTAGC   | 339       | 60      | 55      |
|         |              | MU22   | AAATTATAGGCAAACCTTATT | 58        |         |         |
| MECP2   | variant      | c.2595G>A | 2595G>AmF | GCGCTTGTCTTAAAAGTGGA | 148   | 56      | 59      |
|         |              | 2595G>AmF | GCCTTGTCTTTAAAAGTGGA |           | 54      |         |
|         |              | MU15    | AGAAGTGAAAGGATGAAATG | 58        |         |         |
| c.9961C>G | 9961C>GmF | ACGTTTCAAAAAACCCCCCC | 161   | 64      | 61      |
|         |              | 9961C>GmF | ACGTTTCAAAAAACCCCCCC |           | 64      |         |
|         |              | MU23    | CTGCCCACTTTTCAATGT  | 60        |         |         |

**MECP2** variant

### 3. Results

#### 3.1. Bioinformatics

By computational analysis we identified several putative binding sites of trans-acting factors in the **MECP2** 3' UTR. We identified binding sites for the neurooncological ventral antigen (NOVA), the arginine/serine-rich splicing factors SRp40 and SRp55 and several heterogeneous nuclear ribonucleoproteins (hnRNPs), including the polypyrimidine tract-binding protein (PTB); all these proteins are known to be involved in the regulation of splicing and/or polyadenylation events. Although biochemical evidence of binding needs to be obtained, these findings reinforce the potential biological relevance of this untranslated region and prompted us to proceed to its genetic study.

#### 3.2. **MECP2** 3' UTR variants

Variants in the 3'UTR of the **MECP2** gene were searched in 66 Portuguese patients with a clinical di-
agnosia of RTT or mental retardation by Detection Of Virtually All Mutations – SSCP (DOVAM-S) [18,23]. We identified 5 alterations in the 3’UTR of MECP2 gene: c.1461+9G>A, c.1461+9insA, c.2595G>A, c.9961C>G and c.9964delC (NM_004992.2, Fig. 1 and Table 2). None of the variants is localized in any of the identified putative binding sites. The 1461+9G>A and the 1461+9insA variants (identified by the direct sequencing of exon/intron boundaries in our previous work) were already described in the literature as polymorphisms (see MECP2 mutation database, http://mecp2.chw.edu.au). The c.9964delC variant was previously detected in a Portuguese control population [8], and the c.2595G>A and c.9961C>G variants were identified for the first time in this study and were not found in 218 X chromosomes of a Portuguese control population. The variants c.9961C>G and c.9964delC were both present in the same patient (Table 2); the c.9961C>G variant was also present in the unaffected father of the patient, hence this variation must not be a pathogenic mutation, and the variant c.9964delC was present in the unaffected mother of the patient who had an balanced X-chromosome inactivation pattern in her lymphocytes. We could not test the c.2595G>A variant in the parents of the patient, since their DNA was not available. However, this patient was shown to have another causal mutation in trans (Table 2), a large rearrangement of the MECP2 gene [22], and so this variant is most likely not a pathogenic alteration. Since it is present only in this one patient, in whom a severe clinical presentation was expected, we cannot conclude whether this variant may behave as a modifier of the phenotype.

4. Discussion

The long and highly conserved 3’UTR of the MECP2 gene suggested that mutations in this region could exist and underlie a percentage of the molecularly unexplained RTT cases; from 20% to 70% in classical and atypical cases, respectively. The most interesting RTT group to study was doubtlessly the one without MECP2 mutations in the coding region, although 3’UTR variations could also be modulators of the phenotype caused by MECP2 coding region mutations. Our data suggest that mutations in this region do not account for a large proportion of the RTT cases without a genetic explanation.

In a previous study, a 300-bp segment of the 3’UTR around the 10 kb polyadenylation signal of the MECP2 gene was scanned for mutations in RTT patients and no pathogenic variants were found [4]. Shibayama and colleagues [23] studied a heterogeneous group of patients presenting schizophrenia, autism and other psychiatric disorders and reported that 3’UTR variants in the gene MECP2 seemed to be more frequent in autism patients than in the general population. However, the screening of the entire 3’UTR of MECP2 in an autistic Portuguese population did not show any pathogenic variant that could be responsible for the pathology, and no heightened representation of 3’UTR sequence variants was found in this patient population as compared to controls [8].

In our perspective, the contribution of the MECP2 3’UTR to RTT aetiology cannot be totally excluded without studying a larger number of RTT patients of different populations; analysis should focus in these “blocks” of high conservation and with putative regulatory protein binding sites, for which we have established a large-scale scanning method. Additionally, other methods such as evaluation of mRNA levels and stability and binding of specific proteins by RNA protection assays may also be considered.

Acknowledgements

Mónica Santos is supported by Fundação para a Ciência e Tecnologia (FCT, Portugal) with the PhD fellowship SFRI/BD/9111/2002. Research in Rett syndrome is supported by FSE/FEDER and FCT, grant POCTI 41416/2001.

References

[1] R.E. Amir, I.B. Van den Veyver, M. Wan, C.Q. Tran, U. Francke and H.Y. Zoghbi, Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2, Nat Genet 23 (1999), 185–188.
[2] C.L. Bennett, M.E. Brunkow, F. Ramsdell, K.C. O’Briant, Q. Zhu, R.L. Fuleihan, A.O. Shigeoka, H.D. Ochs and P.F. Chance, A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA->AAUGAA) leads to the IPEX syndrome, Immunogenetics 53 (2001), 435–439.
[3] C.L. Bennett, J. Christie, F. Ramsdell, M.E. Brunkow, P.J. Ferguson, L. Whitesell, T.E. Kelly, F.T. Saulsbury, P.F. Chance and H.D. Ochs, The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3, Nat Genet 27 (2001), 20–21.
[4] V. Bourdon, C. Philippe, O. Labrune, D. Amsallem, C. Arnould and P. Jonveaux, A detailed analysis of the MECP2 gene: prevalence of recurrent mutations and gross DNA rearrangements in Rett syndrome patients, Hum Genet 108 (2001), 43–50.
[5] C.R. Bramham and E. Messaoudi, BDNF function in adult synaptic plasticity: the synaptic consolidation hypothesis, *Prog Neurobiol* **76** (2005), 99–125.

[6] J.M. Chen, C. Ferec and D.N. Cooper, A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes I: general principles and overview, *Hum Genet* **120** (2006), 1–21.

[7] B. Conne, A. Stutz and J.D. Vassalli, The 3′ untranslated region of messenger RNA: A molecular ‘hotspot’ for pathology? *Nat Med* **6** (2000), 637–641.

[8] A.M. Coutinho, G. Oliveira, C. Katz, J. Feng, J. Yan, C. Yang, C. Marques, A. Ataide, T.S. Miguel, L. Borges, J. Almeida, C. Correia, A. Currais, C. Bento, L. Mota-Vieira, T. Temudo, M. Santos, P. Maciel, S.S. Sommer and A.M. Vicente, MECP2 coding sequence and 3′UTR variation in 172 unrelated autistic patients, *Am J Med Genet B Neuropsychiatr Genet* **144** (2007), 475–483.

[9] J.F. Coy, Z. Sedlacek, D. Bachner, H. Delius and A. Pousta, A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3′-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression, *Hum Mol Genet* **8** (1999), 1253–1262.

[10] M.D’Esposito, N.A. Quaderi, A. Ciccodicola, P. Bruni, T. Esposito, M. D’Urso and S.D. Brown, Isolation, physical map, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2, *Mamm Genome* **7** (1996), 533–535.

[11] L.P. Dunne, J. Dubel, G.A. Nasser, T. Ashizawa, P. de Jong and L.E. Vons, Detection of virtually all mutations-SSCP (DOV-AM-S): a rapid method for mutation scanning with virtually 100% sensitivity, *Biotechniques* **26** (1999), 932–934, 938–940, 932–934.

[12] J. Feng, C. Ferec and D.N. Cooper, A systematic analysis of mutations-SSCP (DOV-AM-S): a rapid method for mutation scanning with virtually 100% sensitivity, *Biotechniques* **26** (1999), 932–934, 938–940, 932–934.

[13] A.K. Percy, Rett syndrome: clinical correlates of the newly discovered gene, *Brain Dev* **23** (Suppl 1) (2001), S202–S205.

[14] T.P. Ramun and J.W. Day, Myotonic dystrophy: RNA pathogenesis comes into focus, *Am J Hum Genet* **74** (2004), 793–804.

[15] K. Reichwald, J. Thiesen, T. Wieze, J. Weitzel, W.A. Poustka, A. Rosenthal, M. Plattner, W.H. Stratling and P. Kioschis, Comparative sequence analysis of the MECP2 locus in human and mouse reveals new transcribed regions, *Mamm Genome* **11** (2000), 182–190.

[16] J. Shi, A. Shibayama, Q. Liu, V.Q. Nguyen, J. Feng, M. Santos, T. Temudo, P. Maciel and S.S. Sommer, Detection of heterozygous deletions and duplications in the MECP2 gene in Rett syndrome by Robust Dosage PCR (RD-PCR), *Hum Mutat* **25** (2005), 505.

[17] A. Shibayama, E.H. Cook, Jr., J. Feng, C. Glanzmann, J. Yan, N. Craddock, I.R. Jones, D. Goldman, L.L. Heston and S.S. Sommer, MECP2 structural and 3′-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism, *Am J Med Genet B Neuropsychiatr Genet* **128** (2004), 50–55.

[18] M. Venturin, S. Moncini, V. Villa, S. Russo, M.T. Bonati, L. Larizza and P. Riva, Mutations and novel polymorphisms in coding regions and UTRs of CDK5R1 and OMG genes in patients with non-syndromic mental retardation, *Neurogenetics* **7** (2006), 59–66.

[19] B. Hagberg, J. Aicardi, K. Dias and O. Ramos, A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett’s syndrome: report of 35 cases, *Eur J Pediatr Neurol* **6** (2002), 293–297.

[20] B.G. Weinshenker, D.D. Hebrink, C. Klein, E.J. Atkinson, P.C. O’Brien and C.T. McMurray, Genetic variation in the B7-1 gene in patients with multiple sclerosis, *J Neuroimmunol* **105** (2000), 184–188.

[21] D.G. Wells, X. Dong, E.M. Quinlan, Y.S. Huang, M.F. Bear, J.D. Richter and J.R. Fallon, A role for the cytoplasmic polyadenylation element in NMDA receptor-regulated mRNA translation in neurons, *J Neurosci* **21** (2001), 9541–9548.

[22] X.H. Zhang and L.A. Chasin, Computational definition of sequence motifs governing constitutive exon splicing, *Genes Dev* **18** (2004), 1241–1250.