Structure-functional studies on lncRNA MEG3

Tina Uroda

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Tina URODA

Thèse dirigée par **Marco MARCIA**

préparée au sein du **Laboratoire laboratoire européen de biologie moléculaire**
dans **l'École Doctorale Chimie et Sciences du Vivant**

Caractérisation structurale et fonctionnelle de l'ARN long non codant MEG3

Structure-functional studies on lncRNA MEG3

Thèse soutenue publiquement le **9 mai 2019**,
devant le jury composé de :

Monsieur MARCO MARCIA
DIRECTEUR DE RECHERCHE, EMBL GRENOBLE, Directeur de thèse
Monsieur ALBERTO INGA
PROFESSEUR, UNIVERSITE DE TRENTE - ITALIE, Rapporteur
Madame MAITE HUARTE
DIRECTRICE DE RECHERCHE, UNIVERSITE DE NAVARRE-PAMPELUNE-ESPAGNE, Rapporteur
Monsieur CLAIRE VOURC'H
PROFESSEUR, UNIVERSITE GRENOBLE ALPES, Président
Madame REINI F. LUCO
DIRECTRICE DE RECHERCHE, CNRS DELEGATION LANGUEDOC-ROUSSILLON, Examinateur
Monsieur JANOSCH HENNIG
DIRECTEUR DE RECHERCHE, EMBL HEIDELBERG ALLEMAGNE, Examinateur

Table of Contents

List of Abbreviations

AFM	atomic force microscopy
AUC	analytical ultracentrifugation
BrdU	5-bromo-2’-deoxyuridine
cAMP	cyclic adenosine monophosphate
$CC_{Spearman}$	Spearman's correlation coefficient
cDNA	complementary deoxyribonucleic acid
$C_{K1/2}$	potassium concentration at which MEG3v1 reaches half compaction
$C_{Mg1/2}$	magnesium concentration at which MEG3v1 reaches half compaction
DEPC	diethyl pyrocarbonate
Dlk1	delta like non-canonical Notch ligand 1
DLS	dynamic light scattering
D_{av}	average diameter
D_{max}	maximal distance between two atoms
DMR	differentially methylated regions
DMS	dimethyl sulfate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNMT	deoxyribonucleic acid methyl transferase
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
D1-5	domains 1-5
EDTA	ethylenediaminetetraacetic acid
EdU	5-ethynyl-2'-deoxyuridine
E1-12	Exons 1-12
FAM	fluorescein amidite
FISH	fluorescent *in situ* hybridization
Frag-Seq	fragmentation sequencing
GABA	gamma-aminobutyric acid

Gtl2	gene trap locus 2
HOTAIR	HOX transcript antisense ribonucleic acid
HRF	hydroxyl radical footprinting
H3K27	lysine 27 on histone H3
H1-51	helices 1-51
icSHAPE-seq	*in vivo* click SHAPE sequencing
IG	intergenic
JARID2	Jumonji and AT-rich interaction domain containing 2
J1-16	multi-way junctions 1-16
kb	kilobase
Kd	dissociation constant
kDa	kilodalton
kHz	kilohertz
KO	knockout
lncRNA	long non coding ribonucleic acid
MALAT1	metastasis associated lung adenocarcinoma transcript 1
MDM2	mouse double minute 2 homolog
MEG3	maternally expressed gene 3
miRNA	micro ribonucleic acid
MOPS	3-(N-morpholino)propanesulfonic acid
MS	mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MWCO	molecular weight cut off
M1-5	motifs 1-5
NCBI	National Center for Biotechnology Information
ncRNA	non coding ribonucleic acid
NEB	New England Biolabs
NFA	non-functioning pituitary adenomas
n_{Hill}	Hill coefficient

NMIA	N-methylisatoic anhydride
NMR	nuclear magnetic resonance
NSCLC	non-small cell lung cancer
nt	nucleotide
PARS	parallel analysis of RNA structure
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PRC2	*Polycomb* repressive complex 2
PSD	power spectral density
PTC	papillary thyroid carcinoma
qRT-PCR	quantitative real-time polymerase chain reaction
R	SHAPE reactivity
Rac1	ras-related C3 botulinum toxin substrate 1
Rb	retinoblastoma
RbAp46/48	retinoblastoma protein associated protein 46/48
RBR	ribonucleic acid binding region
R_g	radius of gyration
R_h	hydrodynamic radii
R_{median}	median SHAPE reactivity
RNA	ribonucleic acid
roX1/2	RNA on the X 1/2
rpm	rotations per minute
RT	room temperature
S	Shannon entropy
S_{ave}	average Shannon entropy
S_{median}	median Shannon entropy
SEC	size-exclusion chromatography
SEC-MALLS	size-exclusion chromatography coupled to multi-angle laser light scattering
SEC-SAXS	size exclusion chromatography coupled to small angle X-ray scattering

SHAPE	selective 2’-hydroxyl acylation analyzed by primer extension
SHAPE-MaP	SHAPE and mutational profiling
SLIC	sequence- and ligation-independent cloning
SRA	steroid receptor RNA activator
TR1-6	tandem repeats 1-6
v1,3,9	variants 1,3,9
wt	wild type
Xist	X-inactive specific transcript
1M6	1-methyl-6-nitroisatoic anhydride
1M7	1-methyl-7-nitroisatoic anhydride
3D	three-dimensional
% w/V	percent weight/volume

List of Figures

List of Tables

Abstract

Long non-coding RNAs (lncRNAs) are key players in vital cellular processes, including chromatin remodelling, DNA repair and translation. However, the size and complexity of lncRNAs present unprecedented challenges for mechanistic molecular studies, so that connecting structural information with biological function for lncRNAs has proven difficult so far.

Human maternally expressed gene 3 (MEG3) is an abundant, imprinted, alternatively-spliced lncRNA. During embryogenesis MEG3 controls *Polycomb* proteins, regulating cell differentiation, and in adult cells MEG3 controls p53, regulating the cellular response to environmental stresses. In cancerous cells, MEG3 is downregulated, but ectopic overexpression of MEG3 reduces uncontrolled proliferation, proving that MEG3 acts as a tumour suppressor. Evidence suggests that MEG3 functions may be regulated by the MEG3 structure. For instance, MEG3 is thought to bind p53 and *Polycomb* proteins directly. Moreover, different MEG3 splice variants, which comprise different exons and thus possess potentially different structures, display different functions. Finally, deletion mutagenesis based on a MEG3 structure predicted *in silico* identified a putatively-structured MEG3 motif involved in p53 activation. However, at the beginning of my work, the experimental structure of MEG3 was unknown.

To understand the MEG3 structure and function, I used chemical probing *in vitro* and *in vivo* to determine the secondary structure maps of two human MEG3 variants that differ in their p53 activation levels. Using functional assays in cells and mutagenesis, I systematically scanned the MEG3 structure and identified the p53-activating core in two domains (D2 and D3) that are structurally conserved across human variants and evolutionarily conserved across mammals. In D2-D3, the most important structural regions are helices H11 and H27, because in these regions I could tune p53 activation even by point mutations, a degree of precision never achieved for any other lncRNA to date. I surprisingly discovered that H11 and H27 are connected by “kissing loops”, and I confirmed the functional importance of these long-range tertiary structure interactions by compensatory mutagenesis. Going beyond state-of-the-art, I thus attempted to visualize the 3D structure of a 1595-nucleotide long MEG3 isoform by small angle X-ray scattering (SAXS), electron microscopy (EM), and atomic force microscopy (AFM). While SAXS and EM are limited by currently-insurmountable technical challenges, single particle imaging by

AFM allowed me to obtain the first low resolution 3D structure of MEG3 and reveal its compact, globular tertiary scaffold. Most remarkably, functionally-disrupting mutations that break the H11-H27 “kissing loops” disrupt such MEG3 scaffold, providing the first direct connection between 3D structure and biological function for an lncRNA.

Based on my discoveries, I can therefore propose a structure-based mechanism for p53 activation by human MEG3, with important implications in understanding carcinogenesis. More broadly, my work serves as proof-of-concept that lncRNA structure-function relationships can be dissected with high precision and opens the field to analogous studies aimed to gain mechanistic insights into many other medically-relevant lncRNAs.

Résumé en Français

Les ARNs long non codants (ARNlnc) jouent un rôle clé dans les processus cellulaires vitaux, notamment le remodelage de la chromatine, la réparation de l'ADN et la traduction. Cependant, la taille et la complexité des ARNlnc présentent des défis sans précédent pour les études moléculaires mécanistiques, de sorte qu'il s'est avéré difficile jusqu'à présent de relier l'information structurelle à la fonction biologique pour les ARNlnc.

Le gène 3 humain exprimé maternellement (de l'anglais "maternally expressed gene 3", MEG3), est un ARNlnc abondant, soumis à empreinte parentale et épissé alternativement. Pendant l'embryogenèse, MEG3 contrôle les protéines *Polycomb,* régulant la différenciation cellulaire, et dans les cellules adultes, MEG3 contrôle p53, régulant la réponse cellulaire aux stress environnementaux. Dans les cellules cancéreuses, MEG3 est régulé négativement, mais la surexpression ectopique de MEG3 réduit la prolifération incontrôlée, ce qui prouve que MEG3 agit comme un suppresseur de tumeur. Les données suggèrent que les fonctions de MEG3 pourraient être régulées par la structure de MEG3. Par exemple, on pense que MEG3 se lie directement aux protéines p53 et *Polycomb*. De plus, les différents variants d'épissage de MEG3, qui comprennent différents exons et possèdent ainsi des structures potentiellement différentes, présentent des fonctions différentes. Enfin, la mutagenèse par délétion, basée sur une structure de MEG3 prédit *in silico,* a permis d'identifier un motif MEG3 supposé structuré impliqué dans l'activation de p53. Cependant, au début de mes travaux, la structure expérimentale de MEG3 était inconnue.

Pour comprendre la structure et la fonction de MEG3, j'ai utilisé des sondes chimiques *in vitro* et *in vivo* pour déterminer la structure secondaire de deux variants humains de MEG3 qui diffèrent par leurs niveaux d'activation de p53. À l'aide d'essais fonctionnels dans les cellules et de mutagenèse, j'ai systématiquement analysé la structure de MEG3 et identifié le noyau activant p53 dans deux domaines (D2 et D3) qui sont conservés structuralement dans les variants humains et conservés dans l'évolution chez les mammifères. Dans D2-D3, les régions structurales les plus importantes sont les hélices H11 et H27, car dans ces régions, j'ai pu supprimer l'activation de p53 grâce à des mutations ponctuelles, un degré de précision jamais atteint pour les autres ARNlnc jusqu'ici. J'ai découvert de manière surprenante que H11 et H27

sont reliés par des boucles connectées l'une à l'autre (de l'anglais "kissing loops") et j'ai confirmé l'importance fonctionnelle de ces interactions de structure tertiaire à longue distance par mutagenèse compensatoire. Allant au-delà de l'état de l'art, j'ai donc essayé de visualiser la structure 3D d'une isoforme de MEG3 longue de 1595 nucléotides, par diffusion de rayons X à petit angle (SAXS), microscopie électronique (EM) et microscopie à force atomique (AFM). Alors que le SAXS et l'EM sont limités par des défis techniques actuellement insurmontables, l'imagerie par AFM m'a permis d'obtenir la première structure 3D à basse résolution de MEG3 et de révéler son échafaudage tertiaire compact et globulaire. Plus remarquable encore, les mêmes mutations qui perturbent la connexion entre les «boucles» H11-H27 et qui inhibent la fonction de MEG3, perturbent aussi la structure 3D de cet ARNlnc, fournissant ainsi le premier lien direct entre la structure 3D et la fonction biologique pour un ARNlnc.

Sur la base de mes découvertes, je peux donc proposer un mécanisme de l'activation de p53 basé sur la structure de MEG3, avec des implications importantes pour la compréhension de la cancérogenèse. Plus généralement, mes travaux prouvent que les relations structure-fonction des ARNlnc peuvent être disséquées avec une grande précision et ouvrent la voie à des études analogues visant à obtenir des informations mécanistes pour de nombreux autres ARNlnc d'importance médicale.

1. Introduction

Résumé en Français

Les ARN longs non codants (ARNlnc) sont des ARN de plus de 200 nucléotides de long et qui ne codent par pour des protéines (Cao, 2014). Les ARNlnc remplissent des fonctions biologiques importantes, telles que la régulation de l'expression des gènes et la prolifération cellulaire. Bien que nous connaissions l'importance biologique des ARNlnc, leurs mécanismes d'action sont très mal connus. Si nous avions plus d'informations structurales, il serait possible de comprendre leur mécanisme plus en détail. Cependant, la structure secondaire n'a été cartographiée que pour quelques ARNlnc et il n'existe pas de structure tertiaire d'ARNlnc. Le gène (MEG3), exprimé maternellement, est l'un des ARNlnc intéressants, pour lesquels la relation structure-fonction n'a pas été établie en raison d'une information structurale insuffisante.

Le MEG3 humain est un ARNlnc soumis à emprinte parentale, polyadénylé, alternativement épissé, et codé sur le chromosome 14q32 (Miyoshi et al., 2000; Zhang et al., 2010a). MEG3 se décline en au moins 27 variants. Le variant d'épissage 1 est le plus abondant et compte 1595 nt (537 kDa) (Zhang et al., 2010a). Il est composé de 7 exons. Parmi ces exons, les 3 premiers exons à l'extrémité 5'et les 3 derniers exons à l'extrémité 3' sont communs à au moins 15 autres variants connus de MEG3. La platforme Rfam (Nawrocki et al., 2015) révèle un niveau de conservation dans l'exon 3, commun chez les mammifères, qui dépasse le niveau de conservation typique des ARNlnc (Diederichs, 2014). Après épissage, MEG3 se localise dans le noyau (Cabili et al., 2015). Outre sa localisation cellulaire, son expression dans les tissus et les organes est également strictement régulée, MEG3 étant fortement exprimé dans le cerveau, le système endocrinien et le placenta (Zhang et al., 2003).

MEG3 joue un rôle fonctionnel à la fois dans les stades précoces de développement et dans les cellules adultes. En ce qui concerne la fonction de MEG3 dans le développement embryonnaire, MEG3 joue un rôle dans la régulation de l'expression génique, de la différenciation cellulaire et du développement, en particulier le neuro-développement (Gordon et al., 2010; Kaneko et al., 2014). De plus, dans les cellules adultes, MEG3 agit comme suppresseur de tumeur (Benetatos et al., 2011). L'allèle maternel MEG3 est hyperméthylé et l'expression de MEG3 est perdue dans de nombreuses lignées cellulaires cancéreuses et dans les tumeurs primitives (Zhou et al., 2012). Il est important de noter que les cellules humaines non fonctionnelles d'adénomes

hypophysaires transfectées avec MEG3 réduisent la croissance tumorale chez les souris (Chunharojrith et al., 2015) et que la surexpression ectopique de MEG3 dans les cellules cancéreuses réduit leur prolifération (Zhu et al., 2015). Ces résultats suggèrent que MEG3 agit directement comme suppresseur de tumeur. En effet, MEG3 est impliqué dans la régulation de nombreuses voies de signalisation qui déterminent le développement et la progression des tumeurs (Benetatos et al., 2011; Zhou et al., 2012). La principale voie régulée par MEG3 est la réponse au stress à médiation du facteur de transcription p53. MEG3 active p53 et les gènes cibles activés par p53, induisant ainsi une inhibition de la prolifération et un arrêt du cycle cellulaire (Zhou et al., 2007). Cependant, MEG3 n'active que certains gènes cibles de p53, ce qui indique sa spécificité (Zhu et al., 2015).

Bien que la pertinence physiologique et médicale de MEG3 soit connue, son mécanisme moléculaire reste encore largement inexpliqué, en partie parce que ses propriétés biochimiques et structurales sont méconnues. Bien que la structure secondaire expérimentale de MEG3 soit encore inconnue, les prédictions *in silico* basées sur la séquence suggèrent que MEG3 est hautement structuré (Zhang et al., 2010a). En outre, la perturbation de ces éléments structuraux prédits entraîne une diminution de la capacité d'activer p53 indépendamment de la séquence, ce qui implique que la structure de MEG3 est peut-être plus importante que sa séquence (Zhang et al., 2010a). Pour mieux comprendre le mécanisme moléculaire de MEG3, l'objectif de ce travail est de déterminer la structure expérimentale de MEG3 et d'établir les relations structure-fonction pour cet ARNlnc important.

1.1. Non-coding RNAs

RNA molecules are divided in protein-coding messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs). Although most of the human genome is transcribed, only a small part of it is translated. Among all the ncRNAs only a handful have assigned function and there is still debate in the field whether all these transcripts have a functional role or whether some are transcriptional 'junk' (Palazzo and Lee, 2015). The non-coding portion of the genome has become more prominent throughout evolution indicating its importance in more complex organisms such as humans [Ensembl release 88 (Yates et al., 2016), figure 1]. In humans, transcribed regions which produce noncoding RNAs (ncRNAs) are more abundant than protein coding regions. Non-coding RNAs fulfill various biological functions, such as maintenance of basic cellular processes and regulation of gene expression (Palazzo and Lee, 2015). Non-coding RNAs include housekeeping RNAs such as ribosomal, transfer, small nuclear and small nucleolar RNAs; short regulatory RNAs such as microRNAs, small interfering RNAs and piwi-associated RNAs; and long non coding RNAs (lncRNAs) (Cao, 2014). Small non-coding RNAs are highly conserved and their role in gene silencing is well studied (Brosnan and Voinnet, 2009). Unlike small non-coding RNAs, lncRNAs are poorly conserved and their molecular mechanisms still remain largely unexplained.

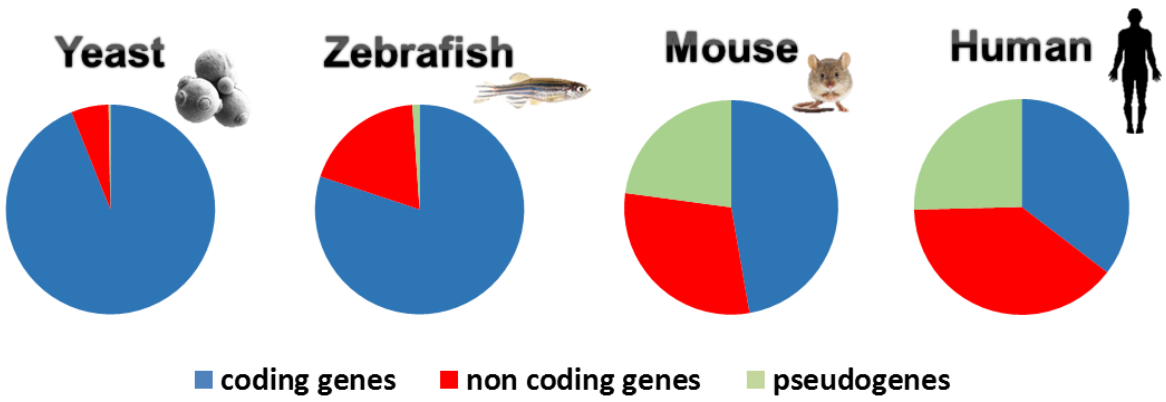


Figure 1: Non-coding genome through evolution.
In blue coding genes, in red non-coding genes and in green pseudogenes in different organisms from evolutionary simpler on the left to more complex on the right (*Saccharomyces cerevisiae, Danio rerio, Mus musculus, Homo sapiens* respectively). Ensembl release 88 (Yates et al., 2016).

1.1.1. Long non-coding RNAs

LncRNAs are more than 200-nucleotides long, non-protein-coding transcripts (Cao, 2014). Among the more than 32,000 human lncRNAs (Volders et al., 2013), a subgroup emerged as particularly suited for mechanistic studies based on their evolutionary conservation (Necsulea et al., 2014), specific cellular distribution (Cabili et al., 2015), specific tissue localization (Kaushik et al., 2013), and association to pathological phenotypes (Sauvageau et al., 2013; Wapinski and Chang, 2011). From the subset of lncRNAs that have been characterized we know that they perform important biological functions, such as genome structure modulation, epigenetic chromatin remodelling, DNA repair and translation (Mercer et al., 2009). Some of the best-studied lncRNAs are X-inactive specific transcript (Xist), HOX transcript antisense RNA (HOTAIR), steroid receptor RNA activator (SRA), and metastasis associated lung adenocarcinoma transcript 1 (MALAT1). LncRNAs fulfill many biological functions, but there is very little known about their mechanisms of action. If we had more structural information it would be possible to understand their mechanism in greater details.

1.1.1.1. Structural studies on lncRNAs

Since lncRNAs are very long *in silico* methods for predicting their structure are still rather unreliable. Moreover, phylogenetic analyses are not very informative since lncRNAs are not generally well conserved across distant species, at least according to our current level of understanding. Instead, for the study of lncRNA secondary structure, a range of enzymatic and chemical probing techniques are used (Blythe et al., 2016; Zampetaki et al., 2018). Enzymatic probing has been done using different enzymes (i.e. nuclease, RNase) that cut RNA molecules at specific sites (double stranded or single stranded regions) and subsequently analyzing the created fragments. Some of the commonly used enzymatic probing techniques are PARS (parallel analysis of RNA structure) and Frag-Seq (fragmentation sequencing). Chemical probing techniques rely on small molecules that are able to react with flexible RNA nucleotides (e.g. single stranded regions) and form a covalent bond that can be subsequently detected. Some of the used chemical probing techniques are dimethyl sulfate (DMS) probing, targeted Structure-Seq, selective 2'-hydroxyl acylation by primer extension (SHAPE), SHAPE and mutational profiling (SHAPE-MaP), in-cell SHAPE-Seq or *in vivo* click SHAPE sequencing (icSHAPE-seq). For

determining the solvent accessible regions of RNA, hydroxyl radical footprinting (HRF) is used. The 3D structure of RNA molecules has so far been studied by X-ray crystallography, SAXS, NMR and electron microscopy but mostly small fragments or motifs have been characterized by those methods. However, secondary structure has been mapped only for very few lncRNAs, namely SRA (Novikova et al., 2012), roX1 and roX2 (Ilik et al., 2013), HOTAIR (Somarowthu et al., 2015), COOLAIR (Hawkes et al., 2016), Braveheart (Xue et al., 2015), Xist (Smola et al., 2016), part of lincRNA p21 (Chillon and Pyle, 2016), RepA (Liu et al., 2017), SPRIGHTLY (Bongyong Lee et al., 2017) and NEAT1 (Lin et al., 2018). In addition, there is no tertiary structure of a full length lncRNA, only of small fragments. For instance, a fragment of ~70 nt from a ~8 kb lncRNA MALAT 1 (Brown et al., 2014) and a fragment of ~20 nt from a ~4 kb lncRNA Gas5 (Hudson et al., 2014) have been crystallized. In addition, the structure of a fragment of 32 nt from a ~0.5 kb lncRNA hTR (Leeper and Varani, 2005) and a fragment of 14 nt from a ~ 17 kb lncRNA Xist (Duszczyk et al., 2011) have been solved by NMR. These 3D structures are clearly insufficient to provide a comprehensive description of the molecular architecture of lncRNAs.

1.2. LncRNA MEG3

One of the interesting lncRNAs for which insufficient structure information is available is maternally expressed gene3 (MEG3). Human MEG3 is an imprinted, alternatively spliced, polyadenylated lncRNA encoded on chromosome 14q32 (Miyoshi et al., 2000; Zhang et al., 2010a). See table 1 for MEG3 nomenclature.

1.2.1. MEG3 evolutionary conservation

LncRNA MEG3 is conserved in at least 30 species of mammals, indicating a higher level of evolutionary conservation compared to other lncRNAs. In 6 of these species, including humans, orangutan, mouse, rat, cow and pig, the full MEG3 transcript sequence is annotated in public databases. However, in other mammals the presence of MEG3 can only be deduced based on genomic alignments. Rfam reveals clear patterns of conservation in exon 3 (Nawrocki et al., 2015). This level of conservation suggests that, unlike other lncRNAs (Diederichs, 2014), portions of the MEG3 sequence and possibly of its structure are functionally important.

1.2.2. Genomic organization of the human MEG3 locus

Human MEG3 is expressed from the Dlk1-MEG3 imprinted locus on chromosome 14q32, which also encodes six other ncRNAs (Anti-Peg11, Meg8, Irm/Rian, AK050713, AK053394, and Meg9/Mirg) and miRNAs, one of which (MIR770) overlaps with a MEG3 intron (Miyoshi et al., 2000). Imprinting of the Dlk1-MEG3 locus is regulated by acetylation of histones and by two differentially methylated regions (DMR), intergenic (IG) DMR and MEG3-DMR (Carr et al., 2007). IG-DMR, which is considered to be the primary imprinting control region, is situated upstream of the MEG3 gene. Instead, MEG3-DMR partially overlaps the MEG3 promoter (figure 2).

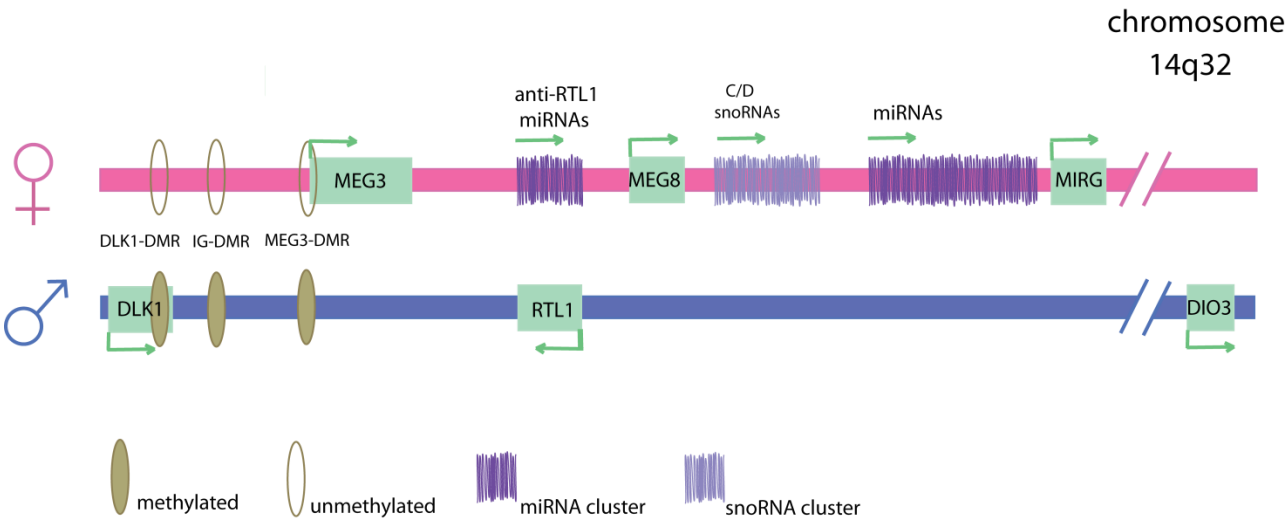


Figure 2: Schematic representation of the DLK1–MEG3 locus on human chromosome 14q32. Maternal chromosome is represented in pink, paternal in blue. Differentially methylated regions are shown as brown ellipses (filled - methylated; unfilled –unmethylated). Green boxes indicate expressed genes; purple lines indicate maternally expressed small RNA clusters. Green arrows indicate the direction of transcription.

Following transcription from the Dlk1-MEG3 locus, MEG3 is spliced (figure 3). The number of MEG3 splicing variants has not been precisely determined yet. 16 variants are annotated in NCBI Gene database, while 27 variants are annotated in Ensembl. MEG3 splicing variants are expressed at different level. Splicing variant 1 (MEG3v1, also described as MEG3) is the most abundant, making 40-86 % of total MEG3 expressed in different human tissues and cell types (Zhang et al., 2010a). Splicing variant 1 is 1595 nucleotides long and is composed of 7 exons. Of these exons, the first 3 exons at 5’ end (E1-2-3) and last 3 exons at the 3’ end (E10-11-12) are

common to at least 15 other known MEG3 variants, whereas central E5 is a variable exon. Splicing variant 9 (MEG3v9) is the second most abundant variant and corresponds to the shortest MEG3 isoform because it lacks varying exons. Splicing variant 3 (MEG3v3) starts ~24-nt downstream of MEG3v1 and contains varying exon E6 along with E1-3, E5, and E10-12. A similar pattern of exon organization and alternative splicing may be also conserved across mammals. For instance, Gtl2, the mouse homologue of human MEG3, is also alternatively spliced. As for human MEG3, Gtl2 splicing variants also possess common exons at their 5' and 3' ends and varying middle exons.

Table 1: MEG3 nomenclature.

	MEG3v1	**MEG3v9**	**MEG3v3**
Ensembl	MEG3-209 ENST00000451743.6	MEG3-217 ENST00000521404.5	MEG3-224 ENST00000554639.5
ncbi	Transcript variant 1 NR_002766.2	Transcript variant 9 NR_046466.1	Transcript variant 3 NR_003531.3
Klibanski lab (Zhang et al., 2010a)	MEG3 Miyoshi: AB032607, AB032607.1	MEG3e GQ183497.1	MEG3a AY314975
In this work	MEG3v1	MEG3v9	MEG3v3
Comments	Ensemble variant has 18 nucleotides less in exon 1 (AGCCCCTAGCGCAGACGG) and 5 more in the last exon (CTGAT) ncbi: same as our construct Miyoshi input has 11 nucleotides less in exon 1 (AGCCCCTAGCG) and longer poly A	Ensemble variant has 18 nucleotides less in exon 1 (AGCCCCTAGCGCAGACGG) and 37 less in the last exon ncbi :longer polyA in the end Klibanski input: 8 nucleotides less in exon 1 (AGCCCCTA), longer poly A	24 nucleotides less in exon 1 (AGCCCCTAGCGCAGACGGC GGAGA) in all sources ncbi and Klibanski input: longer poly A Position 908 (exon 3): with and without T

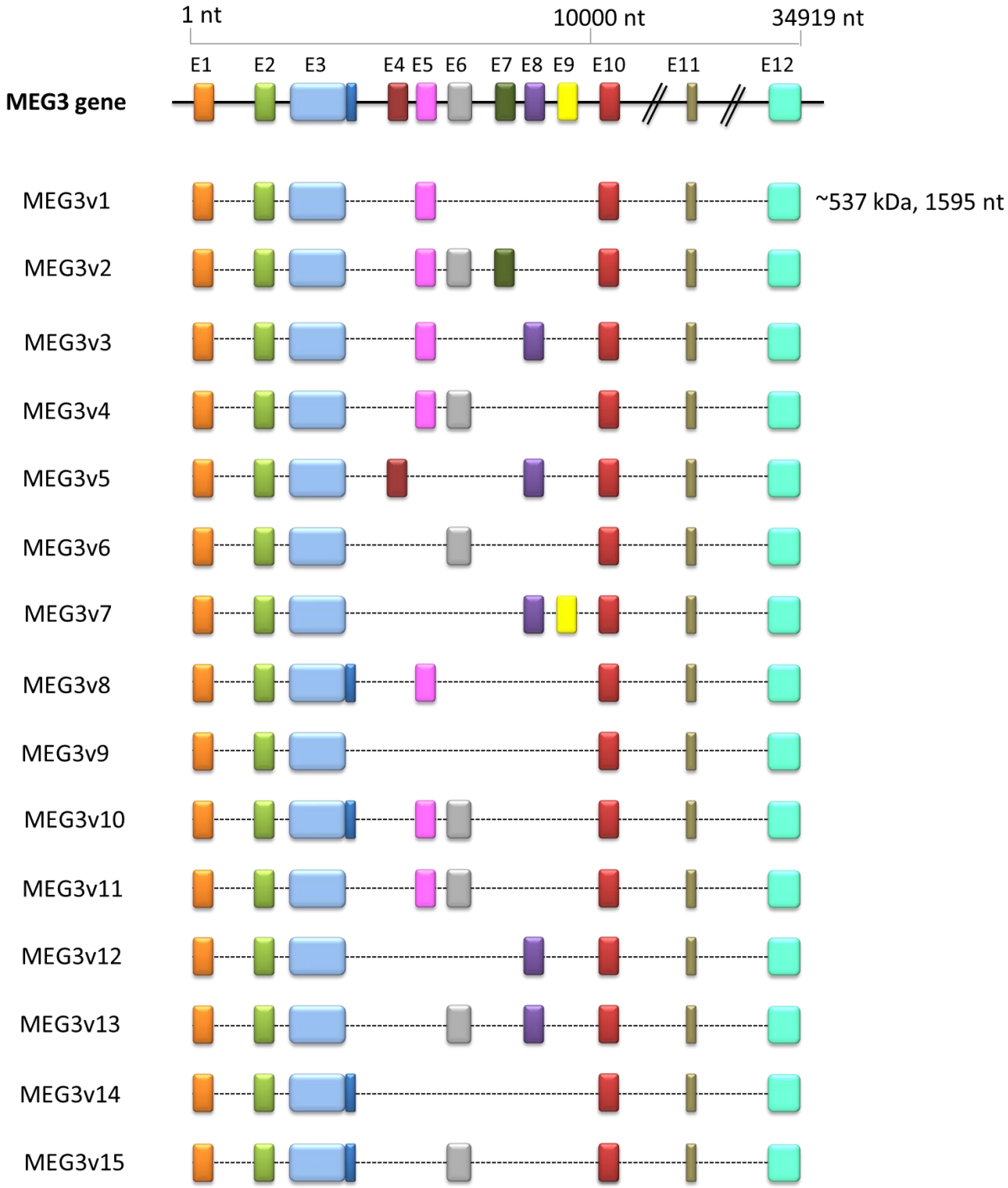


Figure 3: Human MEG3 gene and splicing variants.

Each exon is represented as a colored box. The size of mature variant 1 is indicated in kDa and number of nt on the right.

1.2.3. MEG3 expression in physiological and pathological conditions

Expression of MEG3 from the Dlk1-MEG3 locus is stimulated by c-AMP through a c-AMP responsive element located in a region proximal to the MEG3 promoter. Some micro-RNAs (miRNA-29a, miRNA148a) also regulate MEG3 expression by inhibiting DNMT (DNA Methyl transferase) 1 and 3b and consequently reducing methylation levels at the MEG3 promoter and DMR (Braconi et al., 2011; Yan et al., 2014; Zhao et al., 2006). As a consequence of those regulatory networks, MEG3 expression varies in different tissues, at different developmental stages, and in several pathological conditions. For instance, in adults under normal physiological conditions MEG3 is highly expressed in brain, pituitary gland, adrenal gland and placenta (Zhang et al., 2003). MEG3 is also highly expressed during GABA neuron neurogenesis (Mercer et al., 2008). Instead, in pathological conditions MEG3 expression changes (Benetatos et al., 2011; Zhou et al., 2012). For instance, MEG3 expression is down-regulated in the brain of patients affected by Huntington's disease and in the human islets from type 2 diabetes donors (Johnson, 2012; Kameswaran and Kaestner, 2014). Moreover, MEG3 expression is severely disregulated in various cancers. MEG3 expression is upregulated in adenomyoepitheliomatus lesions and salivary tumors (Benetatos et al., 2011) and down-regulated in tumor cell lines, such as H4 human brain neuroglioma and HepG2 human liver hepatocellular carcinoma, and in primary tumors, such as pituitary tumors (Zhou et al., 2012). Cancers with downregulated levels of MEG3 are very invasive. For instance, brain, breast and lung cancers have incidence rates of 157.6 (per 100,000) and mortality rate of 63.8 (per 100,000) in Europe (Ferlay et al., 2013).

1.2.4. MEG3 cellular localization

Mature spliced MEG3 transcripts are primarily localised in the nucleus of the cell (more than 90 %), although some (less than 10 %) MEG3 transcripts are exported to the cytoplasm, as determined by qRT– PCR (Mondal et al., 2015) and single cell RNA FISH (Cabili et al., 2015). Additional evidence that supports MEG3 dual localisation both in the nucleus and in the cytoplasm is that MEG3 interacts with both nuclear and cytoplasmic proteins, as discovered by MS2-tagged RNA affinity purification and mass spectrometry (Liu et al., 2015). Proteins identified by this technique play a role in different biological processes, including intracellular transport and translation but it is not known if their binding to MEG3 affects their function.

1.2.5. MEG3 functions

The precise cellular functions of MEG3 are not yet fully understood. MEG3 is functionally important at two stages of development (figure 4). In embryos, MEG3 plays a role in the regulation of gene expression, cell differentiation and development, particularly neurodevelopment (Gordon et al., 2010; Kaneko et al., 2014). Instead in adult cells, MEG3 acts as a tumor suppressor (Benetatos et al., 2011).

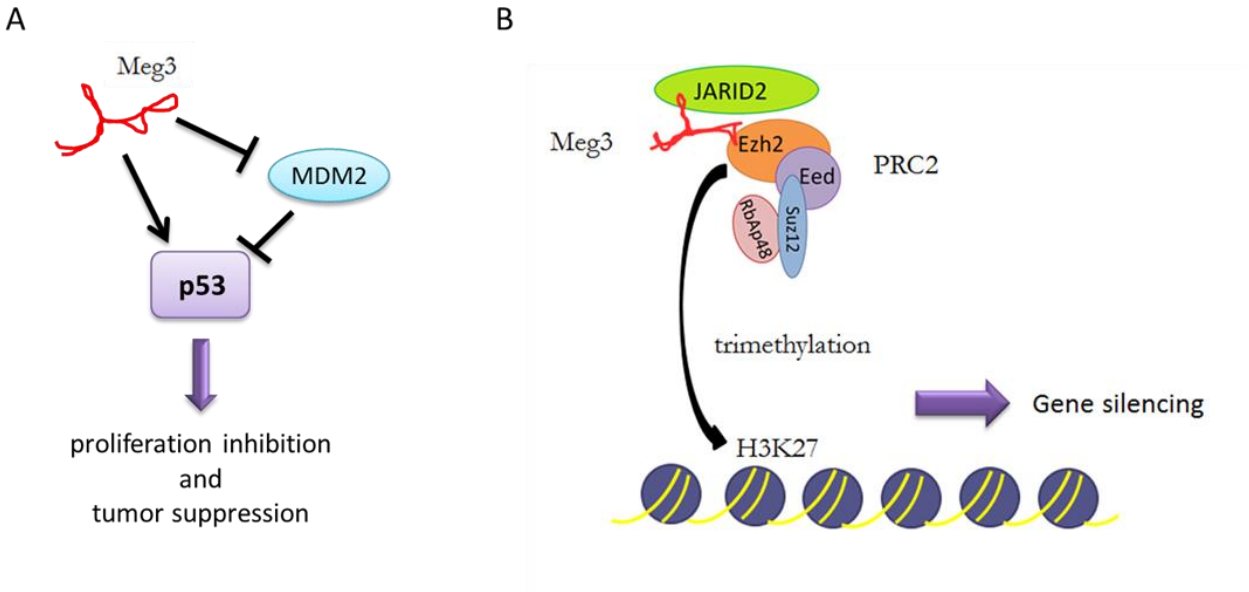


Figure 4: MEG3 functional pathways.
(A) Schematic representation of MEG3 as a tumor suppressor in the p53 pathway. MEG3 activates p53 directly or indirectly by suppressing MDM2 leading to activation of p53 downstream targets and subsequent proliferation inhibition and tumor suppression. (B) Schematic representation of MEG3 interactions with PRC2/JARID2 which leads to trimethylation of lysine 27 on histone 3 and subsequent gene silencing.

1.2.5.1. Development via *Polycomb* gene silencing

At the molecular level, MEG3 exerts its functions by interacting with specific proteins. Recent MS studies revealed that MEG3 interacts with at least 31 proteins (see paragraph 1.2.4 above). Among these interactions, the best characterized ones are those between MEG3 and *Polycomb* group proteins, via which human MEG3 promotes gene silencing (Kaneko et al., 2014). Among *Polycomb* group proteins, MEG3 specifically interacts with *Polycomb* repressive complex 2 (PRC2) and JARID2. PRC2 is a complex composed of 4 subunits: Suz12, either RbAp46 or RbAp48 (RbAp46/48), different isoforms of EED and either Ezh1 or Ezh2 (Ezh1/2). Ezh1 and Ezh2 are the

enzymatic subunits. They are methyltranferases that di- and tri-methylate lysine 27 of histone H3 (H3K27), thus leading to gene silencing. PRC2 has recently been structurally characterized by electron microscopy coupled to chemical cross-linking and mass spectrometry (Ciferri et al., 2012), by X-ray crystallography (Jiao and Liu, 2015) and by cryo-EM (Poepsel et al., 2018). Despite providing insights into the molecular organization of PRC2, such structures do not clarify the key functional properties of the complex. For instance, how does human PRC2 target specific sites of chromatin, considering that it does not possess any DNA binding domain? Additionally, how does PRC2 bind lncRNAs like MEG3, considering that it does not possess any canonical RNA binding motif?

Biochemical and cell-based studies provide only partial answers to those questions. Concerning DNA binding, it is known that chromatin targeting by PRC2 employs accessory proteins. For instance, JARID2, which is a Jumonji family protein and possesses DNA recognition domains, is important for recruitment of PRC2 to specific genes. However, JARID2 does not bind DNA specifically enough to fully explain PRC2 targeting. Thus, other factors must be involved and lncRNAs that directly bind PRC2 and JARID2 are likely to play important roles. For instance, MEG3 may also recognize DNA specifically forming DNA-RNA triplex via a GA-rich region located in exon 1 (Mondal et al., 2015). However, how MEG3 interacts with JARID2 and PRC2 is much less clear, because neither PRC2 nor JARID2 possess canonical RNA binding motifs. As a consequence, PRC2 interactions with lncRNAs are highly debated (Davidovich et al., 2013) and the nucleotides and amino acids directly involved in the interaction are not known. Instead, for JARID2, site of RNA interaction has been mapped more precisely on the protein and correspond to an internal 30 amino-acid region (Kaneko et al., 2014). The RNA binding region (RBR) of JARID2 is not structurally characterized and does not have any sequence homology with other known RNA binding proteins.

1.2.5.2. Tumor suppression

For a molecule to be classified as a tumor suppressor it must be proven that is not functionally active or it is not expressed any more in tumors, that both *in vitro* and *in vivo* re-expression in tumors leads to reduced proliferation and that knockout models in animals lead to tumor development or defects (Hakem and Mak, 2001; Weinberg, 1991; Zhou et al., 2012). In many

cancer cell lines and in primary tumors, such as non-functioning pituitary adenomas (NFA), MEG3 maternal allele is hypermethylated and MEG3 expression is lost (Zhou et al., 2012). Transfecting human NFA cells with MEG3 reduces tumor growth in nude mice (Chunharojrith et al., 2015) and ectopic MEG3 overexpression in cancer cells reduces their proliferation (Zhu et al., 2015). Two MEG3 knock-out mouse models have been created. The first model had a 10kb deletion including MEG3-DMR and the first 5 exons of gtl2 (Takahashi et al., 2009). The second model had a 5kb deletion spanning part of the MEG3 promoter and the first 5 exons (Gordon et al., 2010; Zhou et al., 2010). The two modes exhibit different phenotypes but both show developmental deficiency and/or perinatal death. In conclusion, MEG3 acts as a tumor suppressor because it fulfils the requirements to be classified as such and is involved in regulation of many signaling pathways that determine tumor development and progression (Benetatos et al., 2011; Zhou et al., 2012).

The primary pathway regulated by MEG3 is the p53 cell response to stress (see next paragraph 1.2.5.2.1.) but it also functions in p53-independent manners (figure 5). One putative signaling pathway by which MEG3 inhibits proliferation in the absence of p53 is retinoblastoma (Rb) signaling, because in human pituitary tumors frequent loss of positive regulator of Rb, the $p16^{INK4a}$ is observed coupled to loss of MEG3 expression and $p16^{INK4a}$ and MEG3 expression levels can be correlated (Zhang et al., 2010b). Another proposed mechanism by which MEG3 inhibits cell migration and invasion is by targeting the Rac1 gene. Rac1 gene expression levels decrease when MEG3 is downregulated in papillary thyroid carcinoma (PTC) and downregulated MEG3 in PTC is associated with lymph-node metastasis, while over-expressed MEG3 lowers cell migration and invasion in thyroid cancer (Wang et al., 2015).

Besides regulating cell proliferation, MEG3 also regulates how tumor cells respond to drugs. One pathway by which MEG3 may regulate drug response is the WNT/β-catenin signaling pathway. Deregulation of β-catenin and survivin (a target gene of the WNT/β-catenin pathway) by lower level of MEG3 in NSCLC (non-small cell lung cancer) is associated with higher cisplatin resistance (Xia et al., 2015). MEG3 is also associated with inhibition of angiogenesis in some tumors and in osteoarthritis (Gordon et al., 2010; Su et al., 2015).

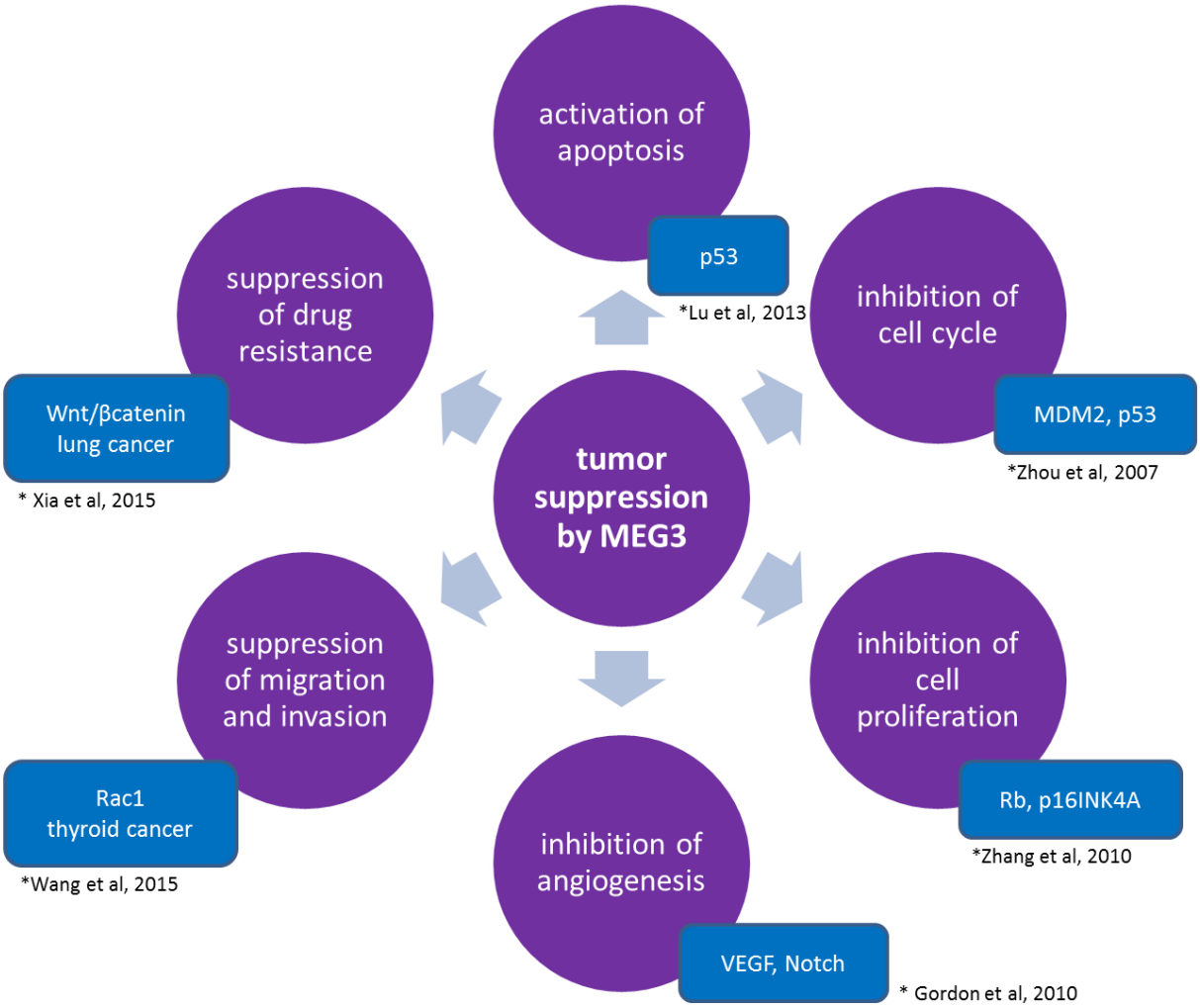


Figure 5: MEG3 tumor suppression.
Chart summarizing the role of MEG3 as a tumor suppressor and different pathways that are affected.

1.2.5.2.1. Tumor suppression via p53 pathway

A number of lncRNAs are involved in the p53 pathway, either by regulating p53 or by being regulated by p53 (Chaudhary and Lal, 2016). For example some of the p53 regulated lncRNAs are lincRNA-p21, PANDA, PINT, TUG1 and NORAD etc., and some of the lncRNAs that regulate p53 are MALAT1, H19, lincRNA-ROR and MT1JP etc. Of those lncRNAs, one of the best studied is MEG3. MEG3 activates p53 and p53 target genes thereby inducing inhibition of proliferation, cell cycle arrest and activation of apoptosis (Zhou et al., 2007). This action of MEG3 leads to selective activation of specific p53 target genes, but the determinants of the observed selectivity are as yet unknown (Zhou et al., 2007; Zhu et al., 2015). MEG3-dependent p53

activation can be tested by a luciferase assay (Zhang et al., 2010a). Using such assay, it was established that the different MEG3 splice variants have different abilities to activate p53 (Zhang et al., 2010a). Moreover, it was shown that in cancer the splicing pattern of MEG3 changes, leading to changes in the p53 stress response (Zhang et al., 2010a). The 27 known splice variants of MEG3 all possess the same stability and half-life in the cell. Yet, they are expressed at different levels and display profoundly different p53 activation abilities (Zhang et al., 2010a). For instance, MEG3v1 possesses intermediate p53 activation ability, MEG3v9 activates p53 more strongly than MEG3v1 and MEG3v3 acts as the weakest p53-activator (Zhang et al., 2010a). These observations led to the hypothesis that structural changes in the MEG3 splice variants induced by the varying middle exons affect the fine tuning of the p53 response by inducing different levels of expression of selected p53 target genes (Zhang et al., 2010a). The precise molecular mechanism by which MEG3 activates p53 is unknown. MEG3 may destabilize p53-ubiquitin ligase MDM2 and *in vitro* and *in vivo* studies suggest that MEG3 stabilizes p53 protein (Zhou et al., 2007; Zhu et al., 2015). Possibly, MEG3 binds p53 directly at its DNA binding domain (Zhu et al., 2015). Moreover, biochemical studies also revealed that disruption of *in-silico*-predicted structural motifs of MEG3 abolishes p53 activation (Zhou et al., 2007). Therefore, it is possible that the structure of MEG3 is important for its function. However, to date there is no systematic study connecting the experimental structure of MEG3 to its functional effects on p53.

1.3. Scientific aims

Sequence analysis and cell biology studies suggest that human MEG3 is highly structured thus raising numerous intriguing questions about how the MEG3 structure may regulate MEG3 functions.

First, MEG3 is one of the lncRNAs with the highest GC content (57 %). Does this nucleotide composition suggest a tendency to form folded, thermodynamically stable base-paired helices?

Second, MEG3 is highly conserved in at least 30 mammals, suggesting that the MEG3 sequence is important for function. Is the MEG3 structure also conserved?

Third, secondary structure predictions of MEG3 performed with the software mfold solely based on its sequence indicate that MEG3 possesses a very intricate molecular architecture, divided into three putative motifs (M1-M3) (Zhou et al., 2007). M1 would include the end of exon 1 and the beginning of exon 2, M2 would encompass part of exon 3, and M3 would encompass the last three exons at the 3' end. Interestingly, since they would encompass the 5'-terminal 3 exons and the 3'-terminal 3 exons, the putative motifs M1, M2 and M3 would be present in at least 15 human MEG3 splice variants. Are these putative M1-M3 motifs really formed by MEG3 *in vitro* and *in vivo*?

Fourth, disruption of putative structural elements of M2 almost completely abolishes the effects of MEG3 on the p53 cell response in a sequence independent manner (Zhang et al., 2010a). Does the MEG3 structure regulate MEG3 function as a tumor suppressor?

Fifth, MEG3 possesses many splicing variants [at least 27 annotated in Ensembl (Yates et al., 2016)] with different expression levels. All variants have the same stability in transfected cells, suggesting that their different expression levels may be due to different splicing efficiency. Possibly, structures formed cotranscriptionally may affect splicing efficiency and thus favor certain splicing variants. Considering that MEG3 splicing variants exhibit different levels of p53 transactivation, is it possible that there is a correlation between exon composition, formation of structured motifs, and MEG3 cellular functions?

To answer all these questions, the aim of my work is to characterize the secondary and tertiary structural architecture of MEG3 and its splice variants *in vitro* and *in vivo* and to establish structure-function relationships for this important lncRNA.

2. Materials and Methods

Résumé en Français

Ce chapitre décrit le matériel et les méthodes utilisés pour obtenir les résultats de mon travail de thèse. En résumé, les ARNs ont été exprimés et purifiés dans des conditions non dénaturantes, comme décrit précédemment (Chillon et al., 2015) avec des modifications mineures. Pour vérifier l'intégrité et l'homogénéité de ma préparation d'ARN, j'ai utilisé plusieurs méthodes biochimiques. Une électrophorèse sur gel natif a été réalisée pour vérifier l'intégrité de l'ARN pendant la production et la purification. La diffusion dynamique de la lumière (DLS) (Patel et al., 2016) a été réalisée pour évaluer la polydispersité des préparations de MEG3. Une chromatographie par exclusion de taille couplée à la diffusion de lumière laser à angles multiples (SEC-MALLS) (Patel et al., 2016) a été réalisée pour déterminer la polydispersité de MEG3 et l'état oligomère.

Pour caractériser MEG3 structuralement, j'ai utilisé les techniques suivantes. L'acylation sélective en 2'-hydroxyle analysée par extension d'amorce (SHAPE) (Wilkinson et al., 2006) a été réalisée pour déterminer les structures secondaires expérimentales. Une sonde chimique *in vivo* a été réalisée pour évaluer l'influence de l'environnement cellulaire sur la structure. Des alignements de séquence et structuraux ont été réalisés pour évaluer le potentiel de conservation évolutive de la structure secondaire obtenue. Des expériences de vitesse de sédimentation par ultracentrifugation analytique (AUC) (Chillon et al., 2015) ont été réalisées pour surveiller le repliement de MEG3. La chromatographie par exclusion de taille couplée à la diffusion de rayons X aux petits angles (SEC-SAXS) (Chen and Pollack, 2016) a été utilisée pour obtenir des informations initiales à faible résolution sur la structure 3D de MEG3. La microscopie à force atomique (AFM) et la microscopie électronique ont été utilisées pour visualiser MEG3 au niveau de molécule unique.

Pour évaluer les corrélations entre les domaines structuraux et le rôle fonctionnel de MEG3 en tant que suppresseur de tumeur, deux types d'analyses fonctionnelles ont été effectuées dans des lignées cellulaires humaines: des analyses du cycle cellulaire utilisant la cytométrie en flux (Lu et al., 2013) et d’une transactivation d'un gène de la luciférase (Zhang et al., 2010a). Une PCR quantitative en temps réel (qRT-PCR) a été réalisée comme expérience de contrôle pour

évaluer l'efficacité de la transfection et pour tester le niveau endogène d'expression de MEG3 dans différentes lignées cellulaires de mammifères.

Toutes les méthodes mentionnées sont expliquées en détail dans le chapitre suivant.

2.1. Cloning and mutagenesis

A plasmid containing the sequence of human MEG3 transcript variant 1 was obtained by gene synthesis [GeneArt (Life Technologies)]. From this synthetic vector, the sequence of MEG3 was amplified by PCR and inserted by sequence- and ligation-independent cloning (SLIC) (Li and Elledge, 2012) into the scaffold of a modified pBluescript vector immediately downstream of a T7 promoter sequence and immediately upstream of an XbaI restriction site. The resulting vector was named pTU1. All pBluescript based vectors were used for *in vitro* transcription. Plasmid pTU2 containing a MEG3 variant 9 without varying exons was created to test influence of varying exons on MEG3 structure. Plasmid pTU2 was created by deleting varying exon 5 from MEG3 transcript variant 1 (Δ936-1049) in plasmid pTU1 with quick change mutagenesis. Plasmids pTU3-pTU7 contain 5 different domains of MEG3, which were determined from secondary structure map, respectively domain1 (2-196), domain 2 (230-410), domain 3 (471-902), domain 4 (951-1113) and domain5 (1116-1486). All domains were amplified by PCR from pTU1 and inserted by SLIC into the scaffold of pTU1 (between T7 promoter sequence and *Xba*I restriction site). Plasmid pTU123 was created by mutating the loop on top of the helix 11 to poly A (365-8xA-374) in MEG3 transcript variant 1 in plasmid pTU1 with quick change mutagenesis.

pCMS-d2-MEG3 was a kind gift of Yunli Zhou (Massachusetts General Hospital, Boston , USA) (Zhou et al., 2007). All pCMS-d2-MEG3 based vectors were used for *in vivo* assays by flow cytometry. Plasmid pTU8 was created by amplifying MEG3 and adding *Sac*I restriction site at 5' and *Not*I restriction site on 3' by PCR from pTU1 and inserting it in pCMS-d2-MEG3 with quick ligation between *Sac*I and *Not*I. Two complementary oligonucleotides SNf and SNr containing a *Sac*I restriction site, 13 nt sequence (5'-GGTTCACTAAACG-3') and *Not*I restriction site were ordered from Eurofins (5'- CCGTTTAGTGAACCGC-3', 5'- GGCCGCGGTTCACTAAACGGAGCT-3'). Plasmid pTU9 was created by annealing SNf and SNr by incubating 2 min at 95 °C and letting it cool down to RT gently and inserting it in pCMS-d2-ME with quick ligation between *Sac*I and *Not*I. MEG3 variants 1 and 9 were cloned in pcDNA3 vector between *Kpn*I and *Not*I restriction sites. Different structural mutants were cloned in pCMS-d2-MEG3 (for flow cytometry assay) and pcDNA3 (for luciferase assay). The presence of the target gene in all plasmids was

confirmed by enzyme digestion or colony PCR and agarose gel electrophoresis. The sequence of all constructs was validated by DNA sequencing (Eurofins). Mach1 competent cells (derivatives of *E. coli*) were transformed with plasmids. Plasmids were extracted with mini and maxi preps (Qiagen) from a single colony. For a list of all cloned vectors see appendix table 5.

2.2. RNA production and purification

MEG3 was expressed and purified under non-denaturing conditions as previously described (Chillon et al., 2015) with minor modifications (figure 6). Briefly, plasmids pTU1-pTU7 and pTU123 were linearized overnight with restriction enzyme XbaI (NEB). The linearized vectors were transcribed *in vitro* with T7 polymerase in various buffers (table 2). Highest transcription yield was obtained in buffer 3 (100 mM $MgCl_2$, 400 mM TrisHCl pH 8.0, 20 mM spermidine, 100 mM DTT) and 4 (80 mM $MgCl_2$, 400 mM TrisHCl pH 8.0, 50 mM DTT, 1 M KCl). Following transcription, template DNA and proteins were removed with Turbo DNase (Thermo Scientific) and proteinase K (Promega) treatment, respectively. Divalent ions were chelated with EDTA in the presence of physiological concentrations of monovalent ions. Such treatment may disrupt the tertiary structure, but preserves the co-transcriptionally generated secondary structure of MEG3 and allows for precise subsequent titration of magnesium concentrations in folding experiments. Following EDTA treatment, samples were rebuffered with filtration buffer (KCl 0.1 M, K-MOPS pH 6.5 8 mM, Na-EDTA 0.1 mM) using Amicon Ultra-0.5 centrifugal concentrators (molecular weight cut-off of 100 kDa) to wash out undesired components of previous reactions. MEG3 samples purified in this way were finally subjected to a polishing size-exclusion chromatography (SEC) step using Tricorn columns (GE Healthcare) self-packed with Sephacryl S500 resin and run in filtration buffer.

Table 2: T7 transcription buffers. Composition of 10x buffers tested for *in vitro* transcription.

Component	Buffer 1	Buffer 2	Buffer 3	Buffer 4
$MgCl_2$ 1 M	220 mM	120 mM	100 mM	80 mM
TrisHCl pH 8.0 1 M	400 mM	400 mM	400 mM	400 mM
spermidine 2 M	20 mM	20 mM	20 mM	-
DTT 1 M	100 mM	-	100 mM	50 mM
NaCl 5M	-	100 mM	-	-
KCl 4M	-	-	-	1 M

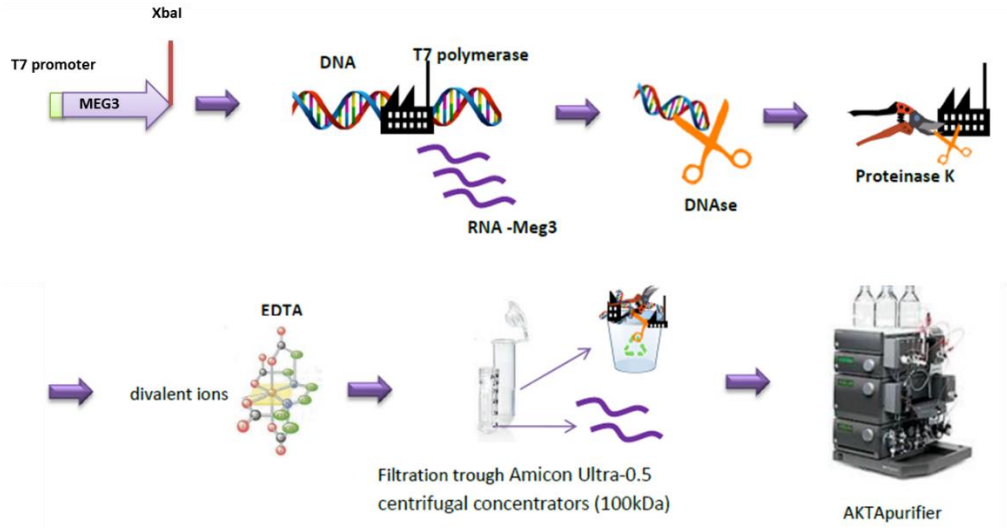


Figure 6: MEG3 production and purification.
Schematic representation of MEG3 production and purification by *in vitro* transcription from a linearized vector in non-denaturing conditions using centrifugal concentrators and SEC.

2.3. Native gel electrophoresis

Native gel electrophoresis was performed to check the integrity of RNA during production and purification. 1 % agarose gels were run in 1x Tris-Borate (TB) buffer (89 mM Tris base, 89 mM boric acid) supplemented with the indicated concentrations of Mg^{2+}. Gels in TB with no Mg^{2+} and with 2 mM Mg^{2+} were run for 45 min at 110 V, and gels in TB with 5 and 10 mM Mg^{2+} were run for 120 min at 80 V. Samples were mixed in a 5:1 ratio with 6x RNA native gel dye (0.5x TB buffer, 40 % sucrose, 0.5 % w/V orange G) before gel loading. Gels were stained with 1x SYBR Safe (Invitrogen) gel stain in 1x TB buffer for 1 h at room temperature before exposure (Invitrogen). 1 % agarose gels were run in 1x TAE buffer for 30 min at 100 V. Quick load purple 2-log DNA ladder (NEB) was used as size reference markers.

2.4. Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) sedimentation velocity experiments were performed to monitor MEG3 folding (Chillon et al., 2015). AUC experiments were performed at the IBS platform that is part of partnership for structural biology. Purified MEG3 was supplemented with varying concentrations of $MgCl_2$ ranging from 0.01 mM to 100 mM or with varying concentrations of KCl ranging from 200 mM to 900 mM. Samples were analysed using Beckman XL-A/XL-I centrifuge with AN-50 Ti rotor (Beckman Coulter). All experiments were performed at

20°C at 25,000 rpm overnight. Data were analyzed with Sedfit using continuous c(s) distribution model (Schuck, 2000).

2.5. Dynamic light scattering

Dynamic light scattering (DLS) was performed to assess polydispersity of MEG3 preparations (Patel et al., 2016). Purified MEG3 samples were analyzed over a concentration range from 0.5 µM to 5.5 µM. Samples were analyzed using Zetasizer Nano S (Malvern).

2.6. Size-exclusion chromatography coupled to multi-angle laser light scattering

Size-exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) was conducted to determine MEG3 polydispersity and oligomeric state (Patel et al., 2016). MEG3 was purified as described (paragraph 2.2.) up to EDTA treatment and rebuffered in filtration buffer. Different aliquots of pure MEG3 were diluted to concentrations of 5-0.32 µM and injected on SEC-MALLS either directly (no magnesium) or after supplementation with 5 and 17.5 mM $MgCl_2$. The same self-packed columns as previously described (paragraph 2.2.) were used.

2.7. Size-exclusion chromatography coupled to small angle x-ray scattering

Size exclusion chromatography coupled to small angle X-ray scattering (SEC-SAXS) was used to derive initial low-resolution information on the 3D structure of MEG3 (Chen and Pollack, 2016). MEG3 was purified as described (paragraph 2.2.) up to EDTA treatment, rebuffered in filtration buffer and filtered using centrifugal filter units with 0.22 µm pore size (Merck Millipore). Different aliquots of pure MEG3 were diluted to concentrations of 5-0.32 µM and injected on SEC-SAXS directly in the absence of magnesium which corresponds to partially-folded MEG3 sample. The same self-packed columns as previously described (paragraph 2.2.) were used. SAXS data were collected at the BioSAXS beamline BM29 at ESRF, Grenoble and analyzed in ISpyB (Delageniere et al., 2011) and using ATSAS modules PRIMUS and DAMMIF (Petoukhov et al., 2012).

2.8. *In vitro* secondary structure probing (*in vitro* SHAPE)

Secondary structure probing of MEG3 variant 3 was performed by Eleni Anastasakou (predoctoral fellow in the Marcia group).

Selective 2'-Hydroxyl acylation Analyzed by Primer Extension (SHAPE) (Wilkinson et al., 2006) was performed to determine the experimental secondary structure of MEG3 (figure 7B). The fraction of MEG3 with highest concentration that eluted from SEC after purification was used for the analysis. To be sure that all RNA was folded to its most compact form, the RNA was supplemented with 17.5 mM $MgCl_2$. MEG3 supplemented with 17.5 mM $MgCl_2$ and native without addition of cations was subjected to chemical modification using 1-methyl-7-nitroisatoic anhydride (1M7), N-methylisatoic anhydride (NMIA), 1-methyl-6-nitroisatoic anhydride (1M6) and dimethyl sulfate (DMS). 1M7, 1M6 and NMIA react with the 2'-hydroxyl group on the RNA nucleotides (figure 7A). Reaction of 1M7 with RNA and formation of adducts is favorable if nucleotides are unpaired and thus more flexible, whereas nucleotides constrained by base-paring form adducts less favourably. Reaction of 1M6 with RNA is favorable if nucleotides are stacking. Reaction of NMIA with RNA is favorable if nucleotides experience slow dynamics. Comparing 1M6 and NMIA reactivity it is possible to detect nucleotides that are involved in noncanonical and tertiary interactions in RNA. DMS methylate N1 of adenine and N3 of cytosine that are not base paired. Modifications were then mapped onto the MEG3 sequence by reverse transcription and fragment analysis. Eight primers positioned every 200 bp of MEG3, were designed and coupled with fluorescent dyes 5-FAM and JOE (primer sequences in appendix table 6). The primer extension reaction was performed using the Omniscript reverse transcriptase (Qiagen) on MEG3 samples treated with 1M7, 1M6, NMIA, DMS or DMSO and EtOH (non-adduct forming control). Reverse transcription stops whenever it encounters a nucleotide that is methylated or modified with 1M7, 1M6, NMIA. Samples were then submitted for fragment length analysis with capillary electrophoresis (Eurofins). QuShape (Karabiber et al., 2013) was used to determine the chemical probing reactivity profiles. Formation of adducts was quantified by comparison between the 1M7-, 1M6-, NMIA- and the DMSO-treated samples and the rate of methylation was quantified by comparison between the DMS- and EtOH-treated samples. Each experiment was done in triplicates. Average values of DMS reactivity values from 3 experiments were self-normalized as described (Chillon et al., 2015). Average values of 1M7 reactivity values from 3 experiments were normalized with "simple2boxplot.py" python script (Rice et al., 2014a) and average values of 1M6 and NMIA reactivity values from 3 experiments were normalized with

"boxplot2simple.py" python script (Rice et al., 2014a). Normalized 1M6 reactivity values were subtracted from the NMIA reactivity values with "differenceByWindow.py" python script (Rice et al., 2014a). Normalized 1M7 reactivity values were classified in 3 groups as follow: 0-0.4 not reactive (most likely base-paired), 0.4-0.85 moderately reactive and >0.85 very reactive (most likely single stranded). The software RNAStructure (Mathews, 2004; Reuter and Mathews, 2010) and the "SuperFold.py" python script (Siegfried et al., 2014) were used to obtain the secondary structure map of MEG3 based on the normalized 1M7 SHAPE reactivity values and NMIA and 1M6 differential SHAPE reactivity values as constrains. Java applet VARNA (Darty et al., 2009) was used to visualize secondary structure using coordinates obtained with RNAStructure software and "SuperFold.py" python script. DMS reactivity values were used as a control to validate the secondary structure map obtained with differential SHAPE with 3 reagents.

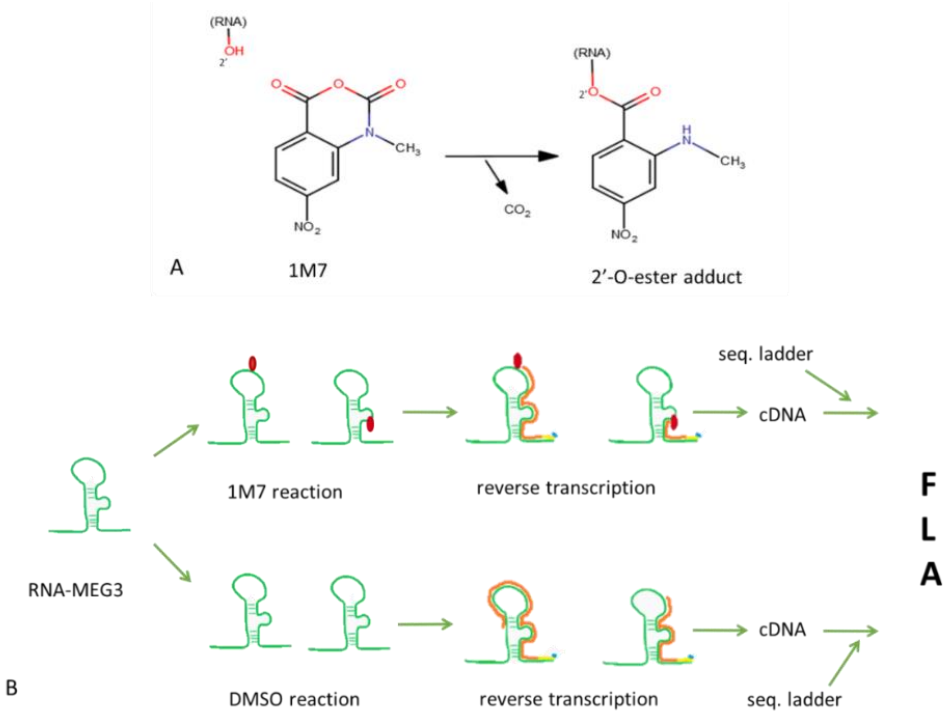


Figure 7: Secondary structure probing by SHAPE.
(A) Structure and reaction mechanism of reagent 1M7. (B) SHAPE pipeline. MEG3 shortly reacts with 1M7 to form adducts and with DMSO as control (no adduct formation), after the adduct forming reaction both MEG3 samples are reverse transcribed with labeled primers and finally samples are send together with sequencing ladders to fragment length analysis (FLA) with capillary electrophoresis.

2.9. Atomic force microscopy

Atomic force microscopy (AFM) experiments were performed in collaboration with Jean-Luc Pellequer and Jean-Marie Teulon (IBS). Power spectral density analysis was performed by Paolo Annibale (MDC, Berlin).

AFM experiments were performed to visualize MEG3 on a single molecule level. MEG3 and group II intron (Marcia and Pyle, 2012) were purified as described and the fraction of MEG3 and group II intron with highest concentration that eluted from SEC after purification were diluted 1:100 times in filtration buffer (KCl 0.1 M, K-MOPS pH 6.5 8 mM, Na-EDTA 0.1 mM), filtration buffer with magnesium (KCl 0.1 M, K-MOPS pH 6.5 8 mM, Na-EDTA 0.1 mM, 10mM $MgCl_2$) or "Milli-Q" water. Poly (A) RNA (GE Healthcare) used as negative control was dissolved in same buffers and "Milli-Q" water at a concentration of 0.3 µg/ml. To obtain denatured samples, MEG3, group II intron and poly (A) were precipitated with isopropanol overnight at -20°C and resuspended in deionized formamide and diluted with ethanol to reach same final concentration as samples diluted in buffer/"Milli-Q" water. A 1 µl, 2.5 µl or 5 µl drop of MEG3, group II intron and poly (A) samples in buffer was deposited on mica, incubated for 3 min, washed with 2 ml of "Milli-Q" water with 200µl drop steps to remove excessive salt crystals, and finally dried with nitrogen gas. A 1 µl, 2.5 µl or 5 µl drop of denatured samples and MEG3, group II intron and poly (A) samples in "Milli-Q" water was deposited on mica, incubated 3 min and dried with nitrogen gas. Imaging was performed on a Multimode 8, Nanoscope V (Bruker) equipped with NanoScope software (Bruker, Santa Barbara, CA). Imaging was done with peak force imaging mode at ~ 1Hz rate, with 512 and 1024 pixel resolution and other parameters were adjusted automatically with ScanAsyst in Air mode. Cantilever ScanAsyst-air (Bruker) with 2nm tip radius, 70 kHz frequency and 0.4 N/m spring constant was used. Images were processed with Gwyddion (Nečas and Klapetek, 2012), and if needed stripe noise was removed using DeStripe (Chen and Pellequer, 2011). Power Spectral Densities (PSD) of the AFM topographic signal were collected in square regions of 250 nm side around each particle of interest (Higuchi, 1988). The power spectral density was collected along the fast scanning axis of the microscope to avoid potential artefacts due to line-to-line offset. The PSD of H11LpA was collected along the y axis, due to a slight resonance of the tip along the fast scanning axis (x).

PSDs of v1 along x and y are nearly identical. PSDs were calculated using the SPM data analysis software Gwyddion (Nečas and Klapetek, 2012), and the PSD for all the particles observed under each given experimental condition were averaged using the software Igor Pro (WaveMetrics, USA). The resulting averaged PSD were plotted against the spatial (angular) frequency and the associated spatial length-scale (Calo et al., 2009). Linear fits to the so-called auto-affine region, i.e. where the PSD frequency dependence is of the form $PSD(f)=a_0 f^{-\gamma}$, are displayed as dashed lines. Fits were performed, within an arbitrarily selected x-range, using a weighted least square algorithm, using the software Igor Pro.

2.10. Electron microscopy

Electron microscopy (EM) experiments were performed in collaboration with Manikandan Karuppasamy (research scientist in the Marcia group).

EM experiments were performed in an attempt to visualize MEG3 on a single molecule level at higher resolution.

2.10.1. Negative staining

Carbon grids were glow discharged just before use. Glow discharge conditions were 25 mA, 30 s discharge, 10 s hold time at 0.3 mbar with negative polarity (Pelco EasiGlow system). 5 µl of different concentrations (0.025, 0.05, 0.075, 0.1 and 0.5 µM) of MEG3 prepared under different Mg^{2+} conditions (5, 10 and 17.5 mM) were applied to glow discharged carbon grids. 5 µl of different concentrations (0.3 and 0.7 µM) of MEG3 D2-3 prepared under different Mg^{2+} conditions (10, 17.5 and 25 mM) were applied to glow discharged carbon grids. The sample was incubated for 1 min to achieve maximum absorption in the carbon and then blotted with filter paper to remove excess sample. Immediately after blotting, 5 µl of 1 % Uranyl formate solution was applied and the grid was blotted after 20 s of staining. The grids were air dried on the bench before they were used for observation in the microscope.

2.10.2. Data collection and 2D processing

Data collection was done using a T12 BioTwin microscope operated at 120 kV with low dose settings. The C2 and Objective apertures were at position 3. The dose used was ~ 5-6 e/A2/s. Images were acquired at a nominal magnification of 49k x using Ceta CCD camera resulting a

pixel size of 2.06 Å. All data processing was done in Relion v2.1 except particle picking. The particles were picked either manually and/or semi-automatically with 'Swarm' option using the program e2boxer.py from EMAN2. The coordinates and micrographs were imported and the contrast transfer functions (CTF) were estimated from CtfFind4 as interfaced in Relion. 2D class averages were done on the extracted particles.

2.10.3. Cryo electron microscopy

1.5 – 3.0 µl of 0.3 µM and/or 0.67 µM of MEG3 in 25 mM Mg^{2+} buffer was applied to freshly both side glow discharged Quantifoil Au 300 mesh R1.2/1.3 grids. 1.5 – 3.0 µl of 0.5 µM and/or 5 µM of MEG3 D2-3 in 0, 5, 10 and 17.5 mM Mg^{2+} buffers was applied to freshly both side glow discharged Quantifoil Au 300 mesh R1.2/1.3 grids. The glow discharge condition was 25 mA, 12s discharge; 10 s hold time at 0.3 mbar with negative polarity (Pelco EasiGlow system). Grids were vitrified using a vitrobot for 2 s blot time with -5 blot force keeping 95 % humidity at 20°C. Cryo grids were mounted on a liquid nitrogen cooled Gatan 626 side-entry holder and observed under a T12 BioTwin microscope operated at 120 kV with low dose settings. The C2 and Objective apertures were at position 3. The dose used was ~ 5-6 e/A2/sec. Images were acquired at a nominal magnification of 49k x using Ceta CCD camera resulting a pixel size of 2.06 Å.

2.11. Mammalian cell lines

HCT-116 (ATCC® CCL247) cell lines were grown in McCoy's 5a medium modified (Life Technologies) supplemented with fetal bovine serum to a final concentration of 10 %. Hep-G2 (ATCC® HB8065) cell lines were grown in Eagle's Minimum Essential Medium (Life Technologies) supplemented with fetal bovine serum to a final concentration of 10 %. WI 38 (ECACC 90020107) fibroblast cell lines were grown in Minimum Essential Medium Eagle (Sigma) supplemented with fetal bovine serum to a final concentration of 10 % and 2mM L-Glutamine.

2.12. *In vivo* structure probing

In vivo structure probing was performed to assess the influence of the cellular environment on MEG3 structure. Human fibroblast cells WI38 were cultured as described. Live cells were collected using a cell scraper, pelleted by centrifugation for 5 min at RT and 2000 rpm, washed

with PBS and supplemented with 900 µl of fresh growth media. Samples subjected to *in vivo* SHAPE probing, were further supplemented with 100 µl of 100 mM or 250 mM 1M7 in DMSO (10 or 25 mM final concentration, as indicated). Negative control samples were treated with DMSO only. Cells were then incubated for 5 minutes at 37 °C. Media was removed and the cells were washed once with PBS before isolation of total RNA by the RNeasy Mini kit (Qiagen), according to manufacturer's instructions. DNA was additionally digested with Turbo DNase I (Thermo Scientific). Total RNA extract was then cleaned using the Zymogen RNA clean and concentrator kit (Zymogen), according to manufacturer's instructions. The integrity of extracted RNA was checked with RNA 6000 Nano chips (Agilent) on Agilent 2100 Bioanalyzer. RNA was reverse transcribed to cDNA with random nonamers (NEB), using SuperScript™ II RT (Invitrogen) in MaP buffer (125 mM Tris pH8, 187.5 mM KCl, 25 mM DTT, 1.25 mM dNTP, 15 mM $MnCl_2$) that introduces mutations at the sites where 1M7 forms adducts with RNA. As a control a parallel reaction was performed without reverse transcriptase. PCR products were cleaned with PCR clean up kit QIAquick (Qiagen), according to manufacturer's instructions. Residual RNA was digested with RNase A (Sigma). cDNA was amplified with 4 sets of primers (see appendix table 6) with Q5 hot start DNA Polymerase (NEB), according to manufacturer's instructions. PCR reactions were cleaned with DNA clean & concentrate kit (Zymogen), according to manufacturer's instructions. Concentration of all fragments was measured with Qubit3 fluorimeter (Invitrogen). Size and purity of DNA fragments was checked by high sensitivity DNA chips (Agilent) on Agilent 2100 Bioanalyzer, according to manufacturer's instructions. All fragments belonging to same samples were mixed to 0.2 ng/µl. Libraries were tagmented and amplified by Nextera XT DNA library prep kit (Illumina), according to manufacturer's instructions. Libraries were cleaned with AMPure XP beads (Backman Coulter), according to manufacturer's instructions. Library concentration was checked with Qubit3 fluorimeter (Invitrogen) and size distribution by high sensitivity DNA chips (Agilent) on Agilent 2100 Bioanalyzer. Libraries were sent for sequencing to GeneCore (EMBL Heidelberg). Data was processed with ShapeMapper.py script (copyright 2017 Steven Busan). Calculated 1M7 reactivity was compared to average *in vitro* 1M7 reactivity (from triplicates, obtained as described in paragraph 2.8.) by deltaSHAPE.py (Smola et al., 2015b).

2.13. MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay

MTT assay was performed to assess the viability of transfected cells (Mosmann, 1983). Cells were transfected as described above and treated with 12 mM MTT (Sigma-Aldrich) 12h after transfection. Cells treated with MTT were incubated 4 h at 37 °C, 5 % CO_2 and subsequently medium was removed and cells were resuspended in 200 µl DMSO and incubated for 10 min at RT. Absorbance was read at 540 nm with microplate reader CLARIOstar (BMG Labtech).

2.14. Luciferase assay

Luciferase assay was performed to assess p53-mediated transactivation of a luciferase gene in transfected cells. HCT-116 cells were seeded at 83000 cells/well in a cell-culture treated 12-well plate (Costar) and transfected after 24h with 115.96 fmol of pcDNA3 vector containing the indicated MEG3 constructs, 50ng of p53-Luc (given by Yunli Zhou,(Zhou et al., 2007)) and 5 ng of pRL Renilla Luciferase Control Reporter Vector (Promega) using Lipofectamine 2000 (Life Technologies), according to manufacturer's instructions. Transfected cells were incubated for 48 h. Cells were lysed with 1x passive lysis buffer provided in the Dual-Luciferase® Reporter (DLR™) Assay System (Promega). Production of Firefly luciferase was measured by adding Luciferase Assay Reagent II (Promega) and measuring luminescence with a microplate reader CLARIOstar (BMG Labtech). The reaction was then quenched and production of Renilla luciferase was measured by adding Stop & Glo® Reagent (Promega) to normalize the Firefly readout values for transfection efficiency.

2.15. EdU incorporation and flow cytometry assay

EdU incorporation and flow cytometry assay was performed by Isabel Chillon Gazquez (postdoctoral fellow in the Marcia group).

EdU incorporation assay was used to test the ability of MEG3 to suppress cell proliferation. HCT116 cells were grown in 6-well plates and transfected with 1 µg of pCMS-d2EGFP-MEG3 or the equivalent number of molecules of pCMS-d2EGFP-empty or pCMS-d2EGFP-MEG3-ΔH11 plasmids with 5 µl Lipofectamine 2000 (Invitrogen). Time-points were taken at 12, 24, 36 and 48 hours after transfection and EdU incorporation was detected using the Click-iT Plus EdU Flow Cytometry Assay Kit (Molecular Probes) following manufacturer’s instructions with some

modifications. Briefly, at each time point, cells were labelled with 10 µM EdU for 1 hour, washed with PBS and trypsinized. The trypsinized cells were washed with PBS and incubated with LIVE/DEAD Fixable Dead Cell Stain (Invitrogen) for 15 min to evaluate the viability of the transfected cells. Cell were then fixed with 4 % paraformaldehyde solution for 15 min and permeabilized with Click-iT saponin-based permeabilization and wash reagent. After all time points were collected, samples were subjected to the Click-iT reaction following manufacturer's instructions and resuspended in Click-iT® saponin-based permeabilization and wash reagent containing 1 µl FxCycle™ Violet Stain (Invitrogen). For the apoptosis assay, 100,000 cells were incubated with 12 µl of Brilliant Violet 421 annexin V (Biolegend) and 10 µl of 0.5 mg/ml of propidium iodide (Biolegend). Compensation controls were prepared from samples stained with one dye at a time. Flow cytometry was performed on a BD LSR II Flow Cytometer (Becton Dickinson) at the Institute of Advanced Biosciences (Grenoble) flow cytometry core facility. Data analysis was performed using the FCS Express 6 package (De Novo Software).

2.16. Sequence and structural alignments

Final sequence and structural alignments were done by Marco Marcia.

Sequence and structural alignments were performed to assess the potential evolutionary conservation of obtained secondary structure. Sequences corresponding to human domain 2 and 3 were identified in other mammals with BLAT (Kent, 2002). Pairwise alignment in Clustal Omega (Li et al., 2015) was used to align 19 mammal sequences (1-3 sequences from 10 orders of mammals) to the sequence of human MEG3. Alignment was adjusted manually to match the secondary structure of human MEG3. Infernal (Nawrocki and Eddy, 2013) was used to build the covariance model, calibrate it and align all the sequences. R2R (Weinberg and Breaker, 2011) was used to score the alignment files produced by Infernal and to produce a visual map of the co-variation model.

2.17. Quantitative real-time PCR (qRT-PCR)

Quantitative real-time PCR (qRT-PCR) was performed in collaboration with Isabel Chillon Gazquez (postdoctoral fellow in the Marcia group).

Quantitative real-time PCR (qRT-PCR) was performed as a control experiment to assess transfection efficacy and to determine the endogenous level of MEG3 expression in different mammalian cell lines. cDNA was generated from 5 µg total RNA by reverse transcription (RT) using random hexamers (Thermo) and SuperScript II reverse transcriptase (Invitrogen). qRT-PCR was performed on a Mx3005P qPCR system (Agilent) and data were analysed using the Pfaffl method (Pfaffl, 2001). The program comprised of 40 amplification cycles with annealing at 62 °C for 30 s, elongation at 72 °C for 30 s and a final temperature ramp for the generation of melting curve. A fragment of 80 bp within domain 1 (exon 2) of MEG3 was amplified using primers 13F_112_D1Ex2 and 14R_191_D1Ex2. A fragment of 101 bp within domain 2 (exon 3) of MEG3 was amplified using primers 15F_230_D2Ex3 and Meg3RT22. A fragment of 91 bp within domain 3 (exon 3) of MEG3 was amplified using primers 01F_MEG3_Ex3 and 02R_MEG3_Ex3. A fragment of 79 bp within domain 4 (exon 5) of MEG3 was amplified using primers 08F_951_Ex5 and 05R_1029_D4. A fragment of 103 bp within domain 5 (exon 10) of MEG3 was amplified using primers 09F_1077_Ex10 and 10R_1179_Ex10. For the primer sequences see the appendix table 6. Primers were designed with Clone Manager Professional Suite (Sci Ed Central) and examined for possible secondary structures with OligoAnalizer 3.1 (Integrated DNA Technologies). Relative quantification was calculated using the Pfaffl method (Pfaffl, 2001). Beta-actin mRNA was used as reference to normalize for total cellular RNA. Neomycin mRNA was used as a reference to normalize for transfection efficiency. Statistical analyses were performed using the Prism 6 package (GraphPad Software).

3. Results

Résumé en Français

En adaptant un protocole de purification non dénaturant établi pour d'autres ARNlnc, j'ai pu obtenir des rendements élevés en MEG3 *in vitro* (~ 3 µM dans 500 µl, soit 0,8 mg au total à partir de 1 ml de réaction de transcription). J'ai utilisé plusieurs tests biochimiques pour confirmer que ma préparation de MEG3 était très pure et homogène. L'obtention de MEG3 homogène m'a permis de déterminer la carte expérimentale de structure secondaire de deux variants d'épissage de MEG3 qui ont des niveaux différents de transactivation de p53 (le variant 1 le plus abondant et le variant 9 induisant un degré plus élevé de transactivation de p53). J'ai déterminé la structure secondaire par acylation 2'-hydroxyle sélective *in vitro*, analysée par extension d'amorce (SHAPE) en utilisant 3 réactifs (1M7, 1M6 et NMIA). J'ai utilisé un quatrième réactif (DMS) pour valider la structure secondaire de v1, que je prends comme isoforme de référence pour la description structurale. Mes données (obtenues par sonde chimique) suggèrent que MEG3 humain possède une structure secondaire robuste dans laquelle la moitié de l'extrémité 5' (D1-D3) est bien définie et structuralement conservée entre plusieurs isoformes d'épissage, et la moitié de l'extrémité 3' (D4-D5) est plus flexible, variable à travers les isoformes d'épissage. La moitié de l'extrémité 3' (D4-D5) est également plus réactive *in vitro* qu'*in vivo*, ce qui indique qu'elle pourrait être impliquée dans la liaison aux protéines *in vivo*. Les domaines D2 et D3 (exon 3) sont la partie la plus conservée de MEG3 en séquence et en structure et possèdent un motif structuré - H11 - invariant chez les mammifères.

Puisque j'ai pu obtenir une homogénéité excepetionnelle pour MEG3, qui n'a jamais été obtenue auparavant pour aucun autre ARNlnc, j'ai aussi cherché à caractériser la structure tertiaire de MEG3. Par AUC, j'ai démontré que l'addition de magnésium induit la formation de particules de MEG3 de plus en plus compactes et j'ai pu visualiser ces particules en 3D par AFM.

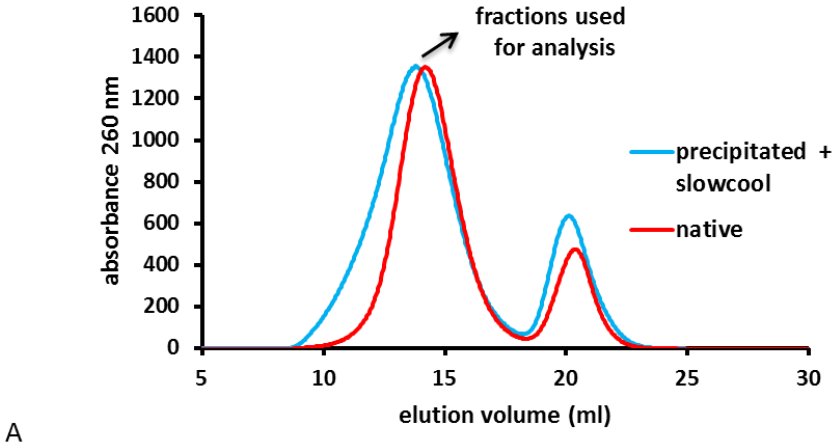
Pour évaluer la corrélation entre les domaines structurels et le rôle fonctionnel de MEG3 en tant que suppresseur de tumeur, j'ai mené une série d'essais à la luciférase qui mesurent la transactivation médiée par p53 par différentes constructions de MEG3. J'ai découvert que le noyau activant p53 de MEG3 est formé de D2 et de D3 et implique plus particulièrement les deux motifs structurels H11 (D2) et H25-H29 (D3). La structure du tronc de H11 hautement

conservée et chaque nucléotide dans la boucle terminale H11 invariante est essentiel pour l'activation de p53. H25-H29 (D3) contient 6 répétitions en tandem qui forment des boucles unies à longue distance avec la boucle terminal de H11.

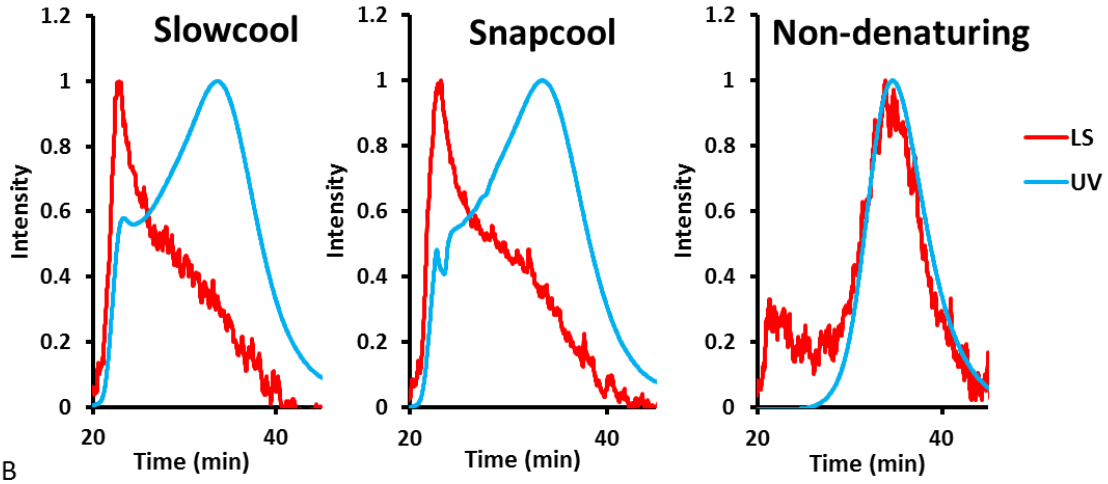
Pour vérifier si la structure de MEG3 était essentielle à son activité, je me suis efforcée de déterminer si les mutants de la boucle terminale H11 (en particulier le mutant nommé H11LpA) perturbaient la structure secondaire et/ou tertiaire de MEG3. En effet, selon l'AUC, j'ai confirmé que le mutant H11LpA n'est pas capable de compacter autant que MEG3 sauvage et j'ai confirmé celle observation avec AFM. En comparant le profil SHAPE du type sauvage MEG3v1 et du mutant H11LpA, j'ai constaté que la plupart des nucléotides ont la même réactivité dans les deux ARN. Fait intéressant, la boucle terminale fonctionnelle H11 est gravement affectée par toutes les positions non réactives dans le sauvage devenant très réactives dans le mutant poly A. Un changement de réactivités entre MEG3v1 et H11LpA se vérifie aussi dans H25-H29. Dans cette domaine structural, la région qui modifie le plus la réactivité globale au SHAPE est l'extrémité 3' du tronc de H27 (nt 857-881).

3.1. MEG3 production

To establish the best way for MEG3 production, I tried 4 different protocols and compared the homogeneity of samples with SEC and SEC-MALLS (figure 8).



A



B

Figure 8: MEG3 production and purification (denaturing vs non-denaturing protocol).
(A) SEC chromatogram. Absorbance at 260 nm is depicted as a function of elution volume. Red line MEG3 sample produced in non-denaturing conditions, blue line MEG3 sample precipitated after transcription, denatured and refolded by slowcool protocol. Second peak on the chromatogram are probably leftover free nucleotides from transcription reaction. (B) SEC-MALLS profile. 0.5µM MEG3 on S500 column (red line light scattering, blue line UV absorption at 260 nm). From left to right: MEG3 sample produced by denaturing slowcool protocol, MEG3 sample produced by denaturing snapcool protocol and MEG3 sample produced by non-denaturing protocol. (See paragraph 3.1 for details).

After transcription, just before running the sample on SEC (see paragraph 2.2) I kept some of the native sample; I denatured 2 samples by heating to 95°C for 4 min and cooled the sample down to refold either quickly by incubating on ice for 5 min (snapcool) or gently resting it at room temperature for an hour (slowcool); finally I precipitated a fourth sample in isopropanol, resuspended it in filtration buffer, denatured it by heating to 95°C for 4 min and gently allowed it to cool down back to room temperature. From the comparison of the SEC chromatograms of native and precipitated/denatured samples we can see that the native protocol results in sharper peak, representing a more homogenous sample (figure 8A). From the comparison of the SEC-MALLS profiles of denatured and native samples we could clearly see that the two denatured and subsequently refolded samples (snapcooled and slowcooled) contained aggregates which interfered with light scattering whereas the native sample had matching peaks of UV absorption and light scattering (figure 8B). Based on those results I concluded that the native sample production protocol is optimal because it gives the most homogenous samples. Adapting a non-denaturing purification protocol established for other lncRNAs, I could obtain high yields of MEG3 *in vitro* (~ 3 µM in 500 µl, total 0.8 mg from a 2 h transcription reaction).

3.1.1. Homogeneity and purity of MEG3

I used several biochemical tests to confirm that my MEG3 preparation is very pure and homogeneous (figure 9). First, I preformed native agarose gel electrophoresis at several check points during purification. In those gels, I could observe two bands corresponding to two different structural conformations of RNA in samples without any divalent cations and a presence of a single migration band when I added Mg^{2+} to the sample (figure 9A). Second, SEC performed directly after purification reveals a high intensity symmetric elution peak (figure 9B). Third, AUC further confirmed that MEG3 is very homogenous at low concentrations of magnesium. A small amount of aggregates starts forming at 7.5 mM Mg^{2+}, likely because Mg^{2+} promotes formation of intermolecular interactions. A single major species of MEG3 can be clearly detected up to ~25 mM $MgCl_2$. At higher magnesium concentrations aggregation is dominant. Fourth, in line with electrophoretic, SEC and AUC results, DLS of MEG3 samples in a buffer without magnesium reveals a single-peak size distribution by intensity, mass and volume

(figure 9C). Analogously, SEC-MALLS of MEG3 in a buffer without magnesium further confirm that in this condition MEG3 does not aggregate or degrade and that it forms monomeric particles (figure 9D). Small amounts of aggregates induced by magnesium have so far precluded accurate determination of MEG3 monodispersity and oligomeric state by DLS and SEC-MALLS in the presence of divalent ions (see paragraph 3.3.2.).

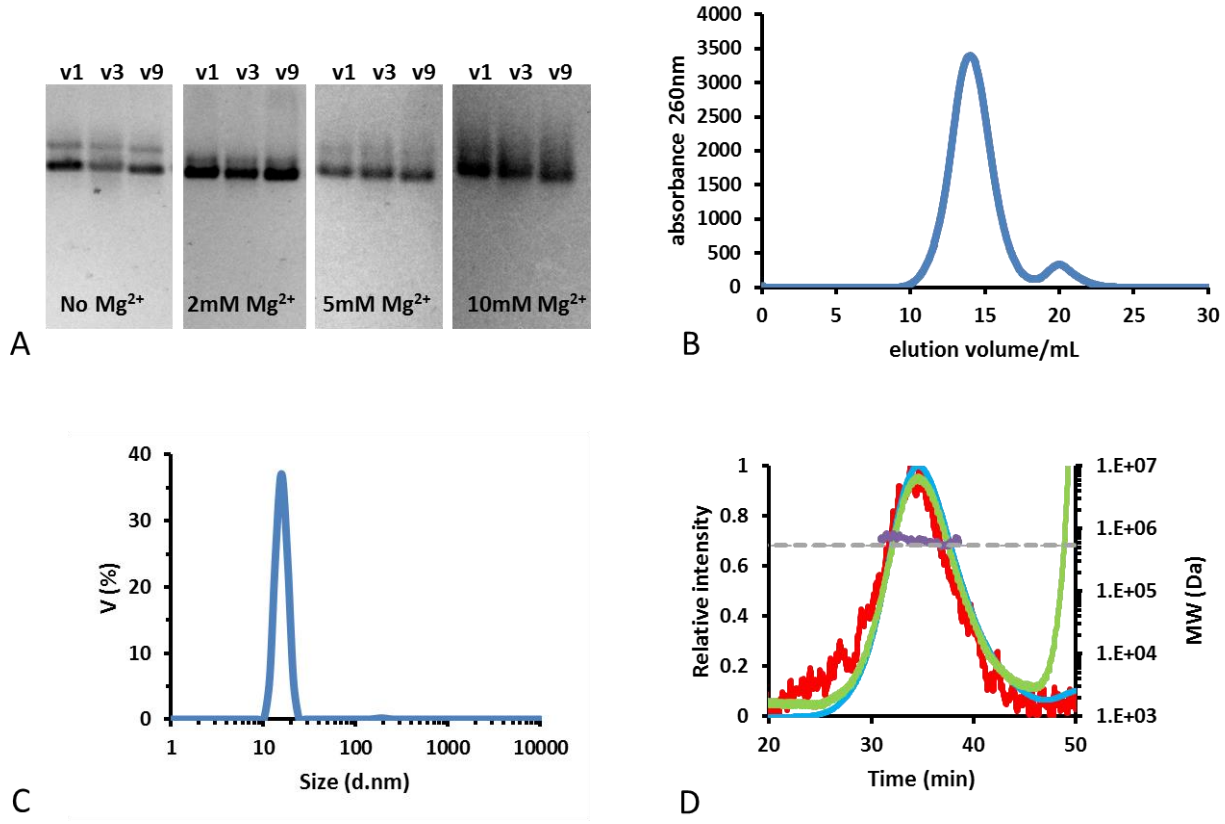


Figure 9: MEG3 purity and homogeneity.

(A) Native agarose gel of MEG3 splicing variants 1, 3 and 9 after *in vitro* transcription with increasing concentration of magnesium from left to right. (B) SEC chromatogram of MEG3v1 after *in vitro* transcription. Absorbance at 260 nm is depicted as a function of elution volume. (C) DLS spectrum of MEG3v1 after *in vitro* transcription. Volume distribution of particle sizes in 0.5 µM MEG3 sample. (D) SEC-MALLS profile of MEG3v1 after *in vitro* transcription. 0.5µM MEG3v1 on S500 column (red line light scattering, blue line UV absorption at 260 nm, green line differential refractive index, purple line calculated molecular weight, dashed line expected molecular weight of MEG3v1).

3.2. MEG3 secondary structure

Obtaining homogeneous MEG3 allowed me to determine the experimental MEG3 secondary structure map by chemical probing (SHAPE) (appendix figures 1, 2 and 3). I determined the structure of MEG3v1 and v9 using 3 reagents (1M7, 1M6 and NMIA). Additionally, I used a fourth reagent (DMS) to validate the obtained map of MEG3v1. Finally, Eleni Anastasakou (predoctoral fellow in the Marcia group) mapped the secondary structure of MEG3v3.

3.2.1. *In vitro* secondary structure probing of MEG3 variant 1

My secondary structure map reveals that MEG3v1 is organized in 5 highly-structured domains (figure 10). Domain 1 (nt 2-196) is arranged in 1 five-way junction; domain 2 (nt 230-410) is arranged in 2 three-way junctions; domain 3 (nt 471-902) is arranged in 5 three-way junctions and 1 five-way junction; domain 4 (nt 951-1113) is arranged in 2 three-way junctions; domain 5 (nt 1116-1486) is arranged in 2 three-way junctions, 1 five-way junction and 1 six-way junction. In total MEG3 forms 16 multi-way junctions and possesses 52 helices (appendix table 2).

Further, comparing Shannon entropy values smoothened over a 55-nt sliding window with SHAPE reactivities, I identified 5 structural motifs with low SHAPE reactivity and low Shannon entropy meaning that they are structured and well-defined (figure 11). Motif 1 (nt 1-410) covers domains 1 and 2 (E1-2-3). Motif 2 (nt 568-692) is part of domain 3 (E3) and is arranged in a three-way junction with 3 helices. Motif 3 (nt 755-885) is also a part of domain 3 (E3) and is arranged in 2 three-way junctions with 4 helices. Motif 4 (nt 955-1409) covers almost whole domain 4 (E5-10). Motif 5 (nt 1127-1475) covers almost whole domain 5 (E10-11-12).

Interestingly, the boundaries of the structural domains of MEG3 correspond to the splicing boundaries of the MEG3 exons (figure 10). Domain 1 is formed by exons 1 and 2, domains 2 and 3 are formed by exon 3, domain 4 is formed by exon 5 and part of exon 10 and domain 5 is formed by exons 10-11-12. Since exons 1-2-3 and exons 10-11-12 are common to most MEG3 splicing variants, this result suggest that domains 1-2-3 and domain 5 may be common to many MEG3 splicing variants, whereas domain 4 may be variable and thus responsible for functional differences between variants. To test this hypothesis, I determined the experimental secondary structure of MEG3v9 by chemical probing using the same three SHAPE reagents, as for v1.

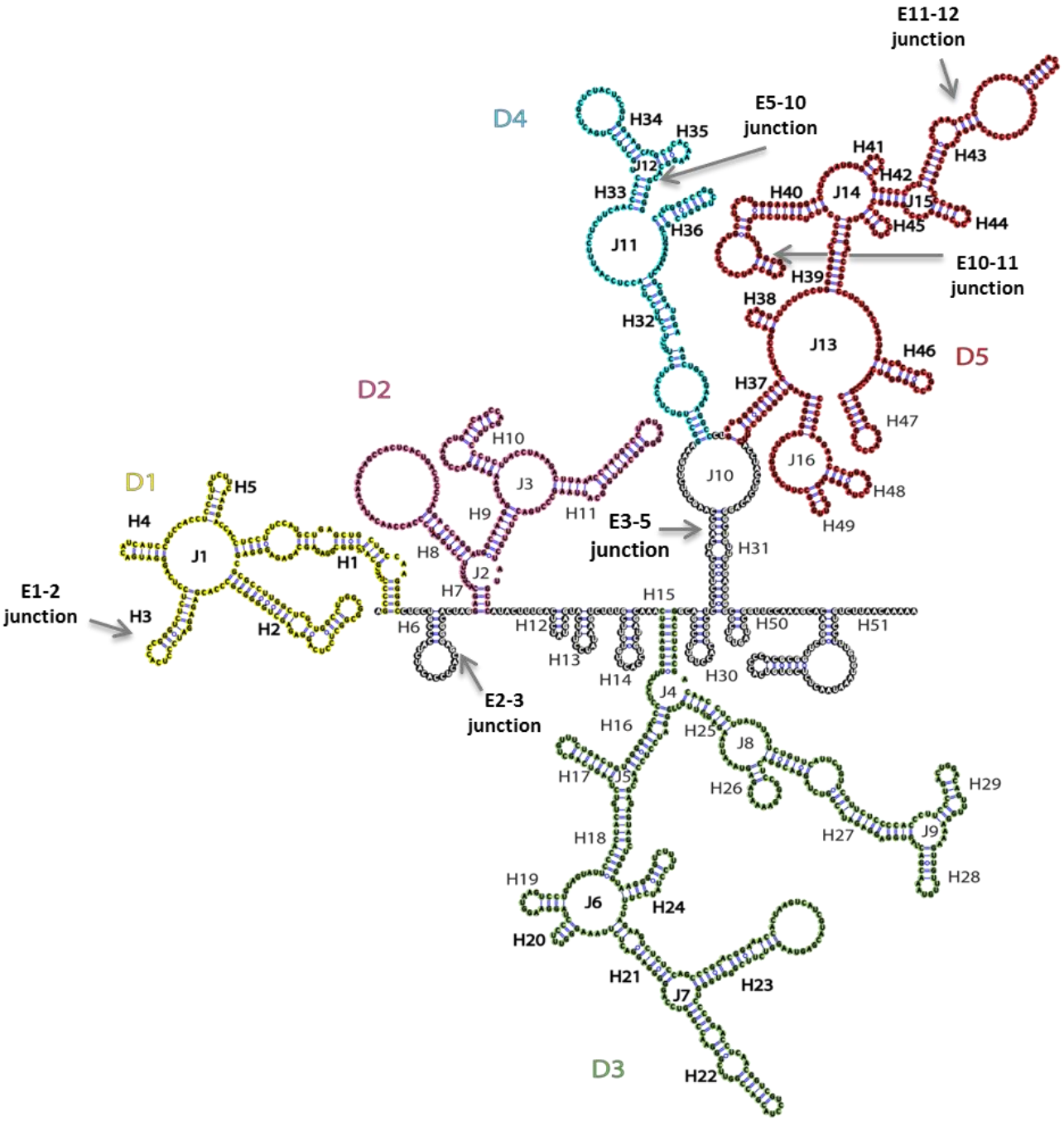


Figure 10: Experimental secondary structure map of MEG3v1 obtained by chemical probing (SHAPE). MEG3v1 secondary structure map divided in 5 domains colored yellow, pink, green, cyan and red respectively. Helixes are marked with H and respective number and junctions with J and respective number. Junctions between exons are indicated with grey arrows.

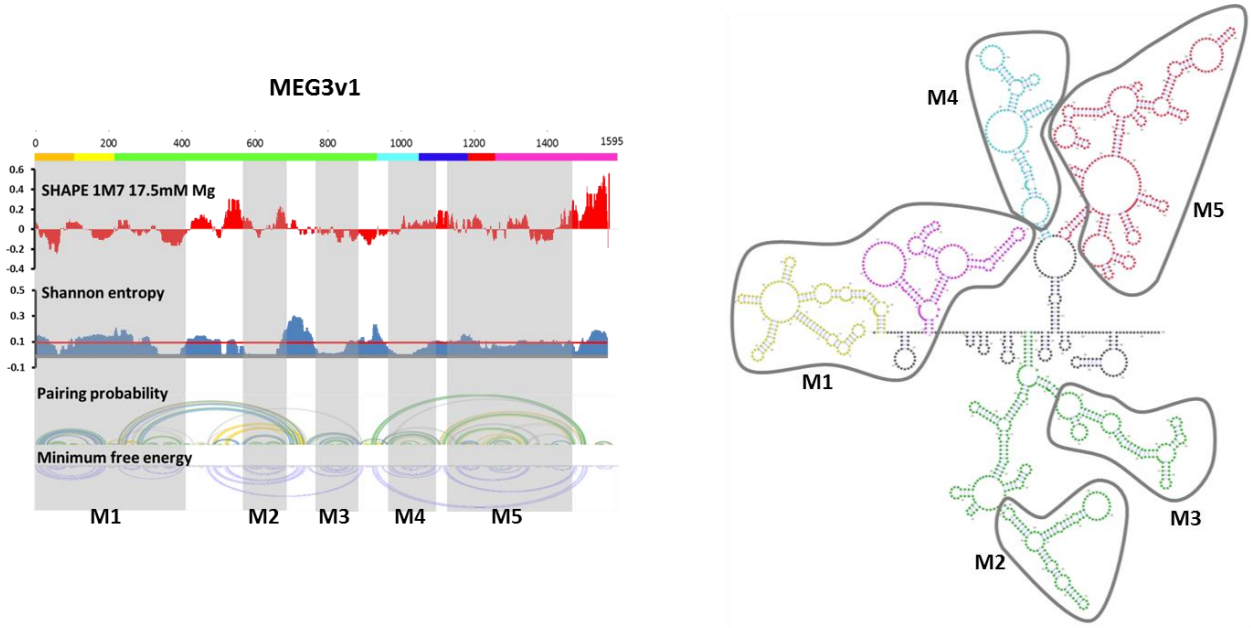


Figure 11: Experimental secondary structure maps of MEG3v1 with indicated structured motifs.

On the left: (1st line) Profile of SHAPE reactivity. Positive values indicate flexible regions, negative values indicate rigid regions. (2nd line) Profile of Shannon entropies (S). The red line marks the median entropy of all nucleotides. Regions with S < median are structurally well-defined. (3rd line) Arc representation of base-pairing probabilities (green arc 80-100 %, blue 30-80 %, yellow 10-30 %, grey 0-10 %). (4th line) Arc representation of the minimum free energy structure. Well-defined structure motifs with low SHAPE reactivity and low Shannon entropy are marked with grey shading. On the right: secondary structure map of MEG3v1. Grey line indicates regions of low Shannon entropy low SHAPE reactivity of respective variants. Nucleotides are colored to represent the domains of MEG3v1.

3.2.2. *In vitro* secondary structure probing of MEG3 variant 9

The secondary structure map of MEG3 variant 9 is organized in four domains (figures 12 and 13). Domain 1 (nt 2-196) is arranged in 1 five-way junction; domain 2 (nt 231-464) is arranged in 1 three-way junction and 1 five-way junction; domain 3 (nt 466-904) is arranged in 5 three-way junctions and 1 five-way junction; domain 4 (nt 905-1431) is arranged in 1 seven-way junction and 2 three-way junctions. In total MEG3v9 forms 12 multi-way junctions and possesses 42 helices (figure 13).

I compared average 1M7 reactivity (from triplicates) of variants 1 with variant 9 by calculating the Spearman's correlation coefficient for all nucleotides that exist in two variants (figure 14). The resulting Spearman's correlation coefficient between variant 1 and 9 of 0.7847 is in line with correlation between two different replicas (0.6774 -0.8662) indicating similar reactivity profile between different splicing variants with same exons. Comparison of the secondary structure maps of different splicing variants reveals that almost all structural motifs within domains 1, 2 and 3 are the same in MEG3v9. On the contrary, secondary structures of the domains 4 and 5 are significantly different. In MEG3v9, D4 and D5 are merged together departing from a common seven-way junction, and D4 is reduced to a single helix, while D5 forms 2 three-way junctions and 7 helices.

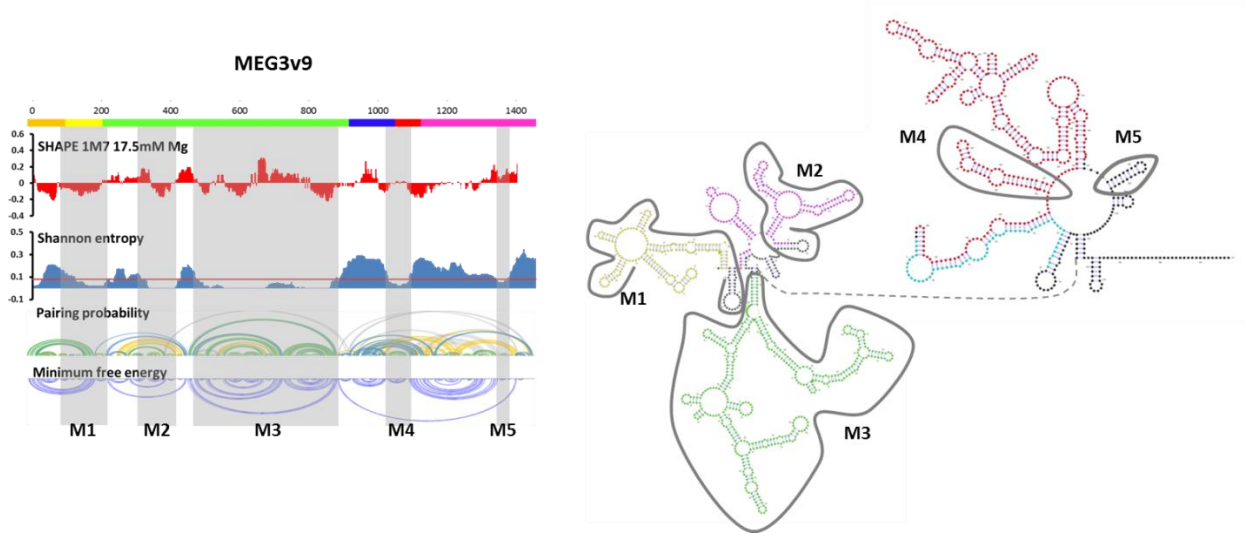


Figure 12: Experimental secondary structure maps of MEG3v9 with indicated structured motifs.

On the left: (1st line) Profile of SHAPE reactivity. Positive values indicate flexible regions, negative values indicate rigid regions. (2nd line) Profile of Shannon entropies (S). The red line marks the median entropy of all nucleotides. Regions with S < median are structurally well-defined. (3rd line) Arc representation of base-pairing probabilities (green arc 80-100 %, blue 30-80 %, yellow 10-30 %, grey 0-10 %). (4th line) Arc representation of the minimum free energy structure. Well-defined structure motifs with low SHAPE reactivity and low Shannon entropy are marked with grey shading. On the right: secondary structure map of MEG3v9. Grey line indicates regions of low Shannon entropy low SHAPE reactivity of respective variants. Nucleotides are colored to represent the domains of MEG3v1.

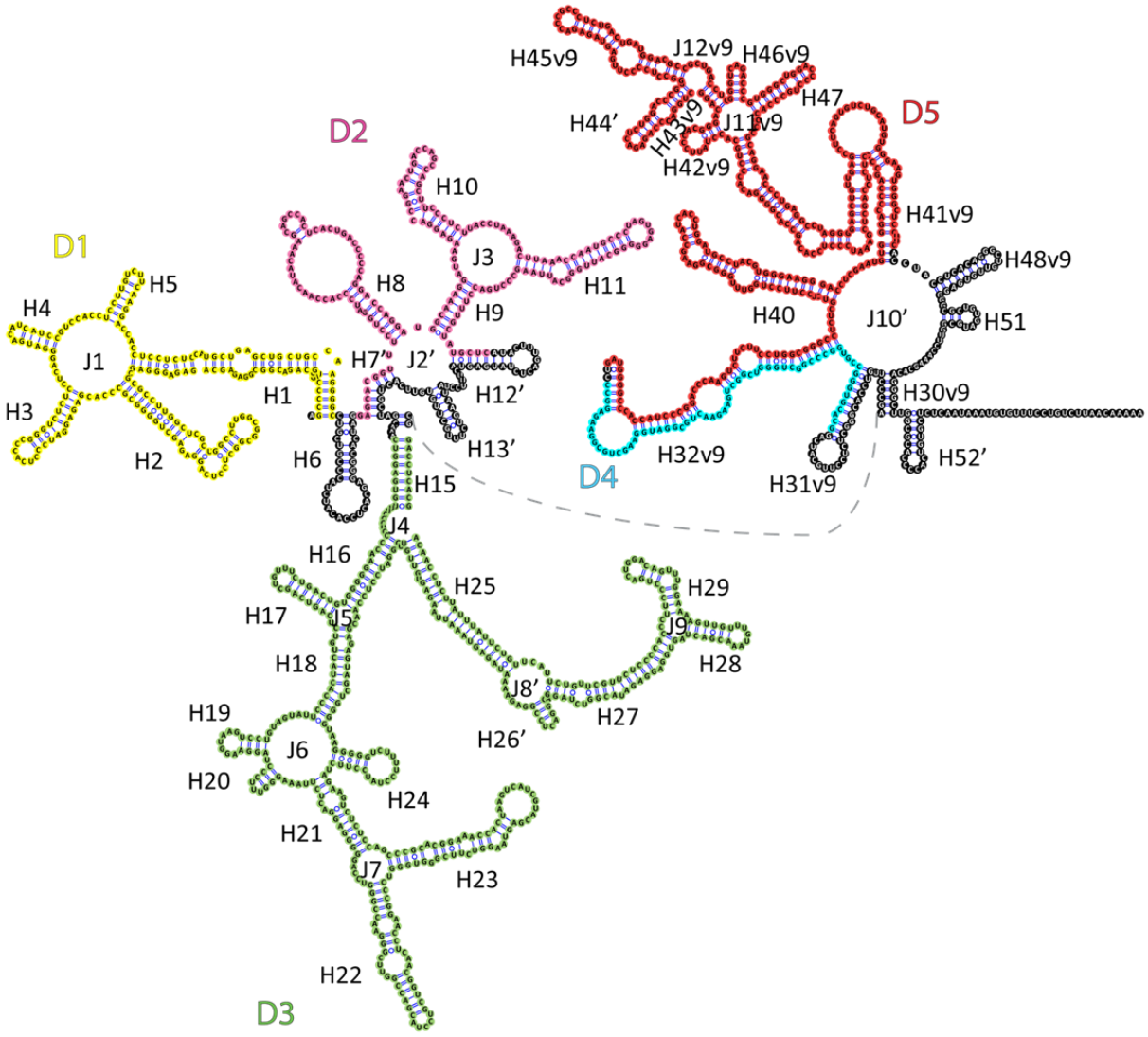


Figure 13: Experimental secondary structure map of MEG3v9 obtained by chemical probing (SHAPE).

MEG3v9 secondary structure map. Nucleotides are colored to represent the domains of MEG3v1 (yellow, pink, green, cyan and red respectively). Helixes are marked with H and respective number and junctions with J and respective number.

Regarding the motifs with low Shannon entropy/low shape reactivity, first 3 motifs that are part of D1-3 in MEG3v1 are also present in v9, whereas regions that correspond to motif 4 and 5 which are part of D4-5 in MEG3v1 are much smaller in v9. Motif 1 (nt 93-203) is part of domain 1 (E1-2) and matches the first part of motif 1 in v1. Motif 2 (nt 295-431) is part of domain 2 (E3) and matches the second part of motif 1 in v1. Motif 3 (nt 466-904) covers domain 3 (E3) and matches motifs 2 and 3 in v1. Motif 4 (nt 1052-1107) is one helix in last domain (E10-11), and matches a small part of motif 5 in v1. Motif 5 (nt 1379-1399) corresponds to one helix in last domain (E12) and it doesn't correspond to any motif present in v1.

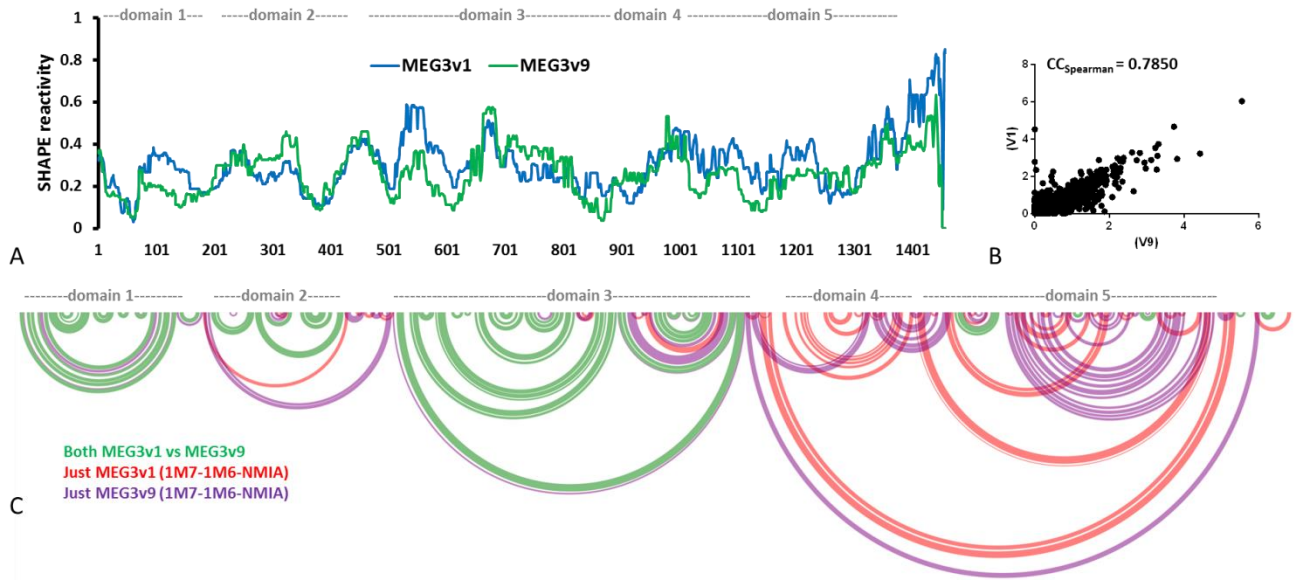


Figure 14: Comparing experimental MEG3v1 and v9 secondary structures obtained by chemical probing (SHAPE).

(A) 1M7 reactivity smoothened over 55 nucleotide window median. X-axes nucleotide numbers corresponding to v9. Domains correspond to v1. Blue line MEG3v1, green line MEG3v9. (B) Correlation graphs of 2 variants with indicated Spearman's correlation coefficient. (C) Arc representation of MEG3v1 and v9 (green arcs = base pair present in both variants, red arcs = base pairs present only in MEG3v1, purple arcs = base pairs present only in MEG3v9). Domains correspond to v1. Missing nt in v9 represented as non-base-paired.

3.2.3. *In vitro* secondary structure probing of MEG3 variant 3

Determining the structure of variant 3 is the subject of the thesis of my colleague Eleni Anastasakou, so a complete description of v3 will not be reported here. However, certain structural features of v3 were useful to derive structural implications for v1 and v9 and to design experiments that are reported in the following sections. Such structural features of v3 are:

- local rearrangements at the 5'-end, due to the missing 5'-terminal 25 nt
- local rearrangements at the base of D3, due to insertion of a two-stem-loop structure formed by E5
- overall D2 and part of D3 is very similar to v1 and 9
- D4-5 are significantly different between splice variants although there are some small motifs formed by common exons E10-12 that are structurally preserved (H44, H47, and H52).

Overall, the regions that differ the most between splice variants correspond to inter-domain likers (H6, H12, H13, H14, H30, and H51). From comparing the 3 secondary structure maps of v1, v3 and v9 I can conclude that insertion or deletions of exon in middle varying region induces structural changes in the 3'-half of MEG3.

3.2.2. Sequence and structural alignments

The Rfam database (RF01872) identifies a motif of MEG3 in 40 mammalian species and 53 sequences. This motif of MEG3 is part of exon 3 and corresponds to part of D3 (more specifically H21-H23) in MEG3v1. I wanted to expand the search to the whole MEG3 domain by domain. I started the search with the structurally most preserved D2-3 (E3). Using the BLAT search on genomic DNA sequences (Kent, 2002), I could identify sequences corresponding to D2-3 in 46 mammals, covering all mammalian orders, except *Monotremata* [orders defined according to (Tarver et al., 2016)]. Although, I could identify the presence of the sequence in DNA of 46 mammals I could find the annotated transcripts (NCBI) of only 6 species (human, cow, mouse, orangutan, rat, and pig). In comparison to other domains, I could identify the

presence of D2-3 in most mammalian species indicating that D2-3 is not only the structurally most preserved but it is also the evolutionarily most conserved region of MEG3. More specifically, I found the presence of D1 (E1-E2) sequence in 33 species. I could not identify D1 in *Prosimians, Eulipotyphla, Xenarthra, Afrotheria*, and *Marsupialia*. I found the presence of D4-5 (E10-12) sequence in 19 species. I could not identify D4-5 in *Xenarthra* and *Marsupialia*. Overall, I could not detect any sequence similarity to MEG3 beyond mammals.

Since D2-D3 (E3) has the most preserved secondary structure between 3 different splicing variants and is the most conserved region of MEG3 as well, I decided to perform covariation analysis on these two domains. Covariation analysis allows to assess potential evolutionary conservation of RNA secondary structures. I selected 19 sequences homologous to D2-3 by choosing 2-3 sequences (species) from each mammalian order. I aligned those sequences using Clustal Omega (Li et al., 2015) and subsequently used that alignment together with secondary structure of MEG3v1 aligned to human sequence as a seed in Infernal (Nawrocki and Eddy, 2013). Final alignment file produced by Infernal contained MEG3 D2-3 sequences from 41 mammalian species. 5 species were not aligned because too divergent. To visualize this alignment I used R2R (Weinberg and Breaker, 2011). R2R revealed a high degree of potential evolutionary conservation of the structure. To assess the statistical significance of such potential structural conservation, I run RScape (Rivas et al., 2017; Tavares et al., 2018) on the alignment produced by Infernal. By doing so I found base pairs with statistically significant covariation and this was reproducible using different windowing and sliding options (Tavares et al., 2018). The statistically significant covariant base pairs are part of H8, H9, H10, and H11 (in D2), and in H17, H19, H21, H22, and H28 (in D3). Within this region, H11 is the most conserved structural element, with part of its stem and entire loop being identical in all aligned sequences (in all 41 mammalian species) and 7 base pairs being significantly covariant (figure 15).

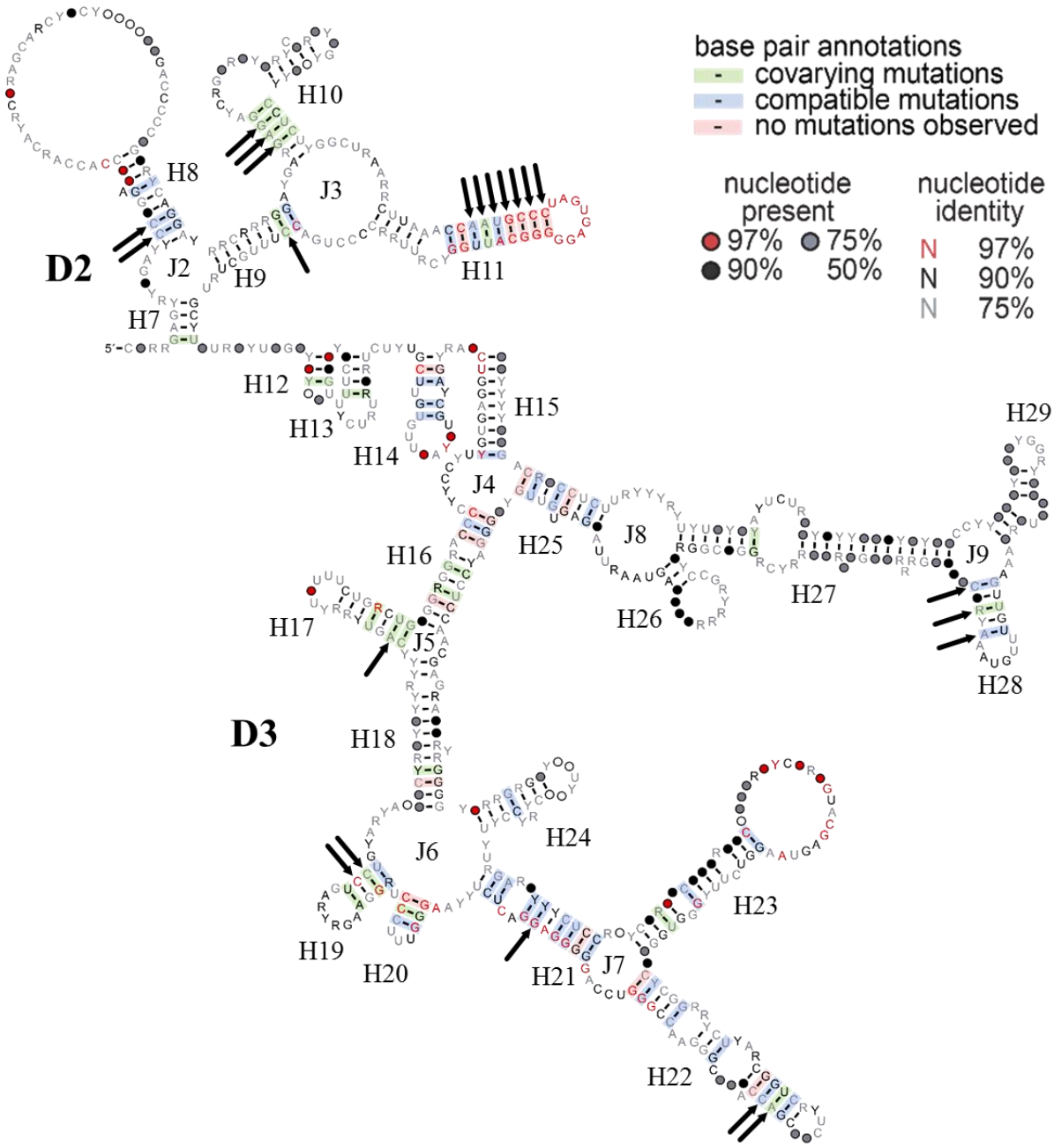


Figure 15: Sequence covariation in MEG3 domains 2 and 3.

R2R plot of 41 D2-D3 (E3) sequences aligned in Infernal [colour legend in the upper right corner (Weinberg and Breaker, 2011)]. Arrows indicate significantly covariant base-pairs (RScape E values < 0.05). D-domain, E-exon, H-helix, J-junction.

3.2.3. *In vivo* chemical probing

Having mapped the secondary structure of MEG3v1 and v9 *in vitro*, I wanted to assess the effect of cellular environment on the structure so I performed *in vivo* chemical probing with 1M7 (Smola et al., 2015b; Watters et al., 2016). I used a WI38 fibroblast cell line that expresses MEG3 endogenously, 68 % MEG3 variant 1, 26 % variant 9, 3 % variant 15, 2 % variant 10 and 1 % variant 8 (Zhang et al., 2010a). I chose this cell line since it expresses MEG3 endogenously at significantly high levels (~ 50 fold less than actin mRNA, corresponding to ~40 copies per cell for MEG3v1, as Isabel Chillon confirmed by qRT-PCR, appendix figure 4). Amplifying the cDNA with primers that anneal to MEG3 sequence and sending the PCR product for sequencing I could identify v1 and v9, which couldn't be separate even by running the PCR product on agarose gel and extracting the DNA from the desired band. Drawback of this cell line is that it expresses multiple splicing variants, and I could not unambiguously separate the reacted fragments produced by MEG3v1 from those produced by MEG3v9. To overcome this problem, I also performed *in vivo* 1M7 probing on HCT116 cells (no endogenous expression of MEG3, appendix figure 4) transfected with MEG3v1. All experiments were done in biological triplicates and 1M7 reactivity was compared to *in vitro* 1M7 reactivity values identifying regions with modified reactivity values between these two conditions (figures 16 and 17). Regions more reactive *in vitro* indicate potential protein binding sites and the regions more reactive *in vivo* indicate conformational change (Smola et al., 2015a). In all biological replicas of both endogenous and transfected MEG3 overall the *in vivo* and *in vitro* 1M7 reactivity profiles were similar for D1-3 but *in vivo* 1M7 reactivity was significantly lower in D4-5 suggesting that the 3'-terminal half of MEG3 is structurally more flexible and prone to conformational changes than the 5'-terminal half. More precisely, whole D4-5 and the small motifs within D3 (H20-22, H28-H29) are less reactive *in vivo* being potential protein binding sites. Whereas H2, H10, H23, and H25-H26 within D1-3 are more reactive *in vivo* being potential places of conformational change upon protein binding in the cell. From *in vitro* and *in vivo* secondary structure studies of MEG3 I can conclude that MEG3 possesses an overall robust secondary structure, with D2-3 representing the structural core and comprising part of the well-defined and structurally conserved 5'-terminal half, and with D4-5 (the 3'-terminal half) being more flexible, variable across splice isoforms and possibly involved in protein binding *in vivo*.

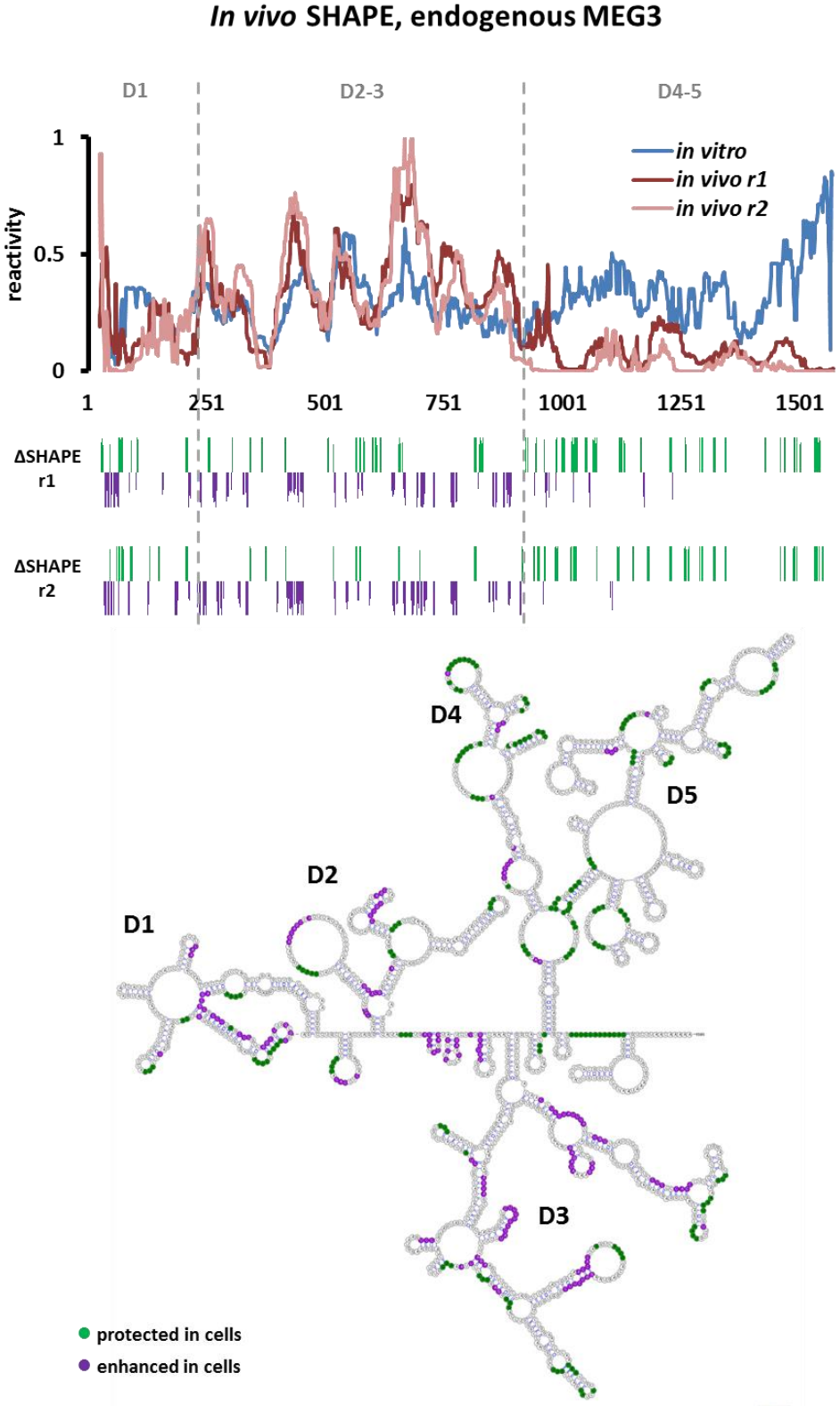


Figure 16: *In vivo* probing of endogenous MEG3 in WI38 cells.

From top to bottom: 1M7 reactivity smoothened over 55 nucleotide window median of two *in vivo* replicas (in dark and light red) and *in vitro* (in blue); ΔSHAPE profiles of 2 replicas (generated by deltaSHAPE (Smola et al., 2015a), regions protected in cells, putative protein binding sites, in green and regions more reactive in cells, putative sites of conformational change, in purple); annotation of the ΔSHAPE profile on the *in vitro* secondary structure map of v1.

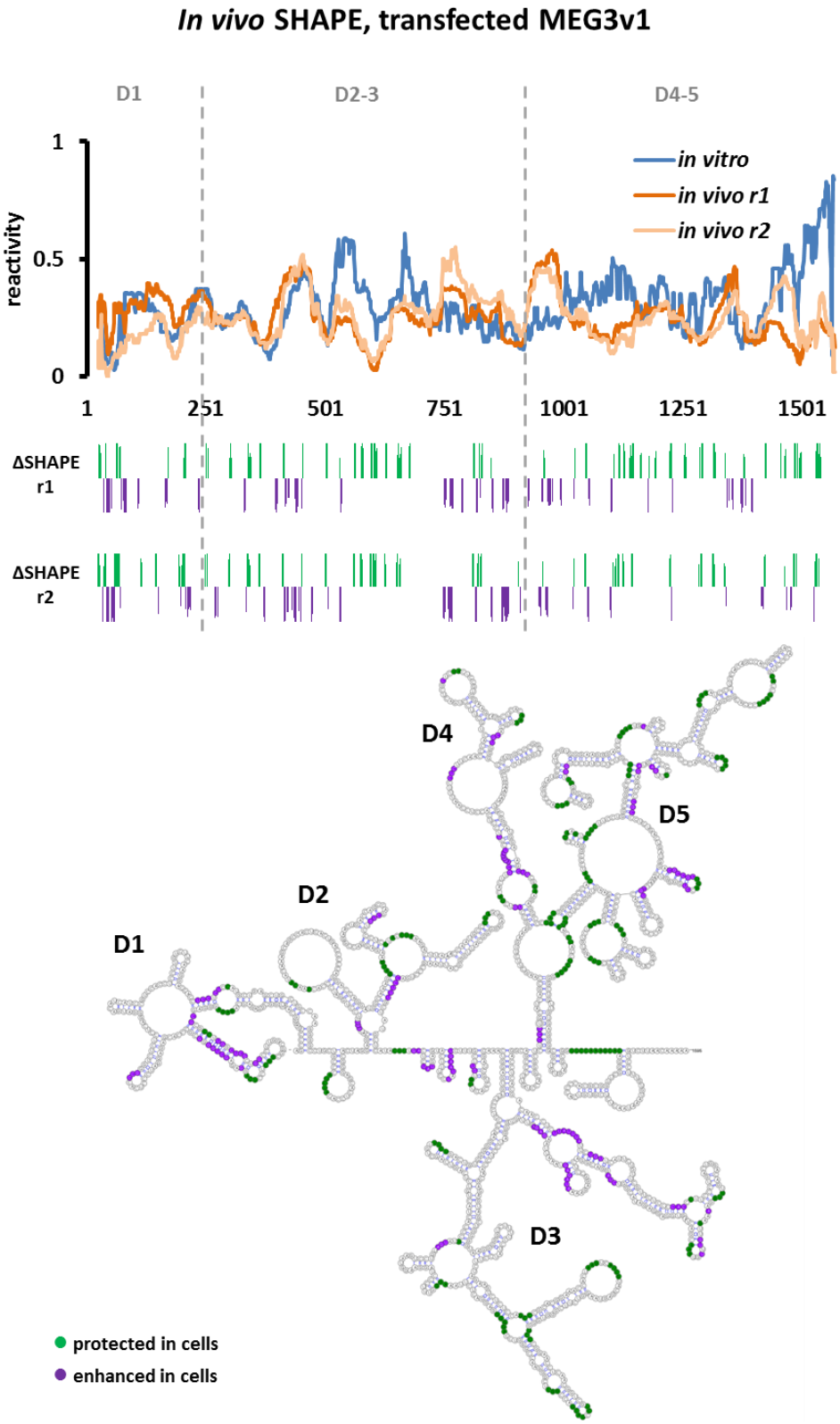


Figure 17: *In vivo* probing of MEG3v1 transfected in HCT116 cells.

From top to bottom: 1M7 reactivity smoothened over 55 nucleotide window median of two *in vivo* replicas (in dark and light orange) and *in vitro* (in blue); ΔSHAPE profiles of 2 replicas (generated by deltaSHAPE (Smola et al., 2015a), regions protected in cells, putative protein binding sites, in green and regions more reactive in cells, putative sites of conformational change, in purple); annotation of the ΔSHAPE profile on the *in vitro* secondary structure map of v1.

3.3. MEG3 tertiary structure

The secondary structure map of MEG3 indicates that MEG3 is a highly structured RNA. First, it possesses many structured motifs (appendix table 2) comparable to other lncRNA (appendix table 3). Second, the structure also allows localization of motifs potentially involved in long range tertiary interactions. For instance, in domain 5 (motif 5) one terminal loop 5'-GGAA-3' (nt 1185-1188) is a well characterized GNRA motif that usually acts as "anchor" during tertiary folding (Jaeger et al., 1994). Third, I can obtain very high homogeneity (figure 9). Based on these considerations, I set out to characterize the MEG3 tertiary structure.

3.3.1. Compaction of MEG3 by AUC

RNA folding *in vitro* can be induced by addition of divalent or monovalent cations (Draper, 2004, 2013; Pyle, 2002). I have established that addition of magnesium induces formation of increasingly compact particles of MEG3 that possess larger sedimentation coefficients and correspondingly smaller hydrodynamic radii (R_h) from 12 nm to 9 nm. MEG3 folds upon addition of magnesium, as indicated by the increase in sedimentation coefficient and by the hypochromicity effect (folded particles display reduced UV absorption because of base stacking) in the AUC analysis. A fit of the magnesium-dependent changes in the R_h of MEG3 using the Hill equation determined that the magnesium concentration for which 50 % MEG3 is folded is $C_{Mg1/2}$ = 6.9±0.35 mM and the Hill coefficient of the fitting equation is n_{Hill} = 3.9±0.6 (figure 18: A, B). However, in the presence of magnesium MEG3 forms some aggregates which makes it unsuitable for analysing with scattering methods such as MALLS and SAXS. Alternatively to Mg^{2+} ions, RNA compaction can also be induced by non-physiological K^+ ions concentration (Woodson, 2005). I determined that potassium compacts MEG3 from R_h = 11.51 (in presence of 200 mM KCl) to R_h = 10.13 nm (in presence of 900 mM KCl) reaching a form which is less compact than with magnesium (figure 18: C, D). Moreover, MEG3 aggregates in the presence of high potassium concentrations which make potassium also unsuitable for studying compact form of MEG3 with scattering methods.

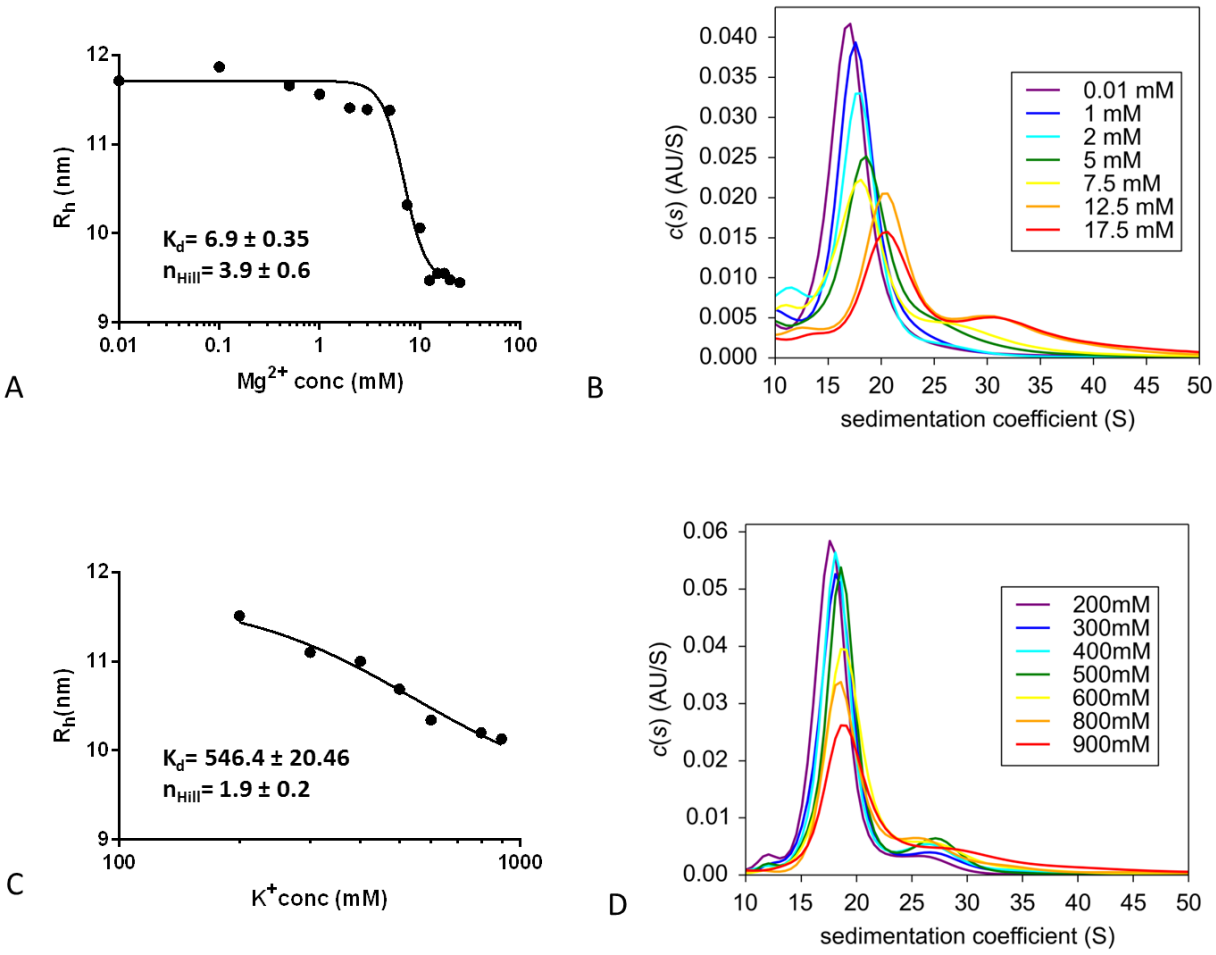


Figure 18: MEG3 folding by AUC.

(A) Hill plot of the hydrodynamic radii of MEG3v1 with different concentrations of magnesium derived from the SV-AUC experiment. (B) SV-AUC profiles of MEG3v1 with different concentrations of magnesium. (C) Hill plot of the hydrodynamic radii of MEG3v1 with different concentrations of potassium derived from the SV-AUC experiment. (D) SV-AUC profiles of MEG3v1 with different concentrations of potassium.

3.3.2. Optimization of folding conditions by MALLS

Since at high concentration of cations I could see that MEG3 aggregates, I tried to find a condition at which MEG3 would not aggregate and thus be suitable to study with light scattering methods. At lower concentration of $MgCl_2$ (5 mM) I could not observe the aggregation by AUC but it was evident with SEC-MALLS. To overcome this problem I tried preparing MEG3 RNA sample in several different ways before injecting it on SEC-MALLS. I

always produced the RNA as described (see paragraph 2.2) stopping just before injecting the sample on the size exclusion column. From this point I tried:

- adding directly to sample desired concentration of $MgCl_2$ or KCl, incubating 15 min on 37°C, filtering the sample with 0.1 µm filter and injecting the sample on SEC-MALLS (figure 19B)
- rebuffering the sample in buffer with desired concentration of $MgCl_2$ or KCl with Amicon tubes with MWCO 100kDa, filtering the sample with 0.1 µm filter and injecting the sample on SEC-MALLS (figure 19C)
- rebuffering the sample in buffer with desired concentration of $MgCl_2$ or KCl by overnight dialysis, filtering the sample with 0.1 µm filter and injecting the sample on SEC-MALLS
- rebuffering the sample in buffer with desired concentration of $MgCl_2$ or KCl with PD-G25 or PD10 columns, filtering the sample with 0.1 µm filter and injecting the sample on SEC-MALLS (figure 19A)
- performing transcription itself with desired concentration of $MgCl_2$, and keeping the concentration of $MgCl_2$ constant by skipping the EDTA step (chelation of divalent cations).

Regarding salt concentrations I screened the following conditions: 5 mM $MgCl_2$, 10 mM $MgCl_2$, 17.5 mM $MgCl_2$, 5 mM $MgCl_2$ + 5 mM Spermidine, 150 mM KCl, 250 mM KCl, 500 mM KCl, 150 mM KCl and 1 M KCl. I also tried changing the concentration of MEG3 in injected sample (0.1-5 µM), and using different size exclusion columns (S-500 and S-1000). Unfortunately, all attempts lead to formation of aggregates that prevented the study of the compact form of MEG3 by light or x-ray scattering methods. I could only obtain suitable SEC-MALLS profiles from samples of MEG3 at 0.5 µM, using a S-500 column, and in the following buffer: KCl 0.1 M, K-MOPS pH 6.5 8 mM, Na-EDTA 0.1 mM. These samples, which do not contain divalent ions and only low concentration of monovalent ions, correspond to partially folded MEG3 samples (figure 19D).

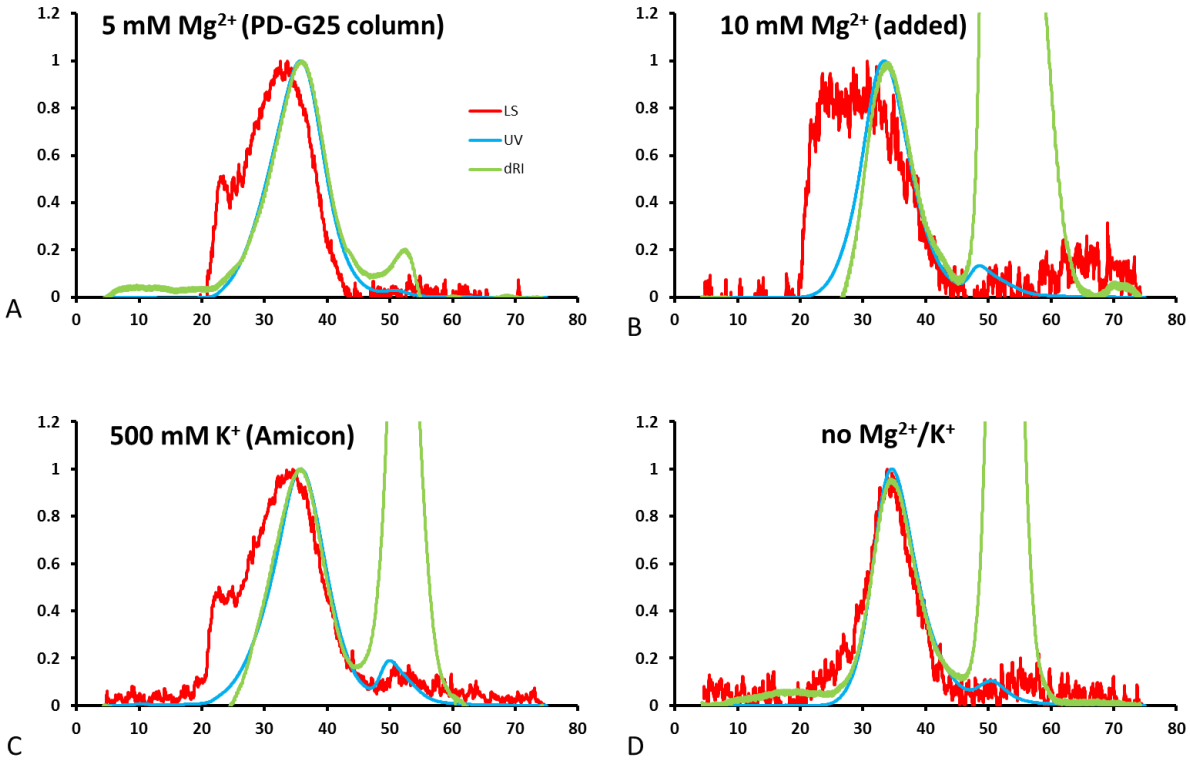


Figure 19: Optimization of folding conditions by MALLS.
SEC-MALLS profiles of MEG3v1 after *in vitro* transcription. 0.5µM MEG3v1 on S500 column (red line light scattering, blue line UV absorption at 260 nm, green line differential refractive index). (A) Sample rebuffered in buffer with 5 mM $MgCl_2$ with PD-G25 column. (B) 10 mM $MgCl_2$ added directly to sample. (C) Sample rebuffered in buffer with 500 mM KCl with Amicon tubes with MWCO 100kDa. All samples were filtered with 0.1 µm filter before injecting it on SEC-MALLS.

3.3.3. Tertiary structure by SAXS

Since SEC-MALLS results support that MEG3 can be studied by light scattering in the absence of magnesium, I performed SAXS studies on such partially folded form of MEG3. Such experiment would provide the dimensions of MEG3 particles in solution and would be very valuable reference to interpret the behaviour of MEG3 in single particle imaging by AFM or EM (see paragraphs 3.3.4. and 3.3.5.). From the linear range of Guinier plot I determined that MEG3 has a radius of gyration (R_g) of 17.78 nm and from pair distances distribution function a maximal diameter (D_{max}) of 70 nm (figure 20). Comparing the radius of gyration from SAXS studies with

the hydrodynamic radius we obtained with AUC and DLS in the same conditions (R_h DLS without magnesium = 14 nm ± 7 nm, AUC 0.05 mM Mg^{2+} = 12 nm) I can conclude that MEG3 in this conditions is not globular but oblate as expected since I was analysing partially folded and not fully folded MEG3 particle (R_g/R_h = 1.27-1.48). Also, the Kratky and the P(r) plots suggest that MEG3 has structured domains connected by flexible linkers, which is expected in these experimental conditions and on the basis of the MEG3 secondary structure maps.

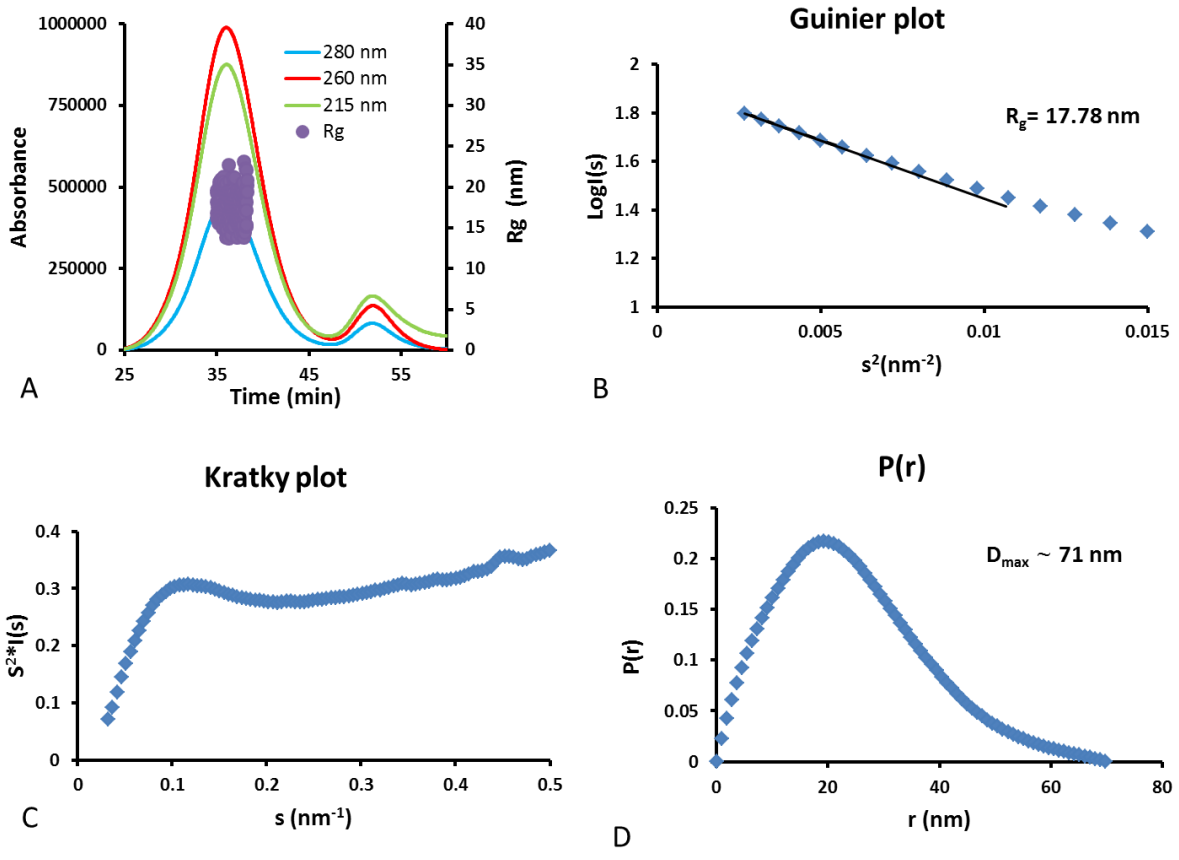


Figure 20: MEG3 folding intermediate visualized in 3D with SEC-SAXS.
(A) UV- absorption elution profile from SEC-SAXS (B) Guinier plot with calculated radius of gyration (R_g ≈ 17.78 nm) (C) Kratky plot (D) Pair distance distribution function indicates that D_{max} ≈ 71 nm

3.3.4. Tertiary structure by AFM

After confirming that MEG3 is compacting in solution I wanted to visualize it with a single molecule technique. AFM has been used for imaging structured RNA molecules (Giro et al., 2004; Hansma et al., 1996; Lyubchenko et al., 2011; Schon, 2016). By AFM in collaboration with Jean-Luc Pellequer and Jean-Marie Teulon (IBS, France), I managed to visualize MEG3 in three different conditions (denatured, partially folded and compact).

First, I needed to optimize the experimental conditions. For this optimization I used partially folded samples, for which I had obtained biophysical parameters (R_g, D_{max}) by SEC-SAXS (see paragraph 3.3.3.). Regarding the surface optimization, I tried to deposit sample on unmodified mica, mica modified with APTES (Lyubchenko et al., 2011) and mica modified with $NiCl_2$ (Lyubchenko et al., 2011). Samples on unmodified mica behaved the best (figure 21).

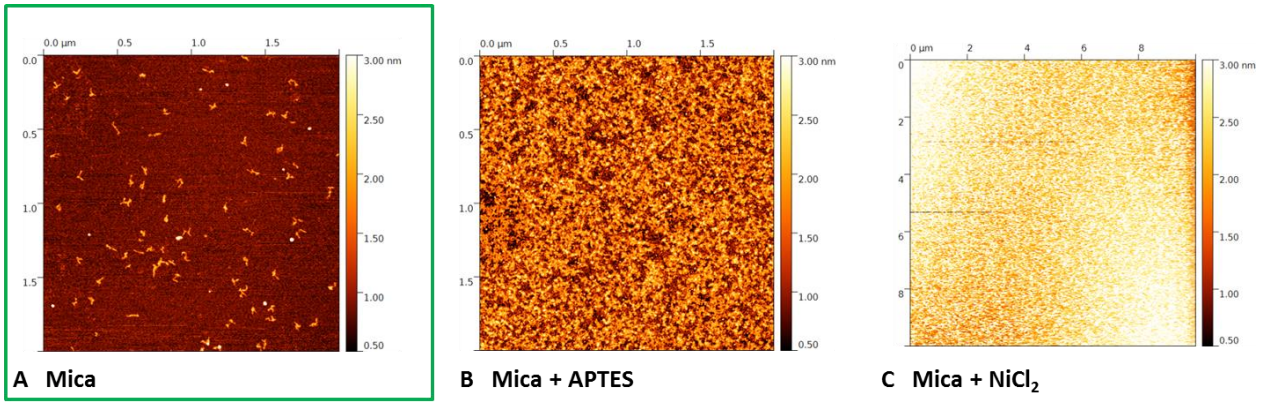


Figure 21: AFM optimization of surface.
AFM images of sample on (A) pure mica surface, (B) mica modified with APTES, (C) mica modified with Ni^{2+}. Green box indicates chosen/optimal condition for AFM imaging.

To further optimize conditions, I tried to visualize samples in liquid (Schon, 2016) and in air (Schon, 2016). Dry samples visualized in air resulted in images of better quality (figure 22). Subsequently, I had to optimize the process of drying of my samples. I tried quick drying with nitrogen gas, incubating for 1, 2, 3, 5 or 10 min before drying with nitrogen gas, and drying samples slowly over night. 3 min incubation and subsequent quick drying turned out to be best approach (figure 22D).

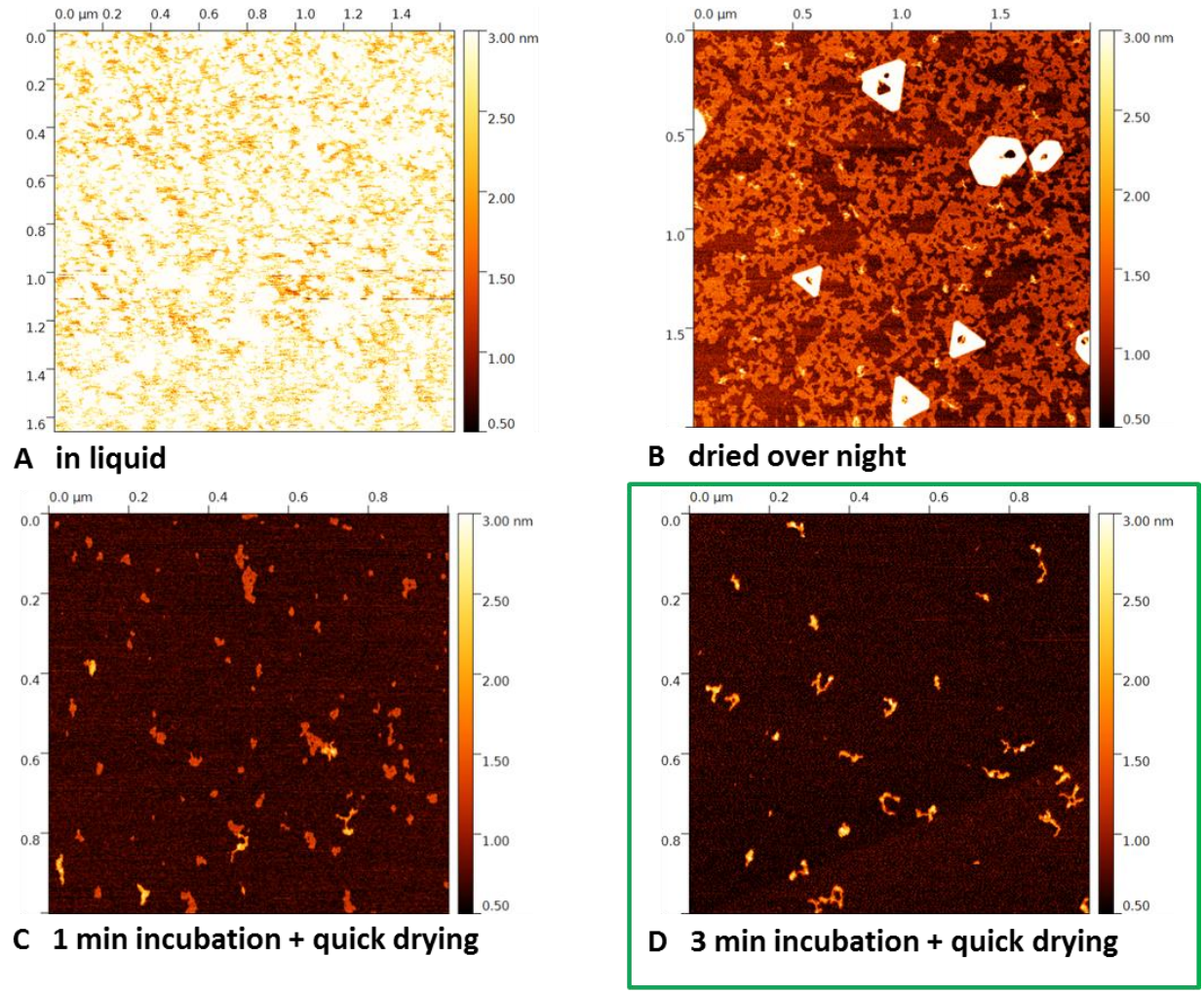


Figure 22: AFM optimization of imaging conditions.
AFM images of sample on mica (A) visualized in liquid, (B) slowly dried overnight, (C) incubated for 1 min before drying quickly with nitrogen gas, (D) incubated for 3 min before drying quickly with nitrogen gas. Green box indicates chosen/optimal condition for AFM imaging.

Finally, I needed to optimize sample preparation of compact samples and denatured samples. I know from my previous AUC studies that MEG3 compacts with addition of magnesium, but magnesium at these concentrations (> 5mM) crystallizes and interferes with imaging. To avoid the formation of crystals on the surface of the mica I tried to bind magnesium on the mica surface and then add MEG3. This approach wasn't successful; I could not visualize uniform single particles (figure 23A). Next I tried to have magnesium in desired concentration in the sample buffer and to remove it after deposition by washing with 2 ml of water following examples reported in the literature (Giro et al., 2004). This approach was successful (figure 23).

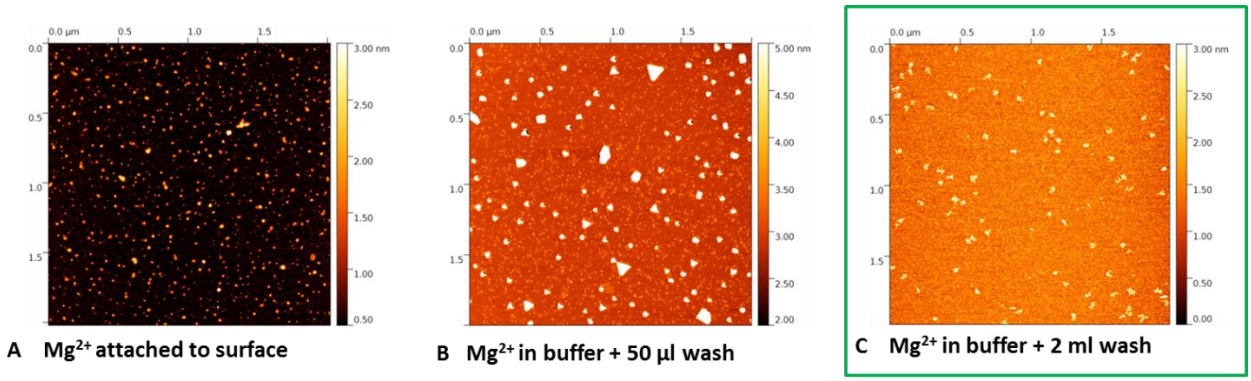


Figure 23: AFM optimization of compact sample imaging.
AFM images of sample on mica (A) Mg^{2+} attached to mica surface, (B) Mg^{2+} in buffer with sample subsequently washed with 50 µl of miliQ water, (C) Mg^{2+} in buffer with sample subsequently washed with 2 ml of miliQ water in 200 µl steps. All three samples incubated for 3 min before drying quickly with nitrogen gas. Green box indicates chosen/optimal condition for AFM imaging.

For the denatured samples I tried heating them for 5 min to 80°C or 95°C, adding 14 M urea and combination of both but this resulted in degraded sample and/or formation of urea crystals (figure 24A-B). Finally I managed to denature my samples and deposit them on the mica by precipitating MEG3 in isopropanol, resuspending it in formamide and diluting this sample with ethanol (figure 24C). Having all the experimental conditions optimized, I managed to visualize MEG3 in three different conditions (denatured, partially folded and compact) (figure 26, appendix figures 5).

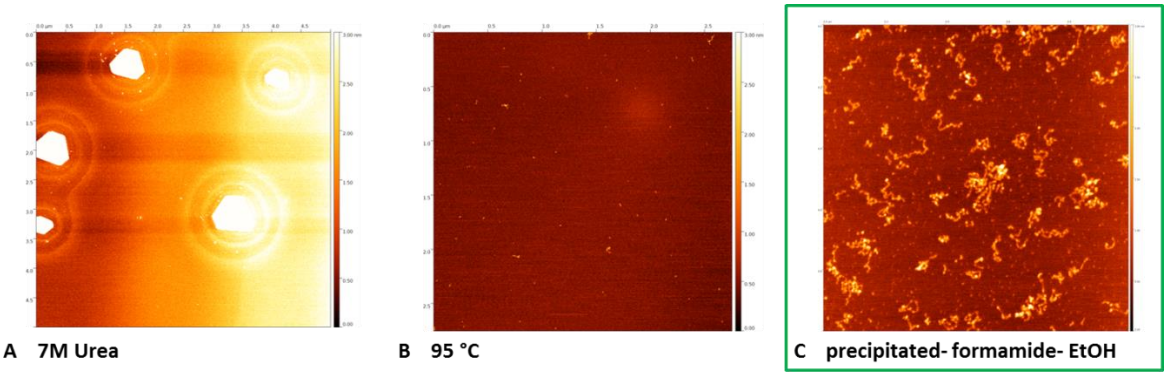


Figure 24: AFM optimization of denatured sample imaging.
AFM images of sample on mica (A) sample mixed 1:1 with 14M Urea, (B) samples heated for 5 min at 95 °C, (C) sample precipitated in isopropanol, resuspended in formamide and diluted with ethanol. Green box indicates chosen/optimal condition for AFM imaging.

I could clearly see that the denatured MEG3 molecule is unstructured because it forms elongated filaments. Next I imaged MEG3 in folding buffer that doesn't contain any magnesium. In these conditions, I could see that MEG3 is folding as the particles become higher and shorter. Interestingly, I noticed recurrent particle shapes in this condition (figure 25). I noticed several higher domains connected with linkers suggesting that MEG3 adopts a secondary structure with a multi-domain organization which is in agreement with my secondary structure map. Previously performed in solution SEC-SAXS studies on such partially folded form of MEG3 confirm that this isn't just the behaviour on the mica surface (figure 20). On AFM images in the presence of magnesium I don't see the multi-domain organization anymore but rather all domains grouped together in more globular particles that are shorter and higher than in previous conditions.

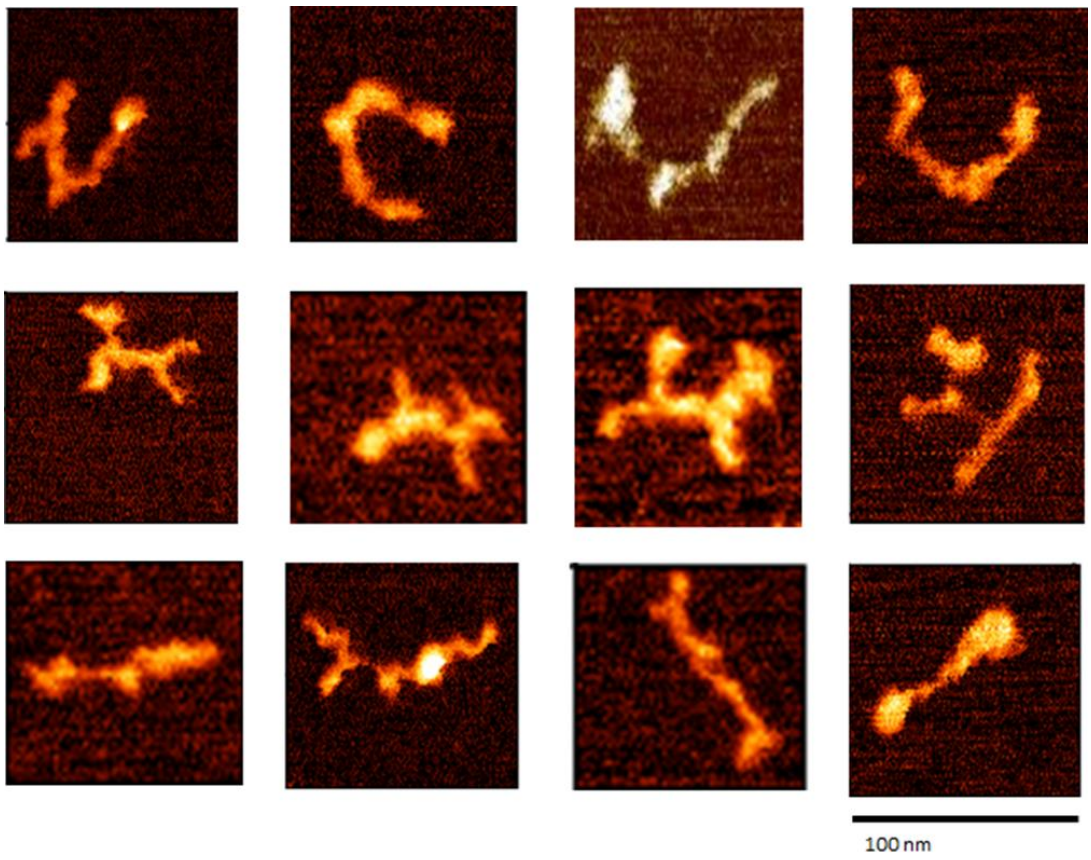


Figure 25: MEG3 folding intermediate recurrent particles visualized by AFM.
MEG3 in the absence of divalent ions. Recurrent particles with globular domains connected with linkers.

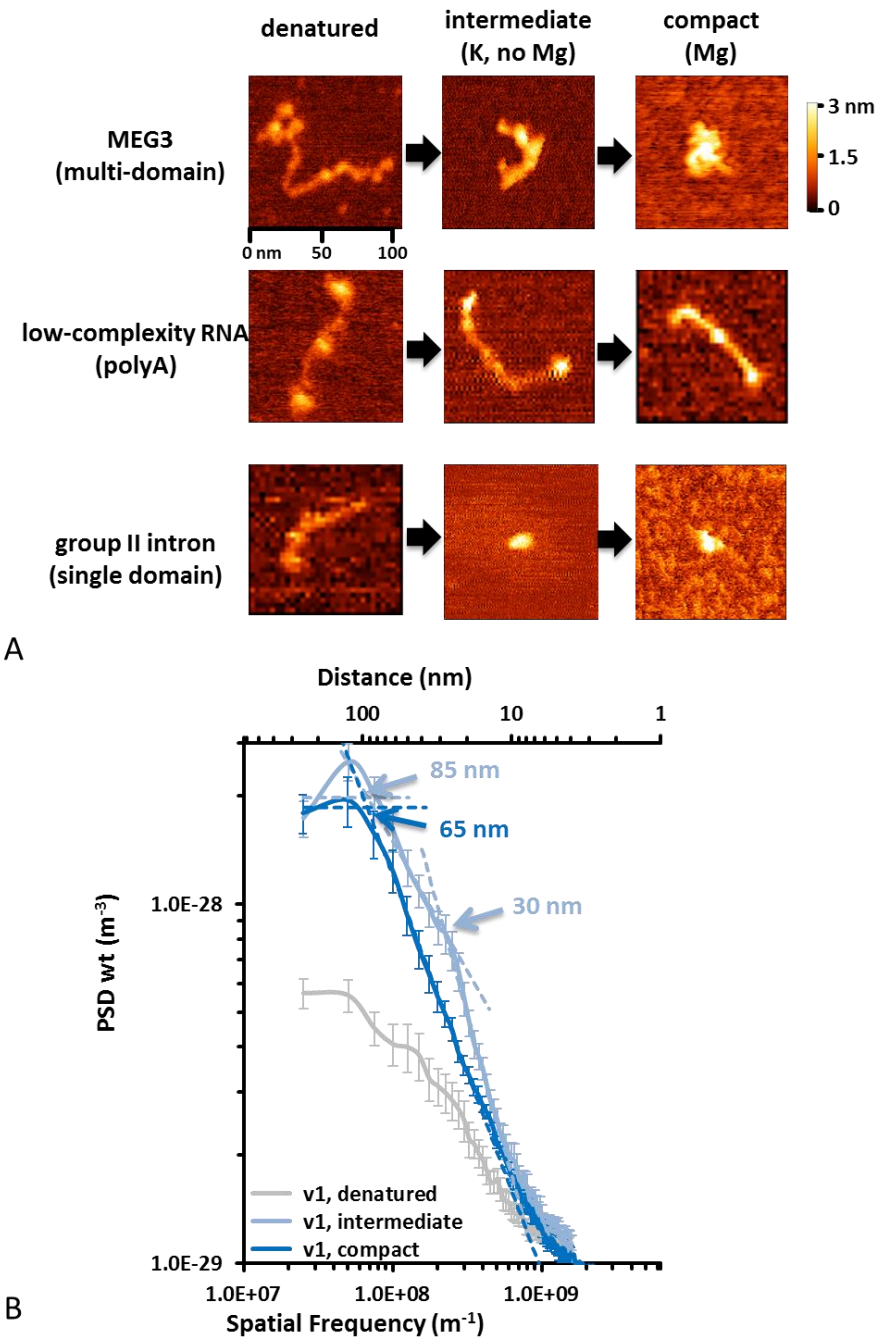


Figure 26: Gallery of particles visualized by AFM.

(A) Top row MEG3 particles in 3 different forms from left to right: denatured, folding intermediate and compact. Middle row RNA homopolymer poly A in same conditions. Bottom raw group II intron in same conditions. (B) Power spectral density analysis of MEG3 particles visualized by AFM. Grey line denatured particles. Light blue line folding intermediate particles. Dark blue line compact particles. Intercepts between linear fits to auto affine regions in the spectra (dashed lines) indicate characteristic spatial frequencies. Error bars indicate standard error mean among the 100 particles analyzed for each condition (out of 110-150 particles imaged in total for each condition).

To statistically represent my data I picked more than 100 particles of MEG3 for each state (denatured, folding intermediate and compact) and in collaboration with Paolo Annibale (MDC, Berlin) did a power spectral density analysis (figure 26B). From that analysis we can see that for denatured form of MEG3 the line doesn't peak at any specific distance indicating that the particles are random length, for folding intermediate we can distinguish two peaks one at 30 nm corresponding to a size of single folded domain and a second one at 85 nm corresponding to a full length molecule, and finally for a compact form we see a single peak at 60 nm corresponding to a size of single fully compacted molecule on top of AFM surface.

To further confirm that the effect of MEG3 is specific and not an unspecific folding behaviour of RNA molecules induced by mica, as a negative control I used poly (A) RNA homoploymer which has been previously characterized by AFM (Hansma et al., 1996). I imaged poly (A) in same 3 conditions as I did MEG3 and in those entire conditions poly (A) remains unstructured as expected (appendix figure 6). As a positive control I used *Oceanobacillus iheyensis* group II intron, for which crystal structures are available (Marcia and Pyle, 2012) and I have access to it in the Marcia lab. As expected *Oceanobacillus iheyensis* group II intron formed one structured domain (appendix figure 7), which was approximately 4 times smaller than MEG3 which is in line with the size of the two (MEG3 is 1595 nt long, *Oceanobacillus iheyensis* group II intron is 394 nt long).

3.3.5. Structural studies with EM

Being successful in imaging single molecule MEG3 with AFM encouraged me to try doing the same but with higher resolution with electron microscope. In collaboration with Manikandan Karuppasamy (research scientist in the Marcia group) I tried imaging particles of MEG3 full length and just the structured core (D2-3) in buffer with and without magnesium with negative stain electron microscopy and cryo electron microscopy. MEG3v1 particles in buffer without magnesium observed by negative stain electron microscopy were flexible as expected, we could observe single particles but they adopted a heterogeneous form (18 – 48 nm) which is in line with what I observe on AFM images in the same condition. To obtain more homogenous sample

we tried imaged MEG3v1 samples in buffer with varying concentrations of magnesium (10-25 mM Mg^{2+}) by negative stain electron microscopy but we could still observe heterogenous particles of different sizes (figure 27). In addition, a lot of small to big size aggregates were seen for all conditions.

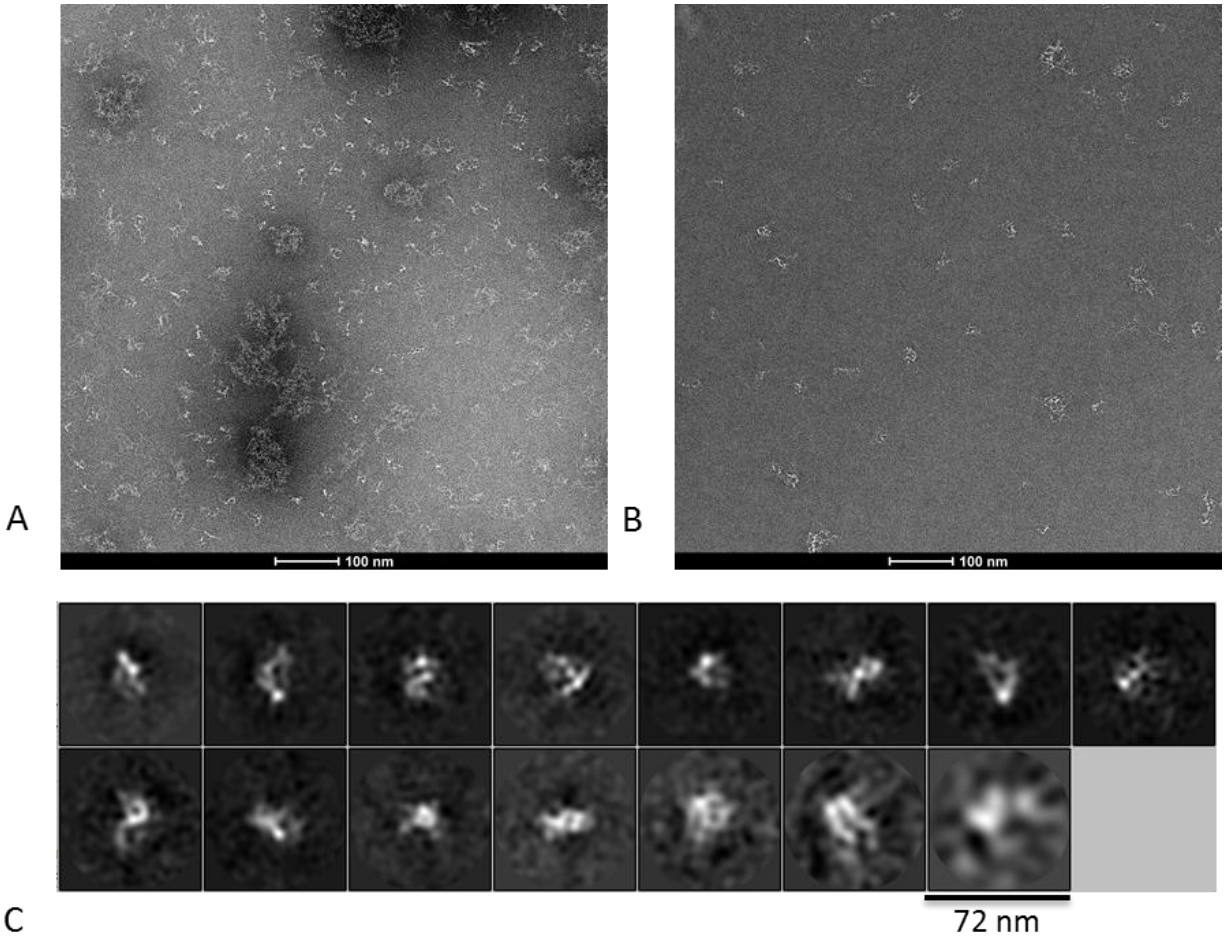


Figure 27: MEG3v1 visualized by negative stain electron microscopy.
(A-B) Negative stain images of 0.05 µM MEG3v1 in the presence of 17.5 mM Mg^{2+} and 0.06 µM MEG3v1 25 mM Mg^{2+}, respectively. (C) 2D classes from automatic picking of particles on negative stain images of MEG3v1 with 25 mM Mg^{2+}.

Subsequently, we tried imaging MEG3v1 samples in buffer with varying concentrations of magnesium (10-25 mM Mg^{2+}) by cryo electron microscopy. In this condition we couldn't observe any single-particles, only aggregates that preferentially form on carbon support and don't fall in the holes in the grid (figure 28).

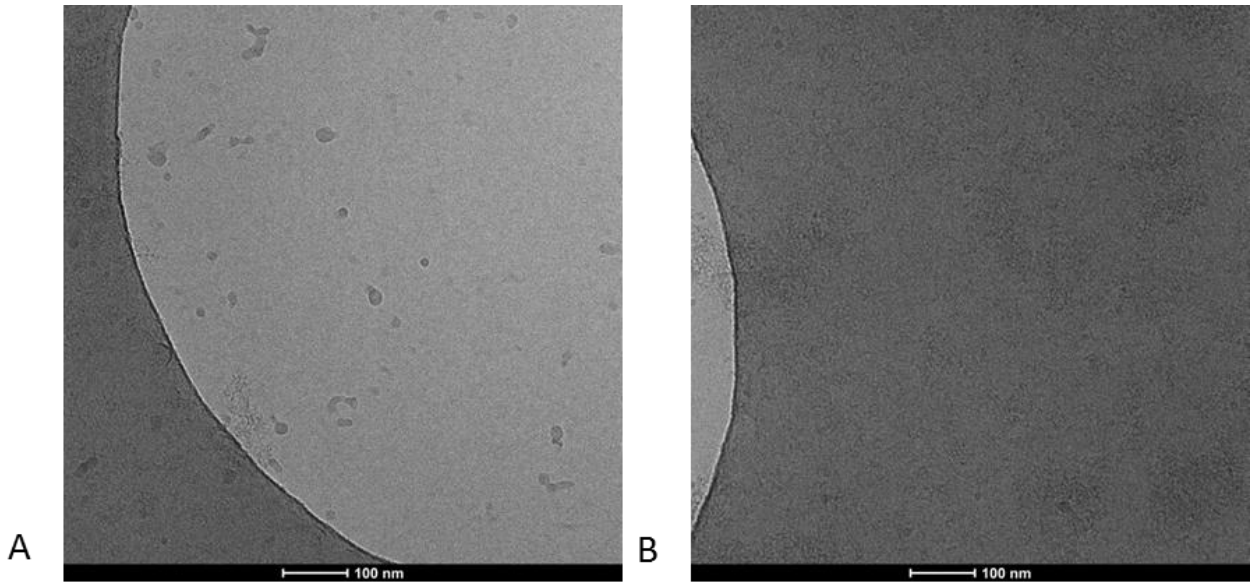


Figure 28: MEG3v1 visualized by cryo electron microscopy.
Cryo-EM images of MEG3v1 with 25 mM Mg^{2+} (A) 0.67 µM sample with zoom in to an empty hole; (B) 0.3 µM sample with zoom in to carbon support with aggregates.

Structured core of MEG3 (D2-3) behaved similar as MEG3v1. By negative stain electron microscopy we observed heterogeneous particles and aggregates (figure 29) and by cryo electron microscopy we observed heterogeneous single-particles only for samples in buffer containing no magnesium and with magnesium we observed different sizes of aggregates (figure 30). To be able to see single folded particles of MEG3 with electron microscopy would therefore require further optimization of sample and conditions. This result is not unexpected because currently there is no EM structure of a lncRNA, nor of any other purified large RNA without proteins to stabilise it on the grid.

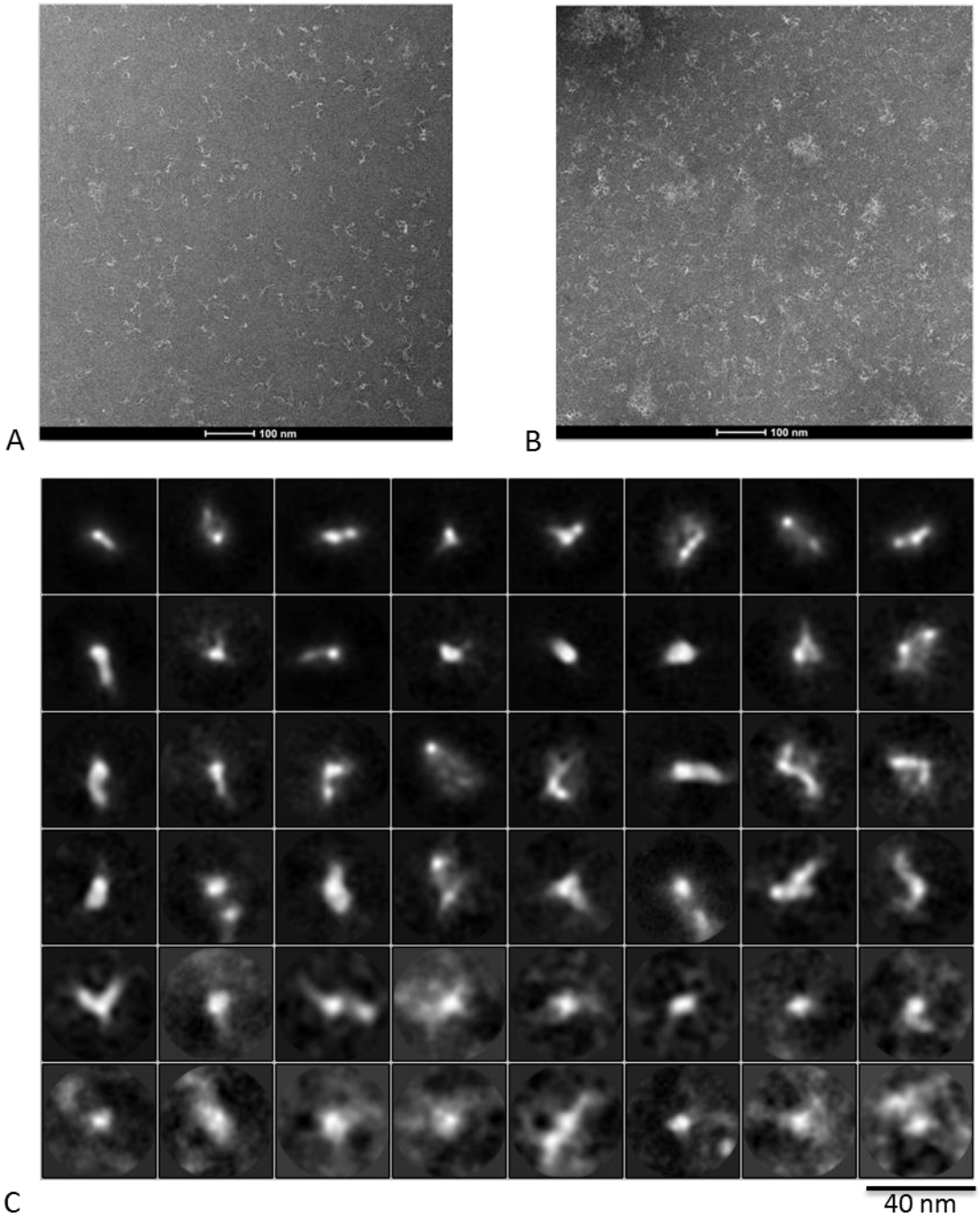


Figure 29: MEG3 D2-3 visualized by negative stain electron microscopy.

Negative stain images of 0.3 µM MEG3 D2-3 (A) without magnesium and (B) in the presence of 25 mM Mg^{2+}. (C) 2D classes from automatic picking of particles on negative stain images of MEG3 D2-3 without magnesium.

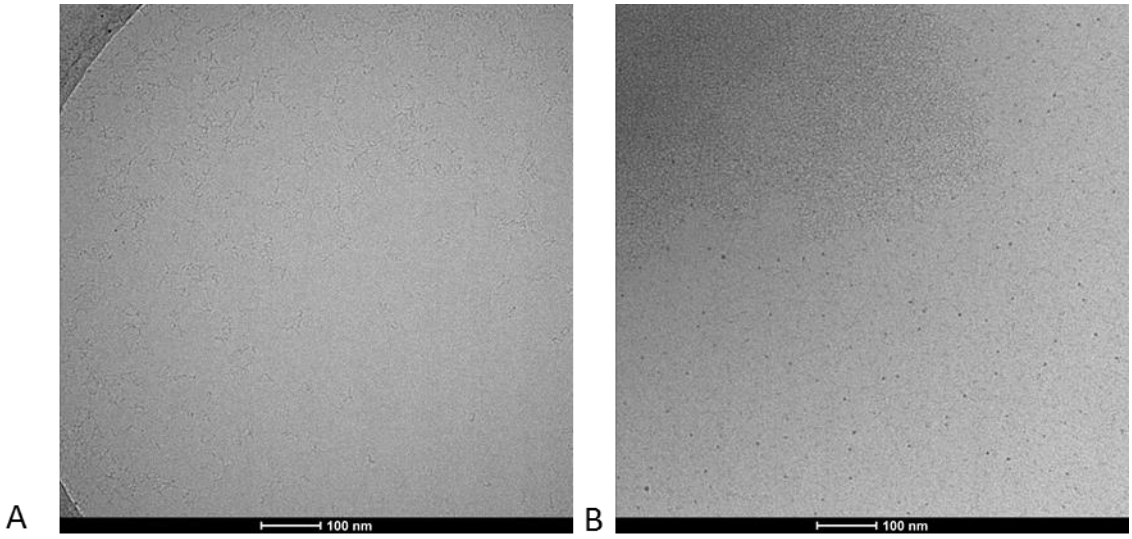


Figure 30: MEG3 D2-3 visualized by cryo electron microscopy.
Cryo-EM images of MEG D2-3 (A) 4 µM sample in buffer without Mg^{2+}; (B) 3 µM sample in buffer with 17.5 mM Mg^{2+}.

Table 3: The hydrodynamic parameters of MEG3v1 from different experiments.

parameter	technique	MEG3v1 state	value
R_g	SEC-SAXS	intermediate (K^+ only)	17.78 nm
R_h	DLS	intermediate (K^+ only)	14 nm ± 7 nm
R_h	AUC	intermediate (K^+ only)	12 nm ± 0.1 nm
R_g / R_h	SEC-SAXS/ DLS, AUC	intermediate (K^+ only)	1.27-1.48
D_{max}	SEC-SAXS	intermediate (K^+ only)	71 nm
D_{av}	AFM	intermediate (K^+ only)	~ 85 nm
R_h	AUC	compact (10 mM Mg^{2+})	10 nm ± 0.1 nm
D_{av}	AFM	compact (10 mM Mg^{2+})	~ 65 nm
R_h	AUC	compact (50 mM Mg^{2+})	8.8 nm ± 0.3 nm
$C_{Mg1/2}$	AUC	Mg^{2+} titration (0-50 mM)	6.9 ± 0.35 mM
n_{Hill}	AUC	Mg^{2+} titration (0-50 mM)	3.9 ± 0.6

3.4. Optimisation of functional assays

In collaboration with Isabel Chillon (postdoctoral fellow in the Marcia group) I found by qRT-PCR two tumor cell lines (HCT-116 and Hep-G2) that express MEG3 in very low level (appendix figure 4), at the limit of detection, but still have preserved p53 pathway which makes them suitable for my studies. I could transfect those cell lines with any desired mutant and look at the effects that they cause without significant interference of endogenously expressed MEG3.

First I tried to use MTT colorimetric assay for assessing cell metabolic activity but this kind of quick experiment wasn't sensitive enough to see the fine differences in activity that mutants would cause and results were too variable for the same condition (figure 31). I tried optimizing it by changing number of cells per well, amount of DNA transfected, time of incubation after transfection and excluding washing steps but none of these conditions gave better results.

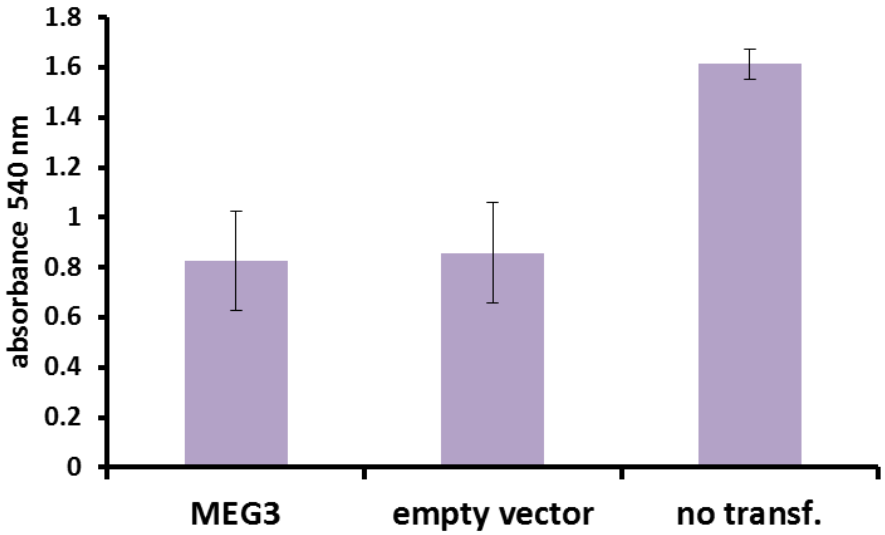


Figure 31: MTT assay.
Average absorbance at 540 nm from 8 samples for each condition. From left to right: transfected MEG3, transfected empty vector and cells that weren't transfected.

As a next step I decided to try functional assays that have been previously used to study MEG3. I referred to the functional test reported by Klibanski group (Zhang et al., 2010a) in which the authors measure the ability of MEG3 to suppress cell proliferation with a BrdU incorporation assay. I modified such test to make it quicker and less prone to subjective interpretations by

using flow cytometer instead of confocal microscopy [flow cytometry has been used before to study MEG3 as well (Lu et al., 2013)]. Additionally, I decided to use a kit that contains EdU (5-ethynyl-2'-deoxyuridine) as an alternative to BrdU (5-bromo-2'-deoxyuridine) which, because of its small size and modified method of detection by "click" chemistry, does not require sample fixation or DNA denaturation (Salic and Mitchison, 2008). Flow cytometry assays have been performed by Isabel Chillon, so the results are not reported in details here. However the experimental protocol is described in the materials and methods section (paragraph 2.15.) for completeness and general conclusions from such experiments are mentioned in the following description of the MEG3 function, wherever relevant to support my conclusions. Most importantly, by flow cytometry we detected that MEG3v1 induces cell cycle arrest in G1 phase and does not induce of apoptosis in HCT116 cells (figure 32).

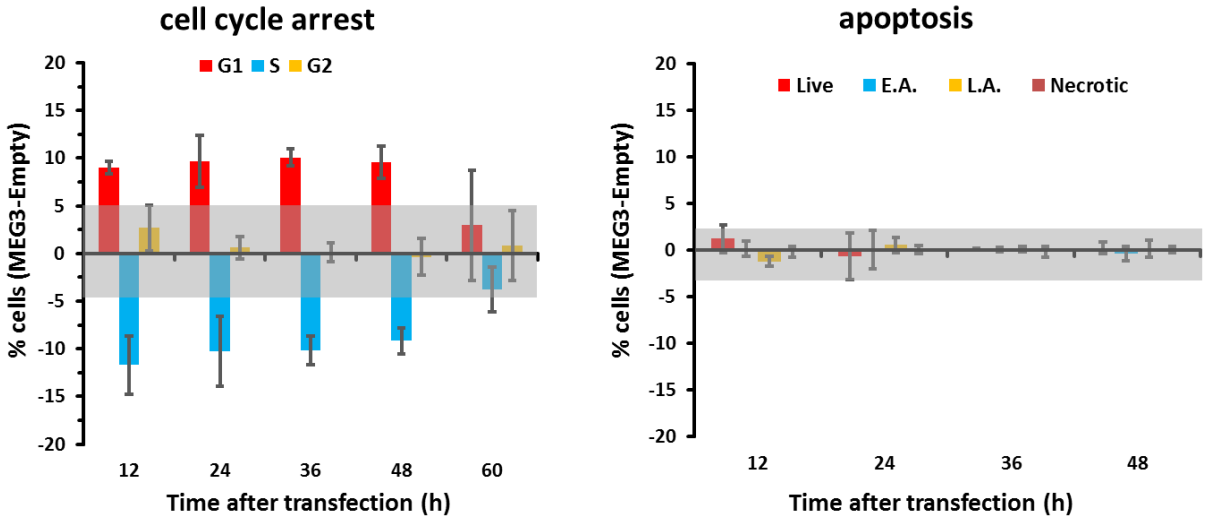


Figure 32: Cell cycle arrest and apoptosis by flow cytometry.
Cell cycle and apoptosis analysis of MEG3v1 in HCT116-p53$^{+/+}$ (E.A., early apoptotic; L.A., late apoptotic). Only GFP^{+} cells depicted. Data were normalized to the signal of corresponding empty vectors. Error bars in all panels indicate standard error of the mean of n = 3 experiments. Experiment performed by Isabel Chillon.

In addition to the flow cytometry assay which detects global cellular effects of MEG3, I also implemented a luciferase reporter assay (Zhang et al., 2010a), which detects MEG3 effects on individual p53 target genes. The luciferase assay is also higher throughput than the flow cytometry assay, allowing for faster screening of MEG3 structural mutants. I adopted the

method from Klibanski lab (Zhang et al., 2010a) with minor adaptations (as described in materials and methods paragraph 2.14.). In the process of optimization (in collaboration with Isabel), we noticed that MEG3 splicing variants stimulate p53-mediated transactivation at different level [as it has been reported before (Zhang et al., 2010a)]. I also showed that the MEG3 effect is exclusively p53-dependent, because no Firefly luciferase production can't be observed in the isogenic HCT116-$p53^{-/-}$ cells (figure 33A). Moreover, we observed that the MEG3 activity depends on the p53 response element located upstream of the luciferase gene in our reporter vector, suggesting a mechanism by which MEG3 may induce the selectivity in the p53 stress response observed previously (Zhu et al., 2015). I tested the activity of the 3 different MEG3 splicing variants (v1, v3 and v9) with 4 vectors containing 4 different p53 response elements (figure 33B). I used two vectors that have been engineered to contain p53 response elements that are activated by p53 but don't exist in this form in living organisms (p53luc and pG-13luc), and two vectors provided by Prof. Alberto Inga (CIBIO, Italy), that have p53-RE of endogenous genes MDM2 and p21 (Menendez et al., 2010). To be more precise I found out that MEG3v9 is strongest, MEG3v1 mild and MEG3v3 weakest activator of p53 when p53luc and PG13luc vectors were used, as it has been reported before for p53luc (Zhang et al., 2010a). Overall all splice variants were more effective on p53luc, whose promoter likely acts as a stronger activator of p53, because it is also more activated by $p14^{ARF}$. But a different behaviour was observed with reporters that contain p53-RE of endogenous genes. The reporter that contains the p53-RE of the MDM2 gene was strongly activated by MEG3v1, MEG3v9 but even stronger by MEG3v3 (which is the weakest activator of p53luc and PG13luc). The reporter that contains the p53-RE of the p21 gene was not activated by MEG3v1, MEG3v9 but it was activated by MEG3v3, although at low levels.

To quantify the expression levels of the transfected MEG3 variants and mutants, I performed qRT-PCR (appendix figure 9).

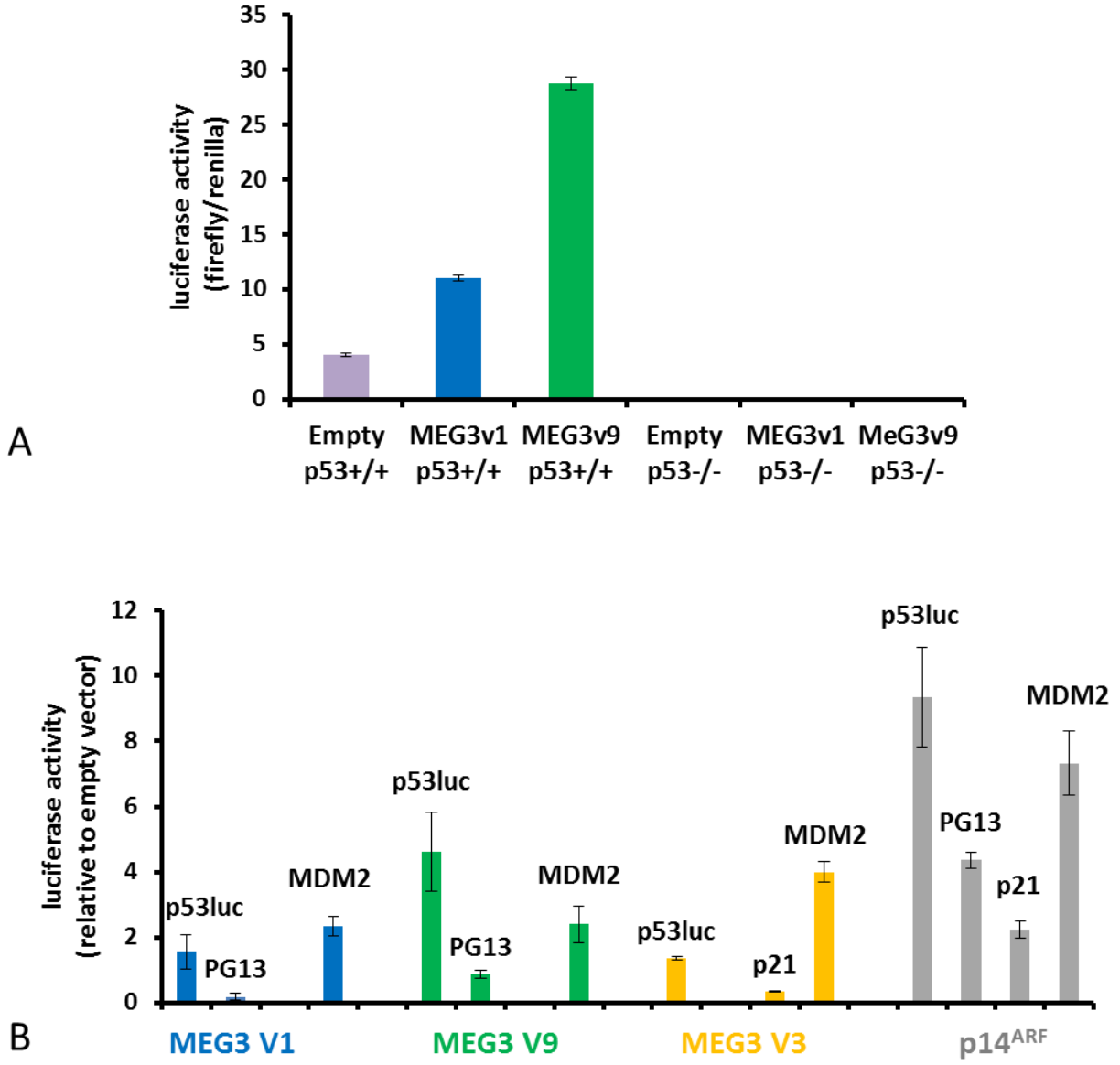


Figure 33: Selective activation of p53 by MEG3 variants.

(A) Luciferase assay performed in HCT116-p53$^{-/-}$ cells (absolute ratio of Firefly luciferase vs Renilla luciferase chemiluminescence). (B) p53 activation by v1, v3, v9, and p14^{ARF} on 4 reporter vectors possessing different p53REs (pG13Luc, p53Luc, pGL-p21, and pGL-MDM2). Data were normalized to the signal of corresponding empty vectors (set to a y-axis value of 0). For this experiment, 500 ng of MEG3 vectors and 50 ng of p14^{ARF} vectors were used for transfection. Error bars in all panels indicate standard deviation of n = 3 experiments.

3.5. Structure-function relationships

To evaluate correlation between structural domains and the functional role of MEG3 as a tumor suppressor I designed and performed series of luciferase assays that measure p53-mediated transactivation by different MEG3 constructs.

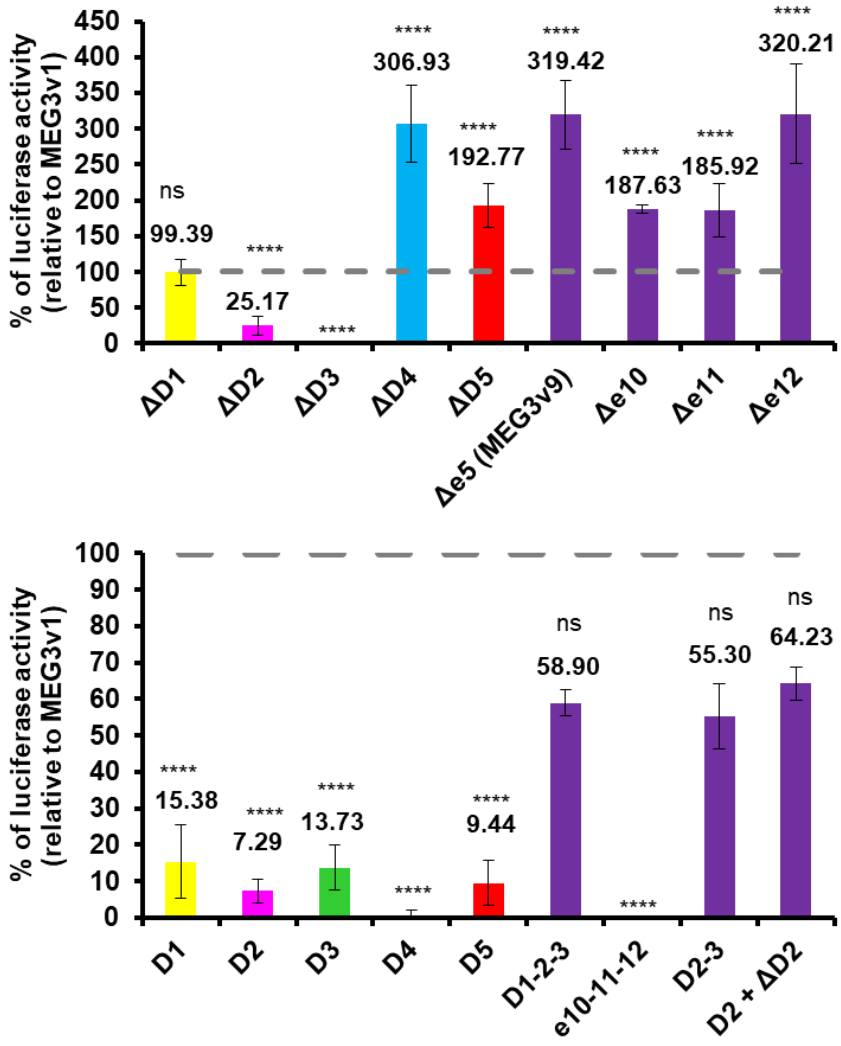


Figure 34: Stimulation of p53-mediated transactivation by luciferase reporter assay (MEG3 domains and exons).

Histograms: bars and data labels represent percentage of p53 activation relative to MEG3v1 wt. Grey line (100 %) indicate the level of activity of MEG3v1 wt. Error bars in all panels indicate standard deviation of n = 3 experiments. Asterisks indicate significant difference in relative luciferase signal with respect to v1 based on one-way ANOVA statistical tests in GraphPad (**** indicates P ≤ 0.0001, ns- non-significant). D-domain, e- exon, Δ- deleted.

3.5.1. Luciferase assay probing

First, I tested the ability of individual domains to stimulate p53-mediated transactivation (figure 34). Individual domains were not able to stimulate p53-mediated transactivation (-1.63-15.38 % of activity relative to MEG3v1). Minimal construct able to do so but not in full capacity (55.30 % of activity relative to MEG3v1) is a domain 2 and domain 3 together, which corresponds to exon 3 that is common to all splicing variants. Not only that domain 2 and 3 are able to stimulate p53 but they are also essential since other part of MEG3 that is common to all splicing variants, namely exons 10, 11 and 12 together, can't stimulate p53 activity at all (-7.81 % of activity relative to MEG3v1) and deleting individually domains 2 and 3 abolishes the activity (respectively 25.17 % and -3.23 % of activity relative to MEG3v1). On the contrary deleting domains 4 and 5 individually or exons 10, 11, 12 stimulates p53-mediated transactivation above the level of stimulation by splicing variant 1 (192.77-319.42 % of activity relative to MEG3v1). Interestingly, D2, which is inactive in isolation, is capable of inducing a partial p53 response when co-transfected with the ΔD2 (64.23 % relative to MEG3v1 wild-type), which are also both inactive *per se*. Therefore, I conclude that both D2 and D3 are needed to induce p53 activation and are capable of acting in *trans* with respect to one another. D2-D3 (E3) thus constitutes the minimal functional unit of MEG3.

Since domain 2 and 3 are essential for MEG3 activity and secondary structure of those domains is well-preserved among mammals I decided to test in greater details the importance of those domains to stimulate p53-mediated transactivation (figure 35). To do so I designed a range of mutants that may disrupt the activity. Within domain 3 deleting helixes 16, 17, 21, 22, 23, 28, 29 doesn't decrease activity significantly (56.64-137.45 % of activity relative to MEG3v1) whereas deleting helices 25-26-27-28-29 almost completely abolishes the activity (18.04 % of activity relative to MEG3v1).

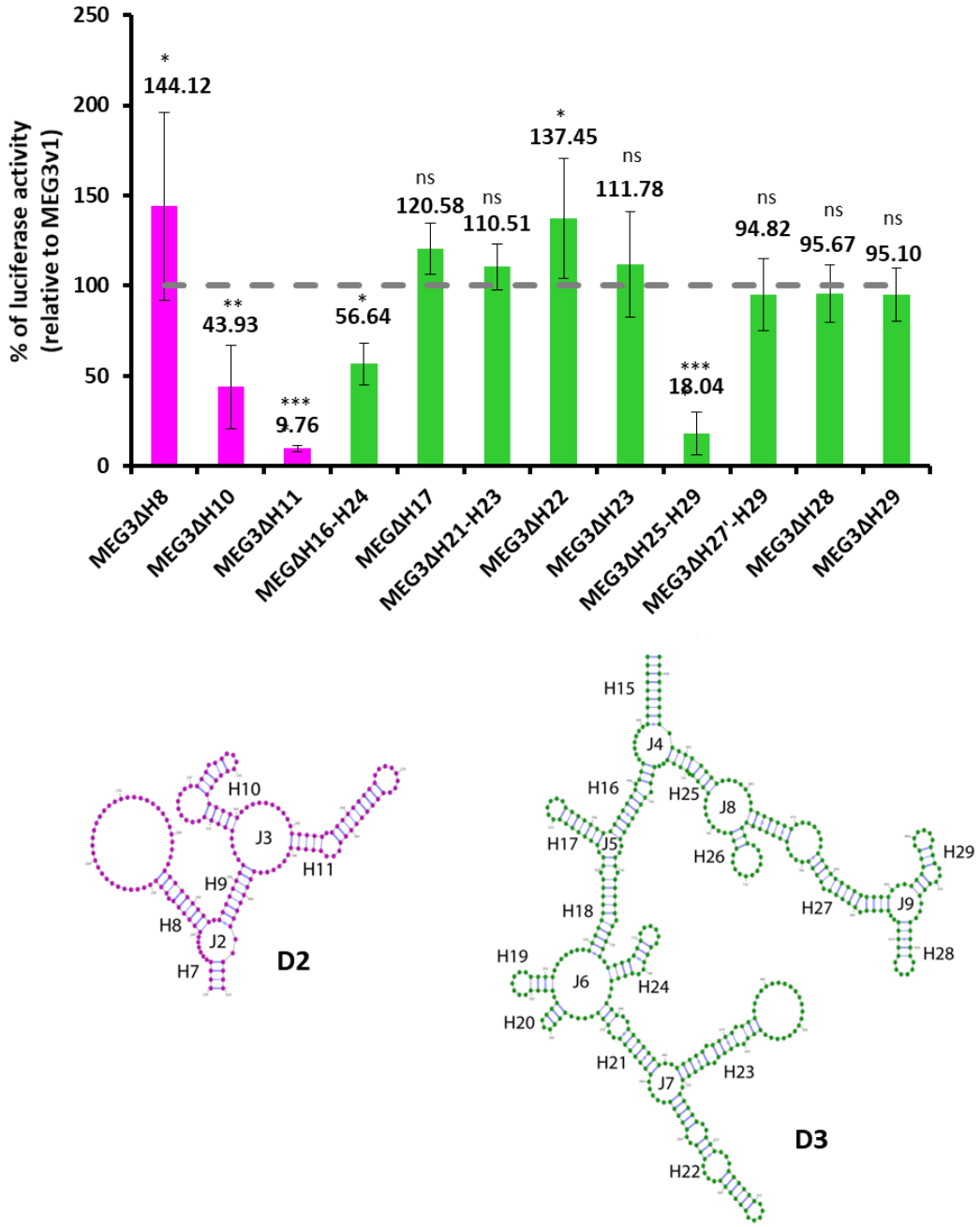


Figure 35: Stimulation of p53-mediated transactivation by luciferase reporter assay (MEG3 D2-3). Histogram: bars and data labels represent percentage of p53 activation relative to MEG3v1 wt. Grey line (100 %) indicate the level of activity of MEG3v1 wt. Error bars in all panels indicate standard deviation of n = 3 experiments. Asterisks indicate significant difference in relative luciferase signal with respect to v1 based on one-way ANOVA statistical tests in GraphPad (ns- non-significant, * indicates P ≤ 0.05, ** P ≤ 0.01 and *** P ≤ 0.001). ΔH27'-H29 mutant- only upper part of H27 deleted (Δ792-862). D-domain, H-helix, Δ- deleted. MEG3v1 secondary structure map of domains 2 and 3 (colored pink and green respectively). D-domain, H- helix, J-junction.

Within domain 2 deleting helix shows 144.12 % of activity while deleting helix 11 decreases activity to only 9.76 %. Finding such small and specific motif important for function I wanted to look into it a bit further (figure 36). Although, deleting the ACG elbow within helix 11 preserves full capacity (130.39 % of activity relative to MEG3v1), deleting the loop on top of helix 11, mutating the loop on top of helix 11 to poly A or breaking the helix 11 by mutating one side abolishes the activity (9.21, -1.08, -0.85 % of activity relative to MEG3v1). By mutating the other side of the helix 11 to restore secondary structure we regain 84.98 % of activity relative to MEG3v1 showing that the activity is structure dependent.

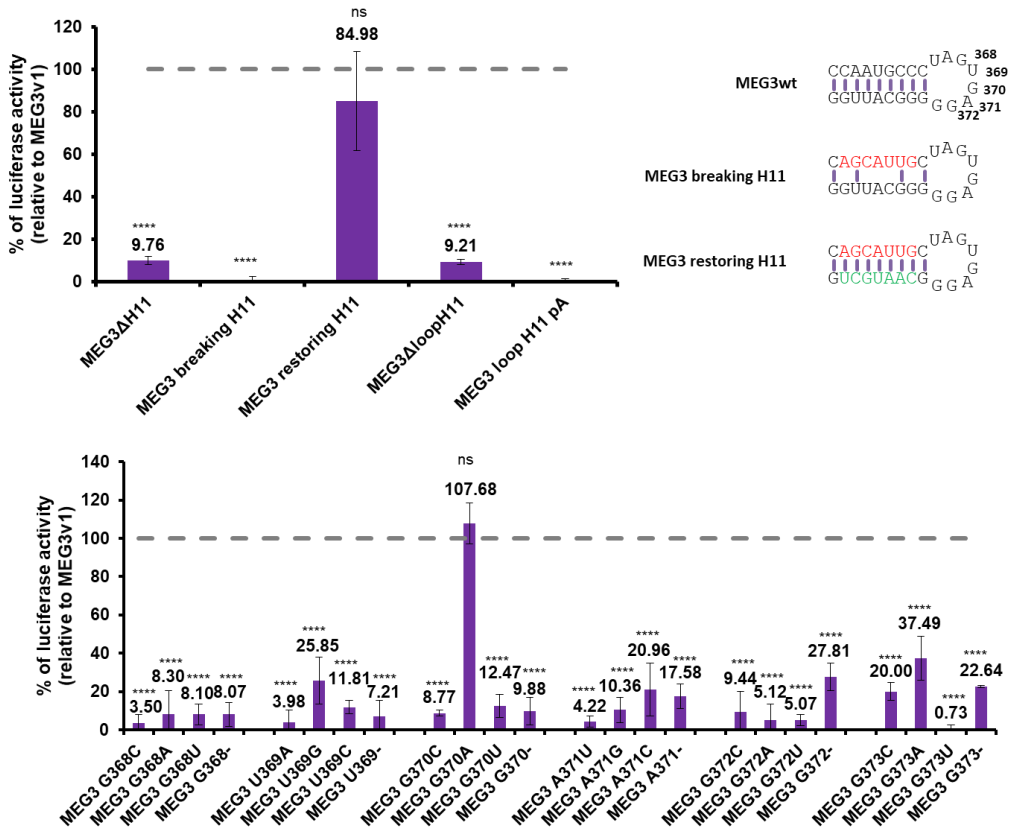


Figure 36: Stimulation of p53-mediated transactivation by luciferase reporter assay (MEG3 H11).

Histograms: bars and data labels represent percentage of p53 activation relative to MEG3v1 wt. Grey line (100 %) indicate the level of activity of MEG3v1 wt. Error bars in all panels indicate standard deviation of n = 3 experiments. Asterisks indicate significant difference in relative luciferase signal with respect to v1 based on one-way ANOVA statistical tests in GraphPad (**** indicates P ≤ 0.0001, ns- non-significant). H-helix, Δ- deleted, pA- poly A. Upper right panel: sequences used to disrupt the H11 stem (H11-5'mut, red nt) and corresponding compensatory mutations (H11-comp, green nt).

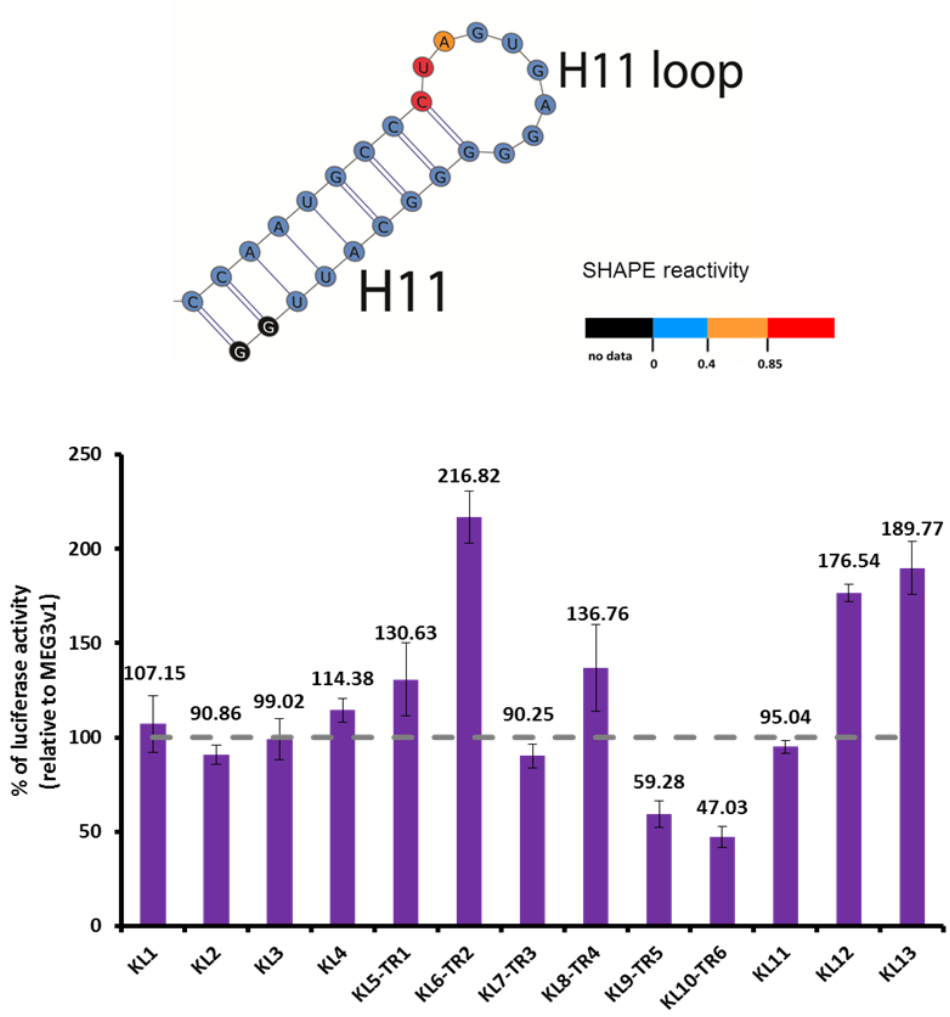


Figure 37: Stimulation of p53-mediated transactivation by luciferase reporter assay (potential kissing loops).

Upper panel: zoom in to MEG3v1 helix 11 colored based on 1M7 (SHAPE) reactivity (legend on the right). Histogram: bars and data labels represent percentage of p53 activation relative to MEG3v1 wt. Grey line (100 %) indicate the level of activity of MEG3v1 wt. Error bars in all panels indicate standard deviation of n = 3 experiments. KL- potential kissing loop, TR- tandem repeat.

To assess the sequence identity importance of the loop on top of helix 11 I tested the range of point mutants (figure 36). From my covariation analysis I know that in more than 97 % of mammals this sequence is completely preserved. Indeed mutating positions 368-372 one by one almost completely abolished the activity, except mutating G at position 370 to A which can activate p53 in full capacity (107.68 % of activity relative to MEG3v1). Since I know from my

chemical probing that nucleotides 368-373 are unreactive (appendix figure 1, figure 37 upper panel) indicating they are not very flexible I tested if there was any potential kissing loop within the MEG3 molecule by mutating all complementary sequences to poly A but none of them had such effect on the activity as helix 11 mutants (figure 37).

I found that there are 6 consecutive motifs (here named tandem repeats) in helix 27 (D3) that are complementary to the H11 terminal loop (figure 38). Since from my previous experiment I know that deleting region H25-H29 almost completely abolishes the activity I decided to look into this tandem repeats. In humans, tandem repeat 1 (TR1) corresponds to nt 857-862, TR2 to nt 862-866, TR3 to nt 866-871, TR4 to nt 871-876, TR5 to nt 875-880, and TR6 to nt 880-884 (figure 38). However, all MEG3 sequences that we identified in mammals possess at least 3 of the 6 TRs that characterize human MEG3 and covariation analyses reveal that the interaction of these TRs with the H11 terminal loop is conserved in evolution (figure 38 and appendix table 4). To test the possibility that one or more of these TRs may pair with loop on top of helix 11 (GUGAG), I made a series of compensatory mutants. I used the MEG3v1-$G^{370}C$ mutant, which is inactive (8.77 % residual activity compared to wild type), and introduced a second point mutation at positions of each TR in H27 that would base-pair with nucleotide 370 in case of kissing loop formation. This approach resulted in the creation of 6 double mutants. Double mutants involving TR2 ($U^{864}G$) and TR6 ($U^{882}G$) did not rescue activity (figure 38). However, interestingly, all other double mutants did compensate activity partially ($U^{860}G$ in TR1 recovered 25 % activity, $U^{869}G$ in TR3 34 %, $U^{873}G$ in TR4 22 %, $U^{878}G$ in TR5 12 % relative to MEG3v1-G370C mutant, figure 38). As well I tested the ability of double mutants to recover the activity using a vector with different p53 response element, the pGL-MDM2. All double mutants did compensate activity partially ($U^{860}G$ in TR1 recovered 5.71 % activity, $U^{864}G$ in TR2 5.64 %, $U^{869}G$ in TR3 17.37 %, $U^{873}G$ in TR4 2.91 %, $U^{878}G$ in TR5 5 %, $U^{882}G$ in TR6 12.91 % relative to MEG3v1-$G^{370}C$ mutant). Interestingly the TR6 ($U^{882}G$) which did not rescue activity with p53-luc and TR3 ($U^{869}G$) which recovered the most of the activity with p53-luc recovered the most of the activity with pGL-MDM2. I also made a mutant that had all of the mutations (MEG3v1-$G^{370}C$ $U^{860,864,869,873,878,882}G$) but it didn't recover the activity possibly because such cumulative mutations destroy the secondary/tertiary structure of MEG3. In paragraph 4.6 future

perspectives I described the current and future efforts being done for further understanding of the interaction H11-H27 and its functional importance.

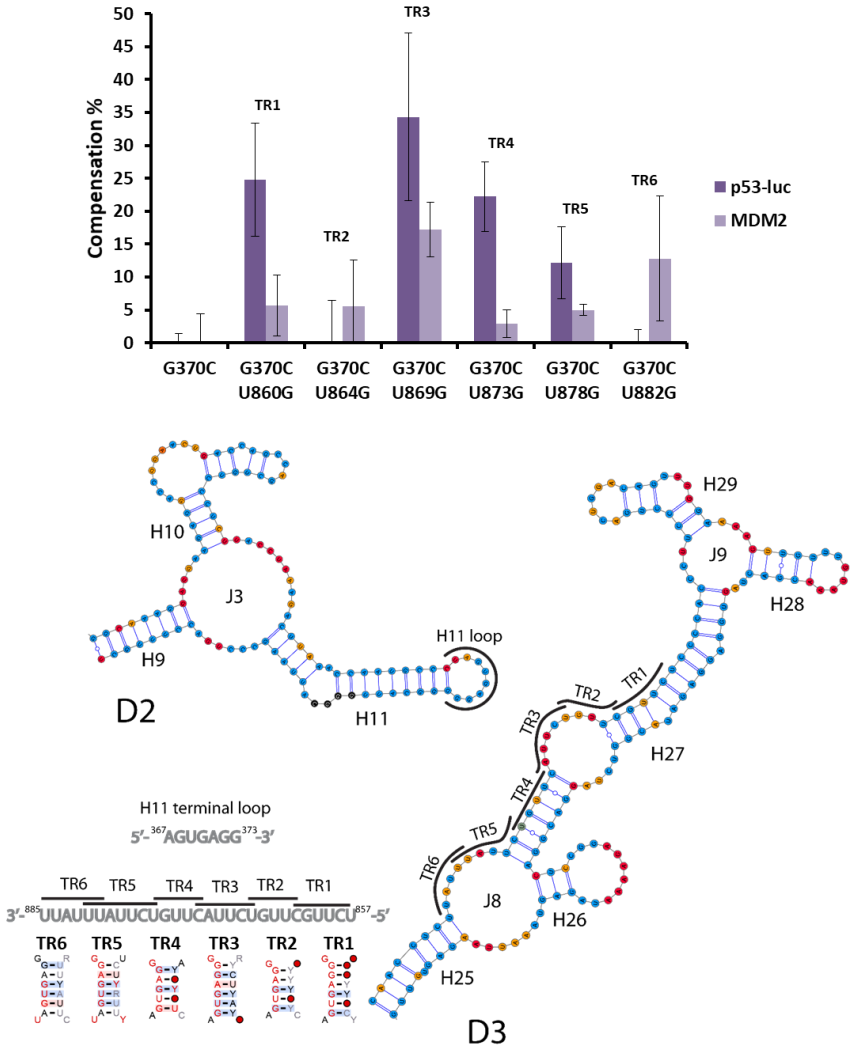


Figure 38: MEG3 H11-H27 interaction.

Histogram: bars represent percentage of p53 activation relative to G370C point mutant (0 %). G370C point mutant and compensatory double mutants. Dark purple bars p53Luc reporter, light purple pGL-MDM2 reporter. Error bars indicate standard deviation of n = 3 experiments. Lower panel: loop on top of helix 11 and tandem repeats in helix 27 depicted on MEG3v1 secondary structure map. Human sequences of the H27 TRs, corresponding base pairing to the H11 terminal loop, and potential covariation of the base-pairing interaction is shown in the bottom left corner. D-domain, H-helix, J-junction, TR- tandem repeat.

3.5.2. Structural studies of the H11LpA functionally inactive mutant

Hypothesising that the structure of MEG3 was essential for its activity, I set out to test whether functionally-impairing mutations in H11 disrupts secondary and/or tertiary structure of MEG3. I wanted to test if the H11 terminal loop is responsible for keeping the MEG3 compact and globular, by forming a kissing loop with TR1-6. To do so I performed the same AUC and AFM experiments that I did with wt MEG3 but this time using mutant that has a poly A loop on top of helix 11. Subsequently I did chemical probing with 1M7 on the same mutant and compared the 1M7 reactivity on a single nucleotide level to see if this mutation causes any local structural changes.

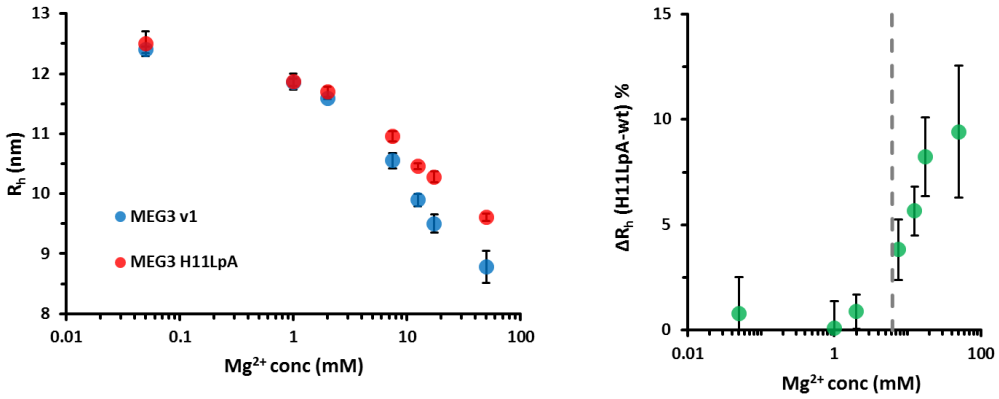


Figure 39: Structural defects of the H11LpA mutant (AUC).
On the left: Hill plot of the hydrodynamic radii of MEG3v1 and H11LpA with different concentrations of magnesium derived from the SV-AUC experiment. On the right: difference in hydrodynamic radii (ΔR_h) between v1 and H11LpA at increasing Mg^{2+} concentrations measured by AUC (the vertical dotted line indicates the $C_{Mg1/2}$ of v1. Error bars indicate standard deviation of n = 3 experiments.

3.5.2.1. Compaction and tertiary structure by AUC/AFM

By doing AUC I have seen that MEG3 H11LpA functionally inactive mutant also compacts with increasing concentration of Mg^{2+} but it never reaches the same level of compaction as MEG3v1 wt (figure 39). More precisely, at the concentrations of magnesium lower than the $C_{Mg1/2}$ wild type and mutant MEG3v1 have the same R_h, but increasing the concentration of magnesium above $C_{Mg1/2}$ MEG3 H11LpA progressively compacts less than MEG3v1 wt (R_h mutant = 10.28 ± 0.09 nm, R_h wild type = 9.50 ± 0.15 nm in 17.5 mM $MgCl_2$ i.e. 8 % difference in compaction).

I could even visualize on a single molecule level such dramatic defect in compaction of H11LpA by AFM in the presence of magnesium (figure 40, appendix figure 8). Statistical PSD analysis on 100 particles confirmed the defect in compaction (figure 40). Indeed, from the PSD analysis I see that the MEG3 H11LpA mutant in the presence of 10 mM Mg^{2+} has two characteristic inflection points at 30 nm and at ~110 nm, suggesting that it adopts a shape more similar to the partially-folded rather than to the fully-compact MEG3v1 wild-type.

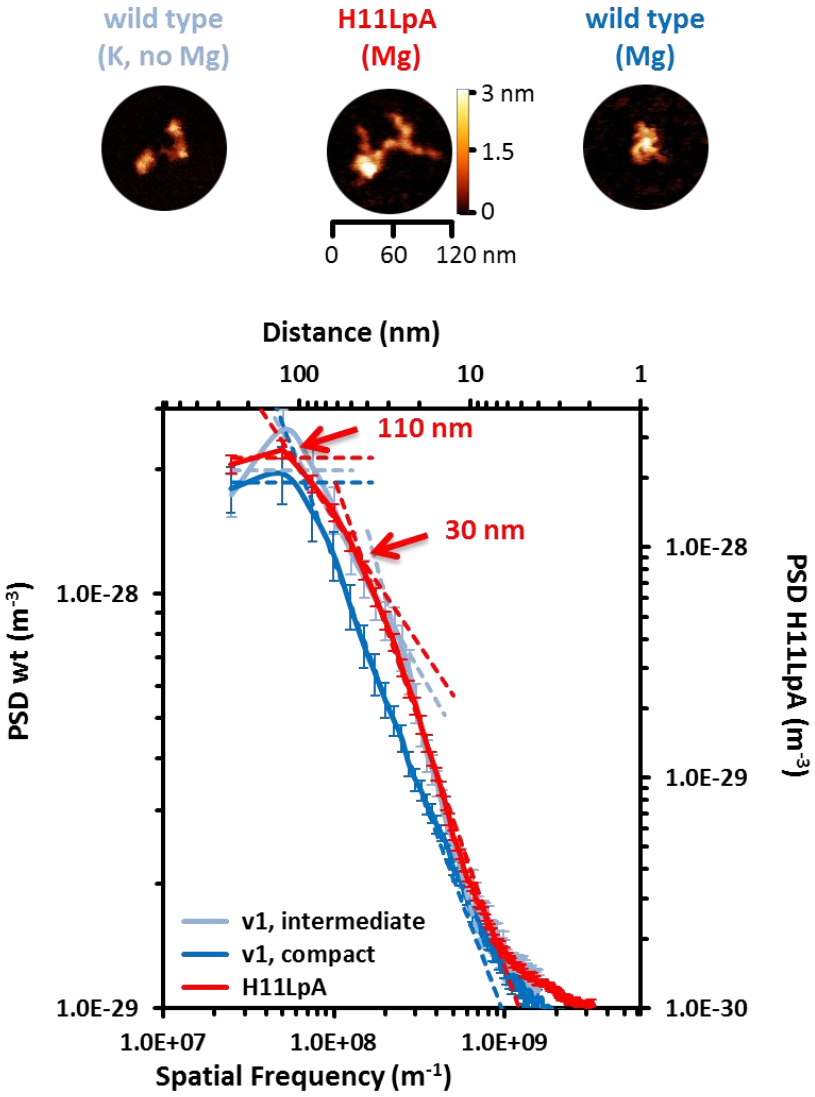


Figure 40: Structural defects of the H11LpA mutant (AFM).
Upper panel: representative AFM particles of MEG3v1 and of H11LpA in K^+ and Mg^{2+} (xy scale bar and z color scale bar are common to all samples). Bottom panel: PSD plot of >100 particles from AFM images of v1 (left y-axis) and H11LpA (right y-axis). Intercepts and error bars same as in figure 26.

3.5.2.2. *In vitro* chemical probing

Knowing that the inactive H11LpA mutant is unable to achieve the tertiary structure compaction as the wild type, I wanted to map such structural difference at the single nucleotide level. I performed chemical probing (SHAPE) with 1M7 and compared it to MEG3v1 wt SHAPE reactivity profile (figure 41). By comparing the SHAPE reactivity profile of MEG3v1 wt and H11LpA mutant, I noticed that most nucleotides have the same reactivity in the two RNAs with the Spearman's correlation coefficient between mutant and wt being 0.7410. This indicates that this mutation doesn't significantly disrupt secondary structure. Although overall SHAPE reactivity was similar, I decided to take a closer look at single nucleotide level. The most significantly affected were the nucleotides making the loop on top of helix 11, those nucleotides are unreactive in wt and they become highly reactive in poly A mutant (figure 41B). I haven't noticed such a dramatic increase in reactivity in any other part of MEG3 that could potentially base pair with loop on top of helix 11 (GUGAG) but 3'-stem side of H27 (nt 857-881), that is made of 6 tandem repeats that could base pair with loop on top of helix 11 (GUGAG), has significantly increased SHAPE reactivity in mutant when comparing the average and median values of reactivity of the region (figure 41C). This region also stands out when comparing the SHAPE reactivity profiles of the MEG3v1 wt and H11LpA full molecules (figure 41A).

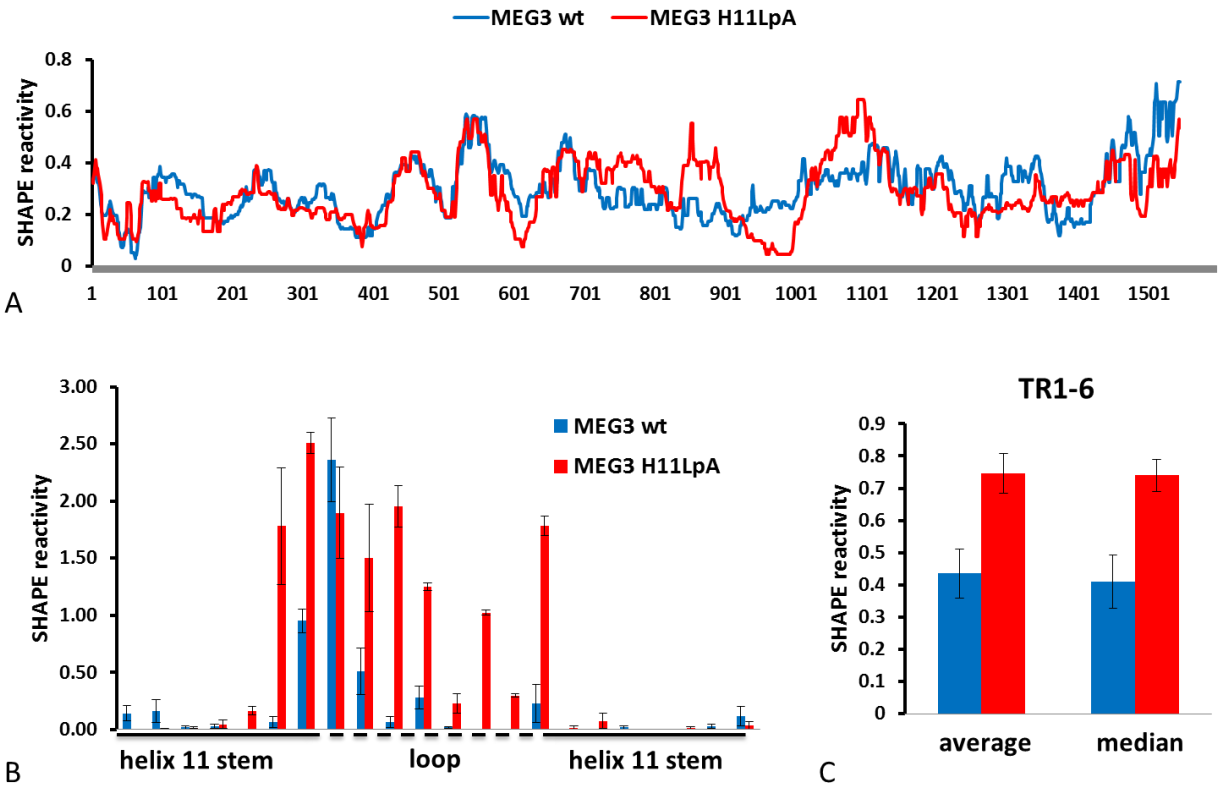


Figure 41: MEG3 H11LpA mutant SHAPE.

(A) 1M7 reactivity smoothened over 55 nucleotide window median. Blue line MEG3v1 wt, red line MEG3 H11LpA mutant. (B) Histogram of 1M7 reactivity of single nucleotides within helix 11. Blue bars MEG3v1 wt, red bars MEG3 H11LpA mutant. Error bars indicate standard deviation of n = 3 experiments. (C) Histogram of average and media 1M7 reactivity of 3'-stem side of H27 (nt 857-881). Blue bars MEG3v1 wt, red bars MEG3 H11LpA mutant. Error bars indicate standard deviation of n = 3 experiments.

4. Discussion

Résumé en Français

Dans ce travail, j'ai déterminé la structure secondaire expérimentale de deux variants d'épissage de l'ARNlnc humain MEG3, j'ai identifié le noyau fonctionnel de MEG3 et j'ai découvert qu'une interaction tertiaire à longue distance dans le noyau est nécessaire pour soutenir la fonction MEG3 et son intégrité structurale. Mon travail est allé au-delà de l'état de l'art dans le domaine des ARNlnc et a soulevé des points de discussion caractéristiques sur les propriétés de MEG3 et des ARNlnc en général.

Premièrement, le niveau élevé de conservation de la séquence et de la structure secondaire fait clairement de MEG3 l'un des ARNlnc les mieux conservés connus à ce jour. Deuxièmement, une caractéristique de nos structures MEG3 modulaires est que les limites du domaine correspondent aux jonctions exon de MEG3, suggérant une corrélation entre l'organisation génétique, l'architecture structurale et la fonction de MEG3. En effet, différents variants d'épissage ont des motifs structuraux communs dans l'exon commun 3 et diffèrent de manière significative en extrémité 3' de la molécule contenant les exons variables. J'ai identifié que cet exon commun 3, qui constitue les domaines 2 et 3, est un noyau structural et fonctionnel de MEG3 et qu'il comprend deux motifs structurés conservés H11 et H27 qui interagissent l'un avec l'autre. Troisièmement, c'est la première fois que nous voyons des mutations en un seul point dans une molécule d'ARNlnc altérant gravement son activité. Quatrièmement, c'est la première fois que nous fournissons des informations sur la structure 3D d'un ARNlnc.

MEG3 se replie en particules globulaires compactes à la fois en solution et sur support AFM, affichant un comportement très différent de celui des ARN de faible complexité et rappelant les ARN hautement structurés tels que les introns du groupe II. Fait intéressant, le mutant H11LpA fonctionnellement inactif n'est pas capable de se compacter, ce qui suggère qu'un échafaudage tertiaire compact régule l'activation de p53 dépendant de MEG3.

Dans leur ensemble, mes données suggèrent un mécanisme basé sur la structure pour l'activation sélective et ajustée de p53 par l'ARNlnc MEG3 et ouvrent de nouvelles possibilités intéressantes de recherche sur MEG3 en particulier et sur les ARNlnc en général.

4.1. MEG3 forms an intricate secondary structure dictated by alternative splicing

Human maternally expressed gene 3 (MEG3) is an abundant, imprinted, alternatively-spliced long non-coding RNA (lncRNA), which controls differentiation and neurodevelopment via Polycomb proteins, and proliferation via p53, acting as tumor suppressor. A combination of *in vivo* and *in silico* studies suggest that these MEG3 functions are regulated by the MEG3 structure (Zhang et al., 2010a). To understand MEG3 structure and function, I obtained the secondary structures of three MEG3 splice variants that have different p53 activation capacity.

I obtained MEG3 secondary structure maps by *in vitro* chemical probing with 3 different reagents (1M7, 1M6 and NMIA) and validated with fourth reagent DMS, *in vivo* chemical probing with 1M7, evolutionary covariation analysis and compensatory structural mutagenesis *in vivo* using a functional assay. My MEG3 structures are different from previously *in silico* predicted MEG3v1 structure (Zhang et al., 2010a) and recently obtained structure by 1M7-only *in vitro* and *ex vivo* probing (Sherpa et al., 2018). This difference are not surprising since *in silico* predictions for such long RNA molecules are still highly inaccurate, and chemical probing with 3 reagents increases the accuracy of RNA structure maps (Rice et al., 2014b) as well long range interaction >600nt are unlikely in RNA molecules [because 99 % of base pairs in long ribosomal RNA molecules are shorter distance (Deigan et al., 2009)] and Sherpa et al. don't use that constrain in building their secondary structure map. In the secondary structure map of Sherpa et al., 5' and 3' end (E1 and E12 to be more precise) base-pair forming a long stem. My functional assays show that this is structurally unlikely since 5' and 3' end of molecule have different effects on MEG3 functionality. To be more precise deleting E1 decreases p53 activation by MEG3 to 70 %, while deleting E12 increases p53 activation to 320 % with respect to wild type. Such opposite effect couldn't be explained if E1 and E12 base-paired with each other.

My structure maps reveal important properties of this lncRNA. MEG3 possesses 5 highly-structured domains with a level of structural complexity comparable to other lncRNAs. More precisely, MEG3v1 (1595 nt) forms 16 multi-way junctions and possesses 51 helices, lncRNA HOTAIR (2148 nt) forms 9 three-way junctions, 5 four-way junctions, 3 five-way junctions, a six-way junction and possesses 56 helices, SRA (873 nt) forms 4 three-way junction, a four-way

junction, a six-way junction and possesses 23 helices, COOLAIR (658 nt) forms a three-way junction, a six-way junction and possesses 10 helices, Braveheart (538 nt) forms a five-way junctions and possesses 9 helices (appendix table 3). Not all lncRNA are as highly structured as MEG3. For instance, XIST motif RepE (861 nt) only forms a three-way junction, a four-way junction and a seven-way junction and possesses only 12 helices. These comparisons suggest that MEG3 is an exceptionally well-structured lncRNA. To additionally validate my secondary structured maps I calculated Shannon entropies (S) of the structures which is a measure of well-definedness (Huynen et al., 1997). Indeed, Shannon entropies (S) of my MEG3 structures are particularly low for all three MEG3 isoforms (S_{ave}=0.125, S_{median}=0.09 for MEG3v1; S_{ave}=0.147, S_{median}=0.092 for MEG3v3; and S_{ave}=0.188, S_{median}=0.081 for MEG3v9; appendix table 1) compared to that of the structures of ribosomal RNA, group I intron, and group II intron (average of representative set of structures S= 0.253 ± 0.078) (Mathews, 2004). This observation, along with the fact that the SHAPE reactivity (R) of MEG3 is also globally very low (R_{median}=0.279 for MEG3v1; R_{median}=0.264 for MEG3v3; and R_{median}=0.264 for MEG3v9; appendix table 1), indicates that the MEG3 secondary structure is overall rigid and well-defined.

Interestingly, one characteristic feature of our modular MEG3 structures is that the domain boundaries match the MEG3 exon junctions, suggesting a potential correlation between MEG3 exon organization, structural architecture and function. This property was observed for another lncRNA, Braveheart (Xue et al., 2016), but not for lncRNAs HOTAIR (Somarowthu et al., 2015) or Xist (Smola et al., 2016). The correspondence of exon junctions and domain boundaries is particularly remarkable for MEG3, which possesses many splice variants (27 in humans) all exhibiting different p53 activation capacities. By comparing the secondary structure maps of 3 different MEG3 variants, I can now rationalize the structural and functional consequences that alternative splicing has on MEG3. In variants 1 and 9, exons 1-2-3, which encompass domains 1-2-3, form the same structure. In variant 3 because the lack of first 24 nt the overall structure is changed but the main structural motifs H9-11 and J3 within domain 2 and H25, H27-29 and J8-9 within domain 3 are preserved. Instead in all three variants, the insertion/deletion of a middle exon determines structural changes in domains 4-5. Functional differences between the MEG3 splicing variants are thus likely to be caused by structural changes in D4-D5, whereas D1-D3

provides the basal scaffold for p53 activation. From my *in vivo* chemical probing experiments, I could clearly see that the domains 4-5 are much less reactive *in vivo* indicating that in cellular environment they are protected by protein binding. Since MEG3 does not possess any sequence elements with known protein binding capacity but it has been proven that MEG3 directly interacts with proteins like the DNA binding domain of p53 and stabilizes p53 protein (Zhou et al., 2007; Zhu et al., 2015), it is possible that the structural motifs that binds p53 or other know protein partners like JARID2 and PRC2 are located in 5'-end of MEG3. Considering both observations, that domains 4-5 differ structurally between variants and that they are potential protein binding sites, it is possible that different variants bind different protein partners in domains 4-5 and thus exhibit different functional proficiency. Interestingly, results from HRF (performed by Eleni Anastasakou) support the hypothesis that D4-D5 may be protein binding sites because they reveal that D4-5 are solvent exposed regions of MEG3.

Besides obtaining insights into the global structural organization of MEG3, my structural probing efforts also helped me localize MEG3 motifs previously proposed to be functional. For instance, nt 19 to nt 38 in domain 1 were predicted to form a triple helix with DNA (Mondal et al., 2015) indicating that domain 1 might be important for localizing MEG3 at specific positions on DNA. Out of the 20 nt that are predicted to form a triple helix with DNA, half of them (10nt) are single stranded in my secondary structure map making it possible to form such interaction. Furthermore, around nt 345 in domain 2 a potential site of contact with PRC2 was proposed (Mondal et al., 2015). Interestingly, from my *in vivo* chemical probing, I also observe that nt 320-345 are less reactive *in vivo* indicating that this is a possible protein binding site. Additionally, average Shannon entropy of my secondary structure from nt 295 to nt 404 is specifically low (S_{ave} = 0.057) indicating that this motif is particularly well-defined. Considering all these observations, it is possible that the motif from nt 295 to nt 404 in domain 2 is responsible for binding proteins like PRC2.

Finally, my evolutionary analysis of the MEG3 sequence and secondary structure shows a surprising degree of conservation. E3, which I found to be the minimal unit of MEG3 needed for p53 activation, can be identified in as early-diverging mammals as *Marsupialia* (i.e. Tasmanian devil), suggesting that this lncRNA originated at least 200 million years ago, thus being about

half as old as p53, which split from the p53/p63/p73 ancestor in cartilaginous fish about 400 million years ago (Belyi et al., 2010). Notably, in early-diverging mammals (*Marsupialia*, *Afrotheria*, and *Xenarthra*), E1-E2 and E10-E12 cannot or only partially be detected (appendix table 4). It is possible that in those mammals the MEG3 sequence is just too divergent from humans and current alignment algorithms cannot identify it with confidence. Alternatively, in those species, MEG3 may have actually been composed of E3 only, which comprises the MEG3 structural and functional core, while other exons may have been acquired later in evolution, to confer further specificity to MEG3. Independent of how it actually evolved, its surprising level of sequence and secondary structure conservation clearly singles MEG3 out as one of the best conserved lncRNAs known to date. In general, structure conservation may be more important than sequence conservation for lncRNA. Some other lncRNAs have conserved parts of its secondary structure despite low sequence conservation, for example COOLAIR (Hawkes et al., 2016), SRA1 (Sanbonmatsu, 2016) and HOTAIR (Somarowthu et al., 2015). As well some other lncRNA like Megamind and Cyrano have function preserved but not the sequence (Ulitsky and Bartel, 2013), this phenomena could also be explained by structure conservation.

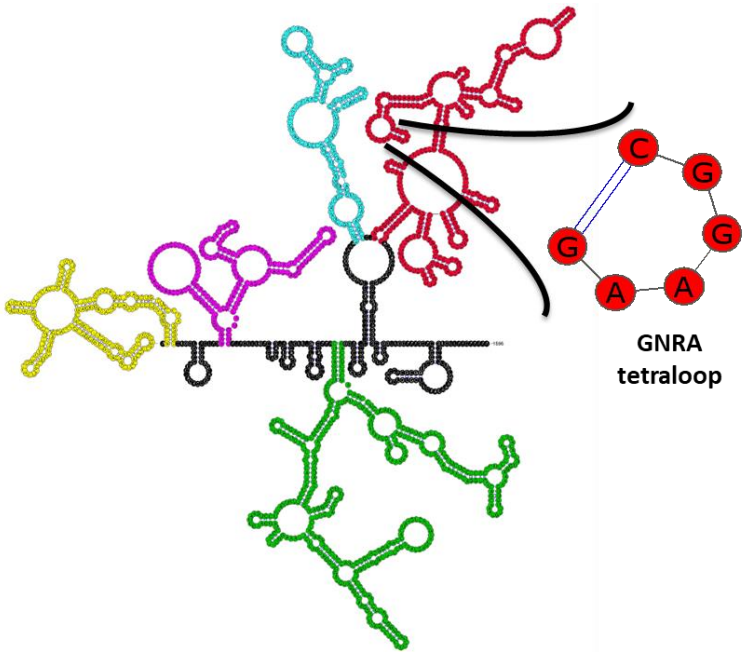


Figure 42: GNRA tetraloop.
MEG3v1 secondary structure map colored by domains with the zoom in to GNRA tetraloop.

4.2. MEG3 folds into compact globular particles

The structure also allows localization of motifs potentially involved in long range tertiary interactions. For instance, in domain 5 (motif 5) one terminal loop 5'-GGAA-3' (nt 1185-1188) is a well characterized GNRA motif (figure 42) that usually acts as "anchor" during tertiary folding (Jaeger et al., 1994). I thus started 3D structural investigation into MEG3, by MEG3 compaction at different ion strengths. Interestingly, I found that MEG3v1 folds in compact particles in solution at near-physiological concentrations of magnesium ions [$C_{Mg1/2}$ = 6.9±0.35 mM, cell concentration of Mg^{2+} ranges from 0.6 mM in free form to 10 mM bound to macromolecules (Ryschon et al., 1996)]. Moreover, the Hill coefficient of MEG3 compaction is 3.9±0.6, significantly higher than 1, thus indicating high cooperativity in magnesium ions binding and compaction of the MEG3 structure. Such value is significantly higher than for other lncRNAs, such as HOTAIR [n_{HOTAIR} = 1.1 ± 0.1, (Somarowthu et al., 2015)] and resembles more closely the folding of large, very well-structured catalytic RNAs, such as the ai5γ group IIB intron [n_{intron} = 2 ± 0.4, (Su et al., 2003)]. Having assessed that MEG3 compacts with addition of Mg^{2+} and considering the exquisite homogeneity of my MEG3 preparation, I then moved to visualize MEG3 in 3D by SAXS and AFM (figure 43). Additionally, the HRF profile of MEG3 (obtained by Eleni) in the presence of mono- but no di-valent ions supports my SAXS and AFM data showing that partial secondary fold forms in the absence of Mg^{2+} and in a modular fashion. These data offer the first low resolution 3D insight into a lncRNA and surprisingly suggest that MEG3 may be globular in its fully folded state. MEG3 folds into compact globular particles both in solution and on AFM support, showing behaviour very different from low complexity RNAs and reminiscent of highly structured RNAs such as the group II intron ribozyme.

In general, proving that lncRNA are structured may help us understand better their mechanism of action and shows that it is possible for a lncRNA to form interactions with other macromolecules based on their 3D structural fold and not their sequence. For example speculated role of lncRNA as a scaffold may be true (Tsai et al., 2010).

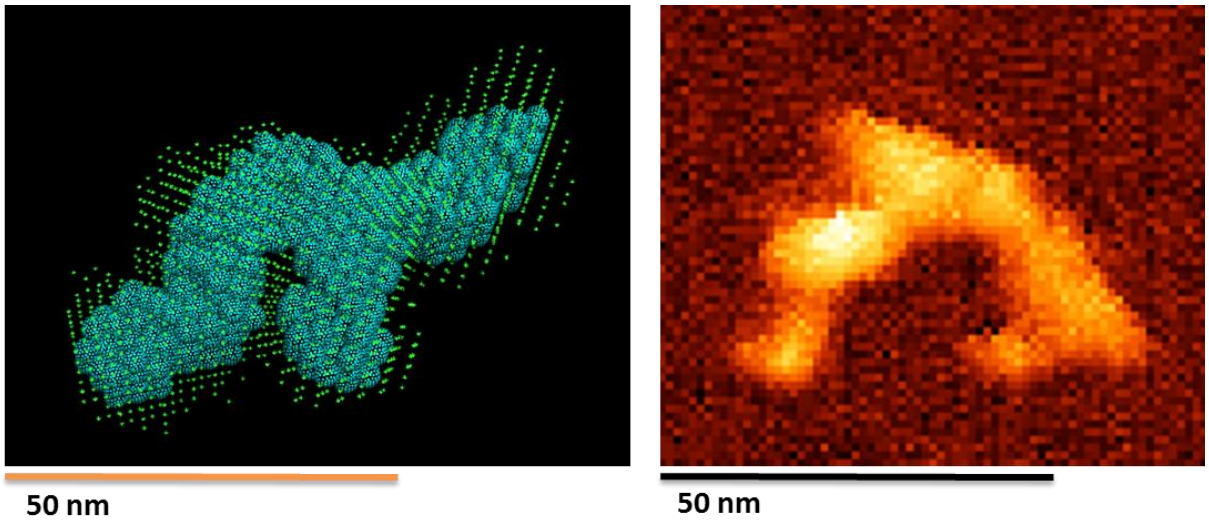


Figure 43: MEG3v1 folding intermediate single particles.
On the left Dammif 3D model generated from the SEC-SAXS experiment. On the right single particle from AFM experiment.

4.3. Structural and functional core of MEG3 comprises two conserved structured domains (D2-D3)

Studying the lncRNA functionality is challenging because of their poor sequence conservation and low levels of expression (Zampetaki et al., 2018). Nevertheless, there are numerous studies on lncRNA functionality, but there are not so many studies on how those functions are modulated and fine-tuned by lncRNA structure. In fact there is no direct experimental proof that connects lncRNA 3D structure with their function. To address this I performed the most systematic and most precise (single nt level) functional probing done so far for lncRNA.

As in existing studies on MEG3 functionality, we observed that MEG3 induces cell cycle arrest in G1 phase but not in G2 phase, but we also noticed that MEG3v1 is able to induce cell cycle arrest much earlier than MEGv3 and MEG3v9 adding another layer to specificity of different slice variants. It is possible that the MEG3v1 with the middle ability to activate p53 acts first and only later on MEG3v3 and v9 get activated, indeed v3 and v9 get expressed at the same level as v1 later on and at 48h post transfection they are more abundant than v1 (as seen from qRT-PCR experiment performed by Isabel Chillon). Further on, we observed that MEG3 seems not to induce apoptosis in HCT116 cells, whereas apoptosis was previously reported in SPC-A1 and A549 cells (Lu et al., 2013). It is still debatable in the field if lncRNA are functionally selective

and specific or if they act promiscuously or not have a functional role at all (Palazzo and Lee, 2015). My research shows that MEG3 is indeed selective and specific in its mode of action. I could see selectivity of MEG3 in HTC116 (it induces cell cycle arrest but not apoptosis) and specificity in recognizing between p53 REs.

Further using luciferase assays, I could probe my MEG3 structure and identify the structural motifs that play a role in MEG3-dependent p53 activation. From my current mutagenesis and functional studies, I can conclude that both domain 2 and 3 are essential and sufficient to activate p53, although at lower level than full length MEG3v1. This observation could explain the basal level of activity that all splicing variants exhibit, since domain 2 and 3 correspond to exon 3, an exon that is common to all splicing variants. Importantly, exon 3 is also well preserved in all mammals where MEG3 can be identified suggesting that the minimal functional core of MEG3 may have been preserved over evolution and that MEG3 is likely functional in all mammals where it is expressed. Domain 1 does not have a pronounced influence on the activation of p53. On the other hand deleting domains 4 or 5 or individual exons that comprise them increases p53-mediated transactivation by MEG3 above the level of splicing variant 1. In fact one of those mutants with deleted middle exon 5 is variant 9 that indeed shows higher level of p53-mediated transactivation. Hence, my current model is that domains 2 and 3 confer a basal activity to MEG3, while the rest of the molecule modulates such activity and this is what causes the different levels of activity of the different splicing variants.

In this study, though, I focused on the core of MEG3 (domains 2 and 3) to start understanding how the minimal functional portion of this lncRNA works. I did a sequence and structure alignment of these regions and found that the whole region is well preserved in MEG3 from other organisms. Particularly well conserved are helix 11 and helix 25. Interestingly the sequence of the terminal loop of helix 11 is exactly identical in at least 34 mammalian MEG3 sequences, covering 10 out of 12 orders of mammals. This analysis encouraged me to direct my attention to these structural regions and indeed preserved secondary structure of helix 11 and sequence of related loop showed to be essential. This is the first time that we see single point mutations in a lncRNA molecule abolish its activity completely. I tested a set of point mutations and found that GUGAG sequence on position 368-372 is absolutely necessary. GURAG is an RNA sequence typical of introns and possessing affinity to U1snRNP (Dhir et al., 2010) and the

GUGAG sequence is present in MEG3, although in exon not in intron. So I performed RT-PCR and sequencing to exclude the possibility of my point mutants not being expressed or aberrantly spliced. Indeed all my point mutants are expressed at the same level, although some are slightly less expressed (2-fold) that MEG3v1wt.

4.4. A compact tertiary scaffold regulates MEG3-dependent p53 activation

Since nucleotides in GUGAG terminal loop are non-reactive in chemical probing, indicating that they are not very flexible and rather constrained by base pairing I wanted to explore the possibility that GUGAG base pairs with different distal parts of MEG3 bringing them together and thus keeping the overall MEG3 structure compact and globular. Indeed, mutating the loop on top of the H11 terminal loop to poly A induces pronounced differences in the hydrodynamic properties of MEG3 (ΔRh ~ 8 % between v1 and H11LpA). I additionally confirmed inability of H11LpA mutant to compact by visualizing it on single particle level by AFM and performing power spectral density analysis on images. Interestingly, on a single nt level overall SHAPE reactivity doesn't change between inactive H11LpA mutant and MEG3v1 wt indicating that the overall secondary structure is preserved. But H11 terminal loop becomes very reactive in mutant suggesting that in wt those nucleotides form intramolecular interactions and those interactions are possibly responsible for MEG3 compaction which is essential for its activity. I haven't notice such a dramatic increase in reactivity in any other part of MEG3 that could potentially base pair with helix 11 terminal loop (GUGAG). But within H25-H29, the region that changes overall SHAPE reactivity the most is the 3'-stem side of H27 (nt 857-881). Similarly, H25-29, particularly in the 3'-side of the H27 stem, also changes in HRF reactivity (obtained by Eleni). 3'-stem side of H27 is made of 6 distal tandem repeats that could potentially base pair with GUGAG terminal loop in human. Occurrence of 3 to 6 tandem repeats is also observed in all other mammalian species where we could identify MEG3 sequence. Additionally, in humans the compensatory mutations at the position of every individual TR are able to restore partially function that is lost in $G^{370}C$ H11 point mutant. This result supports the theory that the tandem repeats in H25-H29 region of MEG3 interacts with the loop on top of H11 and likely forms alternative, mutually exclusive kissing loops. This redundancy, existence of 6 tandem repeats, might be beneficial and functionally important in RNA as it is for proteins (Andrade et al., 2001).

In proteins such repeats may create a larger binding surface areas for cofactors (Andrade et al., 2001), similar scenario is possible in MEG3 case. It is possible that loop on top of helix 11 (GUGAG) interacts dynamically with TR1-6 creating slightly different conformers and modulating the surface of MEG3, which can promote different interactions with p53 itself or other cofactors involved in the p53 pathway and thus regulate a specific response on different p53 target genes.

In conclusion, the 3D structure of MEG3 formed by interaction between H11 and H27 is essential to activate p53. Therefore, my work establishes proof-of-concept that the 3D structure of an lncRNA is directly linked to its biological function, resolving a long-standing debate in the field.

4.5. A structure-based mechanism for selective and fine-tuned p53 activation by lncRNA MEG3

My biochemical and biophysical characterization of human MEG3 integrated with the existing cellular and medical characterization start to elucidate the mechanism of action of MEG3 as a tumor suppressor through the specific and selective activation of the p53 pathway. I propose the following model (figure 44). In physiological conditions MEG3 is expressed in the nucleus of certain tissues, such as the brain, pituitary and adrenal gland. Different splicing variants are present in different percentage in different cell types but MEG3v1 is always the most abundant. In MEG3v1, the terminal loop of H11 (D2) dynamically interacts with TR1-6 on H27 (D3) forming a sliding kissing loops and that interaction is responsible for compaction of the MEG3 core (D2-3) leading to formation of a functional tertiary structure. This core is present in all splicing variants and it is essential for function. Around the D2-3 core, D4-5 form the solvent exposed portion of MEG3. This outer layer differs between splicing variants and serves to protect the functional core as well to fine tune the selectivity and specificity of different variants possibly by interacting with different cofactors and the proteins present in the cellular environment. Thus, the action of MEG3 also depends on the cellular environment. When the levels of p53 protein in the cell are low the effect of MEG3 of p53 pathway is probably also reduced. Instead, when the cells are under stress and cellular levels of p53 are high and other co-activators are produced, particularly $p14^{ARF}$ or $p16^{INK4a}$ that are functionally related to MEG3 (Zhang et al.,

2010b), MEG3 is active. Once correctly folded, MEG3 activates p53 response in the cell possibly by a direct interaction with DNA binding domain of p53 (Zhu et al., 2015). p53 induces selective expression of some of its target genes (Zhou et al., 2007) which leads to cell cycle arrest at G1 phase and/or apoptosis depending on the cell type. This mechanism contributes to tumor suppression, restoring the physiological conditions under stress, and thus not surprisingly the loss of MEG3 expression leads to development of tumors (Cheunsuchon et al., 2011).

Screening for structure-disrupting mutations in the MEG3 gene, particularly in the two key functional motifs H11 and H27, may serve as a useful biomarker for identifying patients with increased cancer susceptibility.

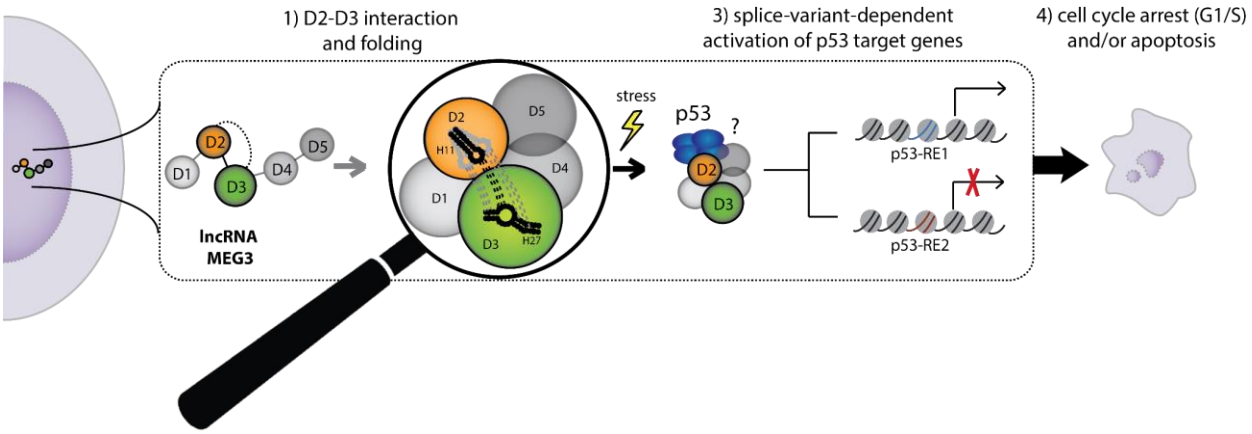


Figure 44: Model for MEG3-dependent p53 activation.

In the nucleus of healthy cells (sketched on the left), MEG3 is transcribed and folds guided by the H11-H27 interaction between D2 and D3 (dotted arrow). Under stress (yellow lightning bolt), compact MEG3 activates p53 inducing selective expression of certain p53 target genes. The resulting effect of MEG3-dependent p53 activation is cell cycle arrest at the G1/S checkpoint and/or apoptosis, depending on the cell type (sketch of an apoptotic cell on the right).

4.6. Future perspectives

My work just started to explain the potential mechanism of action of this important lncRNA MEG3, and opens a lot of possibilities for future research. This project could be continued in

many different ways, and I'm going to discuss just some of the numerous possibilities that I am currently pursuing or that I envisioned but won't be able to perform by the end of my PhD.

To confirm the interaction between H11 and H27 I designed a series of new mutants including the second set of compensatory mutations on H27 that compensate for $A^{371}U$ mutation on H11 loop; some triple mutants that have $G^{370}C$ and two compensatory mutations on H27 (TR3 + TR1-6) to see if they are additive; mutant that has $G^{370}C$ and all the potential compensatory mutation on H27 that wouldn't disrupt the secondary structure; and finally a mutant that has $G^{370}C$ and all the potential compensatory mutation on H27 together with compatible mutations on 5' end of helix 27 to preserve the secondary structure. Subsequently to catch the interaction *in vitro* and *in vivo* RNA crosslinking could be done. *In vitro* short-range crosslinking using thionucleotides with primer extension mapping could be used (Harris and Christian, 2009), for the detection of interaction *in vivo,* one of the existing protocols could be adopted with modified detection step to only check for this specific region (Aw et al., 2016; Nguyen et al., 2016; Sharma et al., 2016). All three protocols have been used to capture the full RNA interactome but they also captured some intramolecular lncRNA interactions, such as for MALAT1 (Aw et al., 2016; Nguyen et al., 2016; Sharma et al., 2016). As for the *in vivo* studies presented in this work WI38 fibroblast could be used to check the endogenous MEG3 and MEG3v1 transfected in HCT116 to unambiguously catch just one splicing variant. Additionally, to confirm *trans* interaction between D2 and D3 *in vitro* one could run native TB agarose gels with radioactively labelled D2 and ΔD2.

To test if MEG3 is directly binding p53, one could do *in vitro* electrophoretic mobility shift assay (EMSA) with p53 constructs containing DNA binding domain that has been previously shown to bind MEG3 (Zhu et al., 2015) and MEG3 wt, representative mutants. To test if MEG3 is interacting with some other proteins one can do a RNA pull down with MEG3 wt and representative mutant and compare the proteomes of two (Castello et al., 2016). For example HyPR-MS was used to catch at the same time lncRNA-protein interactomes for MALAT1, NEAT1, and NORAD (Spiniello et al., 2018). By doing the RNA pull down and analysing the proteome we can also see if MEG3 interacts directly with p53 *in vivo*.

To determine connection of MEG3 with different signalling pathways involved in tumor progression or suppression, one can check if over- or under- expression of certain proteins is correlated with transfection of different MEG3 constructs. To be more precise, one could do a western blot with antibodies against p53, MDM2, GDF-15, Rb, $p16^{INK4a}$, Wnt/β-catenin, Notch, Rac1 and VEGF on control cells (transfected with empty vector) and on cells transfected with MEG3 and selected mutants. Depending on which pathway we want to test we should use a cell line that has desired pathway preserved.

Since to date we were not successful to obtain a high resolution 3D structure of full length MEG3 alone, finding the region of MEG3 that is binding selected proteins would be useful. Ribonucleoprotein complexes between MEG3 and a certain protein could potentially be a good candidate for structural studies by cryo-EM or X-ray crystallography. So far there is no 3D structure obtained for RNA alone with cryo-EM but there are several examples of obtaining a structure of RNA when putting it on a grid together with a protein that stabilises it. Best examples are numerous structures of ribosomal subunits (for example PDB ID: 1c2w) but as well there are other examples such as group II intron complexed with its reverse transcriptase (PDB ID: 5g2y) [information obtained from PDBe (PDBconsortium, 2019)]. Obtaining a high resolution 3D structure of MEG3 would tell us more details on the molecular mechanism of its action. To identify RNA and protein regions that are interacting one could do cross-linking MS experiments, for example one of the recently developed protocol (Trendel et al., 2019). One practical example of mapping a protein binding sites on lncRNA is PAR-CLIP and mapping of CLIP-seq data done to identify sites on NEAT1 that are binding selected proteins (Yamazaki et al., 2018). More low-throughput but simpler to perform would be to do EMSA with full-length wild type MEG3 and MEG3 mutants (existing ones and designing the new ones) and with wild type and mutant proteins (Hellman and Fried, 2007).

Acknowledgements

I would like to thank my supervisor Dr. Marco Marcia for giving me an opportunity to work on such interesting project, guiding me through it and at the same time giving me the freedom to do it my way.

I would like to thank my thesis advisory committee members: Dr. André Verdel (IAB Grenoble), Dr. Alex Bateman (EMBL EBI Hinxton), Dr. José Antonio Márquez (EMBL Grenoble) for their input, advice and discussions during my TAC meetings.

I would like to thank my jury members: Prof. Claire Vourc'h, Dr. Maite Huarte, Dr. Alberto Inga, Dr. Reini Fernandez de Luco and Dr. Janosch Hennig for evaluating my thesis.

I also want to thank Ombeline Pessey for excellent technical assistance; Eleni Anastasakou for *in vitro* SHAPE of MEG3v3 and HRF of MEG3v1; Isabel Chillon for flow cytometry, RT-PCR and optimization of functional assays; Mani Karuppasamy for EM imaging; and all members, current and past, of the Marcia lab for helpful discussion.

I would like to thank Jean-Marie Teulon and Dr. Jean-Luc Pellequer (Institut de Biologie Structurale, Grenoble) for AFM data acquisition; Dr. Paolo Annibale (Max Delbruck Center, Germany) for AFM PSD analysis; Dr. Alberto Inga (Center for Integrative Biology, Italy) for pGL-p21/MDM2 RE vectors; Dr. Christine Ebel and Aline Le Roy (Institut de Biologie Structurale, Grenoble) for assistance with SV-AUC; Prof. Marc Jamin (University Grenoble Alpes, IBS) and Dr. Caroline Mas (PSB, Biophysical platform) for their assistance with SEC-MALLS; Dr. Adam Round and scientists at the BioSAXS beamline BM29 (ESRF, Grenoble) for assistance with SAXS data collection and processing; Samira Acajjaoui and Dr. Montserrat Soler Lopez (ESRF Grenbole) for support with DLS; the Chemical Biology Core Facility (EMBL Heidelberg) for synthetizing the SHAPE reagent 1M7; Dr. Yunli Zhou (Massachusetts General Hospital, Boston) for sharing the pCMS-d2EGFP-MEG3, pCI-control, pCI-MEG3, p53-Luc,pCMVbeta, and pCI-p14ARF vectors; Prof. Anna Pyle and Rafael Tavares (Yale University) for sharing RScape scripts before publication; Dr. Malte Paulsen (EMBL Heidelberg), Dr. Mylène Peyzet (IAB Grenoble), and Dr.

Jean-Philippe Kleman (IBS Grenoble) for support with flow cytometry; and the EMBL Gene Core Facility (EMBL Heidelberg) for performing high-throughput sequencing related to *in vivo* SHAPE; Stuart Fisher for adaptation of Superfold.py script.

I would like to thank Danielle Desravines and Luc Sinardet for helping with translation of chapter summaries to French.

I would like to thank my master thesis mentor dr.sc. Krešimir Pavelić for encouraging me to da a PhD.

I would like to thank all EMBL Grenoble PhD students that overlapped with me and Branislava Ranković for sharing all ups and downs of a PhD path with me.

I would like to thank Maria Lukarska for forwarding me all the thesis advices and for a constant supply of chocolate.

I would like to thank my gym partner Michael Hons for helping me keep my body and mind in a good shape.

I would like to thank my "lab babies" Francesca Chandler, Lena Anastasakou, Anna Hoefler and Coltrane Morley Wiliams for all the needed coffee breaks and weekend fun.

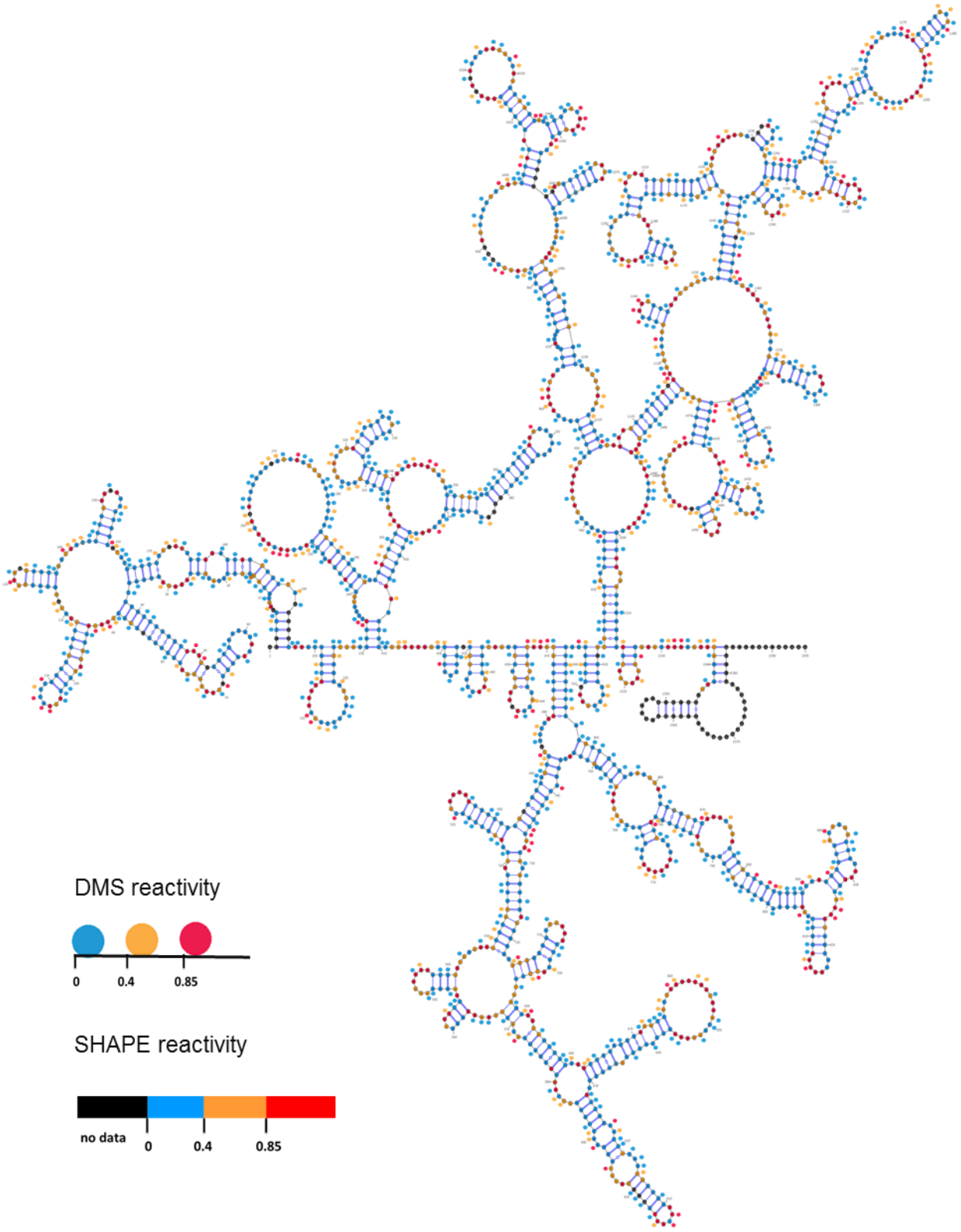
I would like to thank to my "collègue de travail" Ombeline Pessey for strictly 'professional' relationship and help to integrate in France (especially teaching me about the polite complaining).

I would like to thank my Grenoble family (Irene Garcia Ferrer and Marc Sansa) for adopting me like their own.

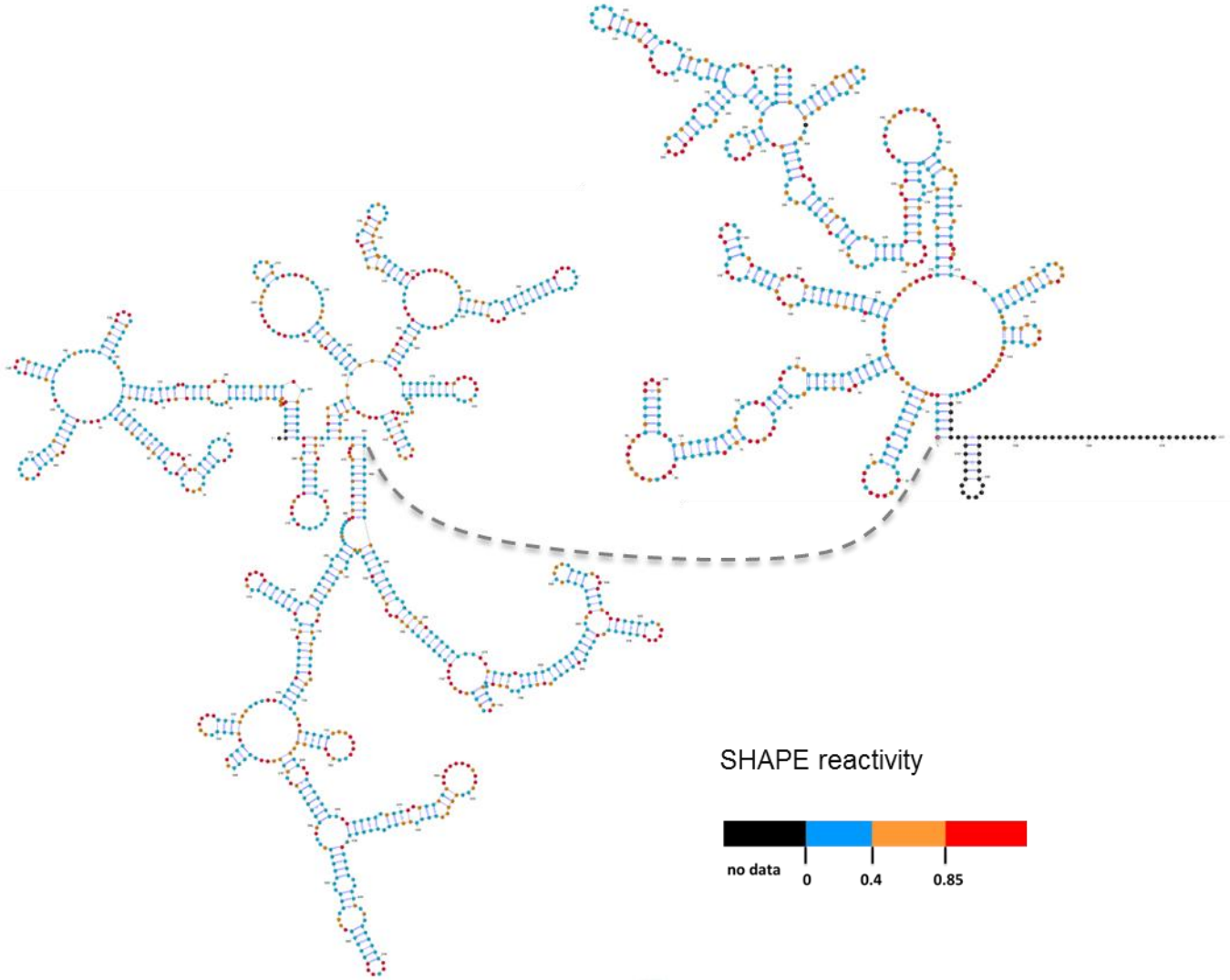
Zahvalila bi se i svim mojim curama Kristini Vukušić, Ingi Kavazović, Valentini Marković Jelušić, Tamari Relić, Tjaši Juraković, Tamari Kopjar i ponajviše Petri Kozina te Aniti Radojević što su uvijek tu za mene ☺

Na kraju jedno veliko hvala mojoj obitelji mami Ljiljani, tati Damiru i bratu Luki za njihovu podršku i što su mi omogućili da budem ovdje <3

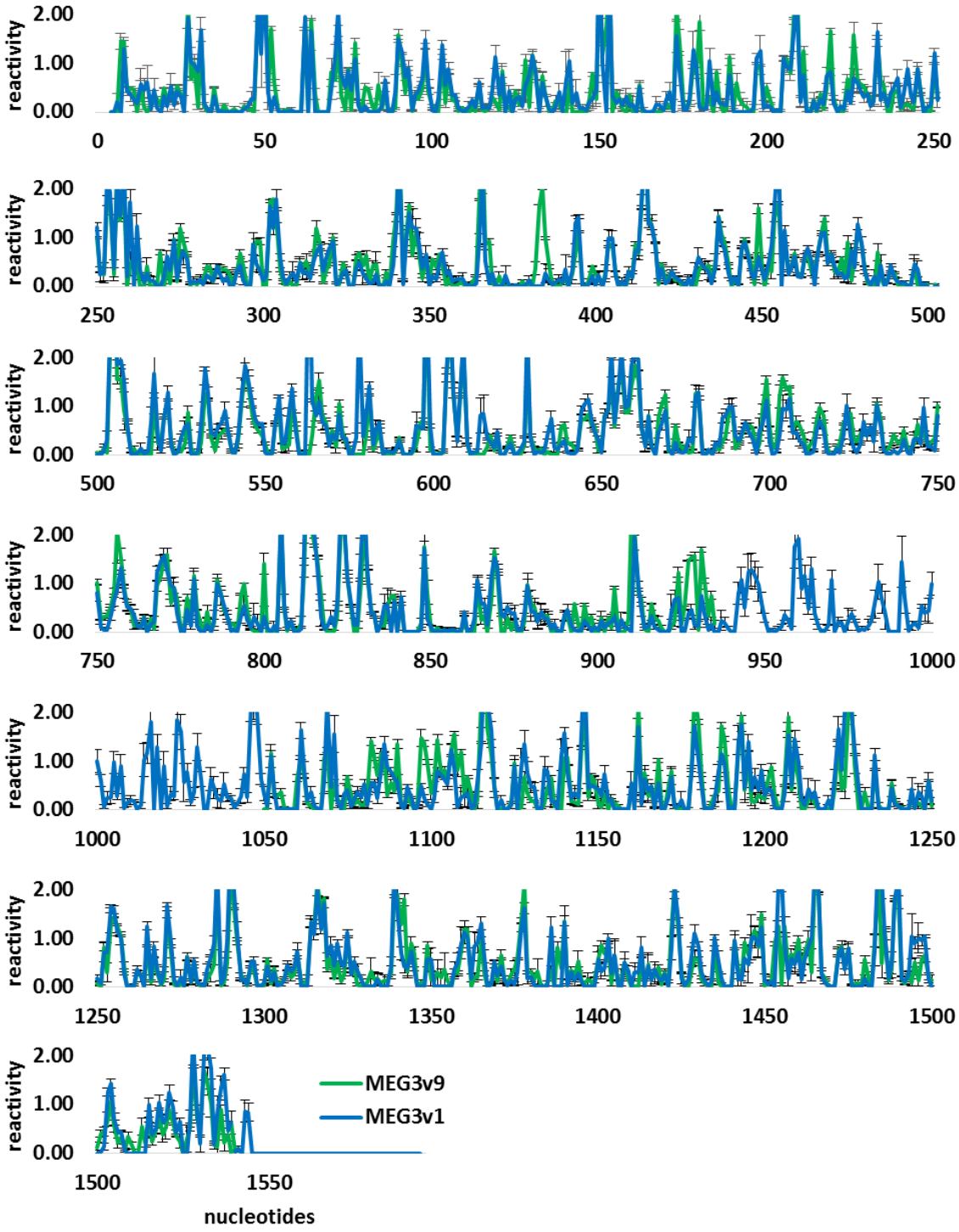
Appendix



Appendix Figure 1: MEG3v1 secondary structure map by chemical probing (SHAPE and DMS). Colour map depicted in lower left corner.



Appendix Figure 2: MEG3v9 secondary structure map by chemical probing (SHAPE).
Colour map depicted in lower right corner.



Appendix Figure 3: *In vitro* SHAPE probing of v1 and v9.

The graph reports 1M7 reactivity values of individual nucleotides. Error bars indicate standard error of the mean from n = 3 experiments. The y-axis report SHAPE reactivity. Nucleotides with very high reactivity values (> 2) are excluded for clarity of visualization. The x-axis reports nucleotide numbers for the v1.

Appendix Table 1*: In vitro* chemical probing (SHAPE) statistics.

$CC_{Spearman}$ between replicas		
	MEG3v1	
r1 vs r2	**r1 vs r3**	**r2 vs r3**
0.708	0.677	0.746
	MEG3v9	
r1 vs r2	**r1 vs r3**	**r2 vs r3**
0.866	0.832	0.853
$CC_{Spearman}$ between splicing variants		
v1 vs v9	**v1 vs v3**	**v9 vs v3**
0.784	0.639	0.695

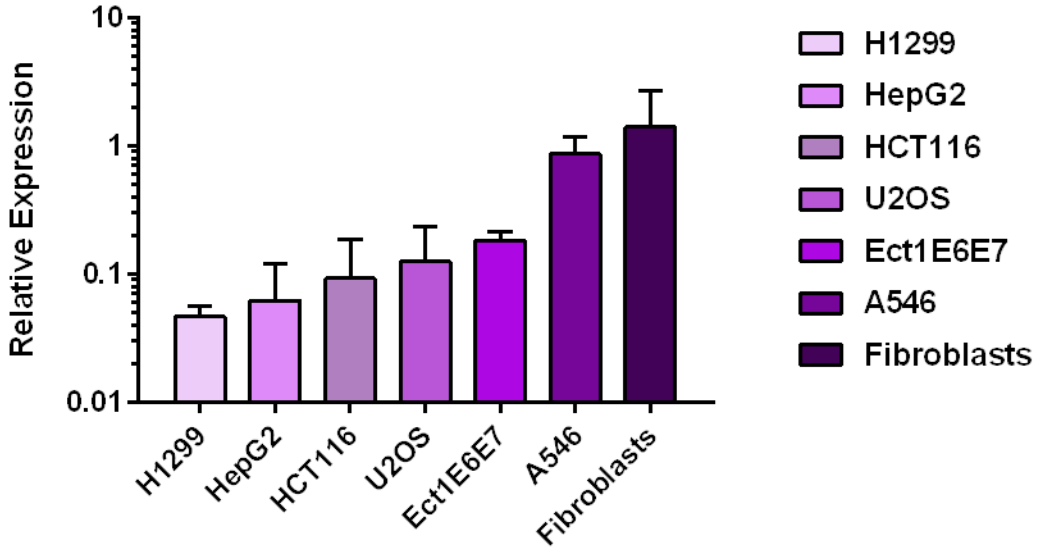
Snannon entropy and SHAPE reactivity		
	MEG3v1	
S_{ave}	**S_{median}**	**R_{median}**
0.125	0.09	0.279
	MEG3v9	
S_{ave}	**S_{median}**	**R_{median}**
0.188	0.081	0.264

Appendix Table 2: Secondary structure motif present in MEG3v1 secondary structure map.

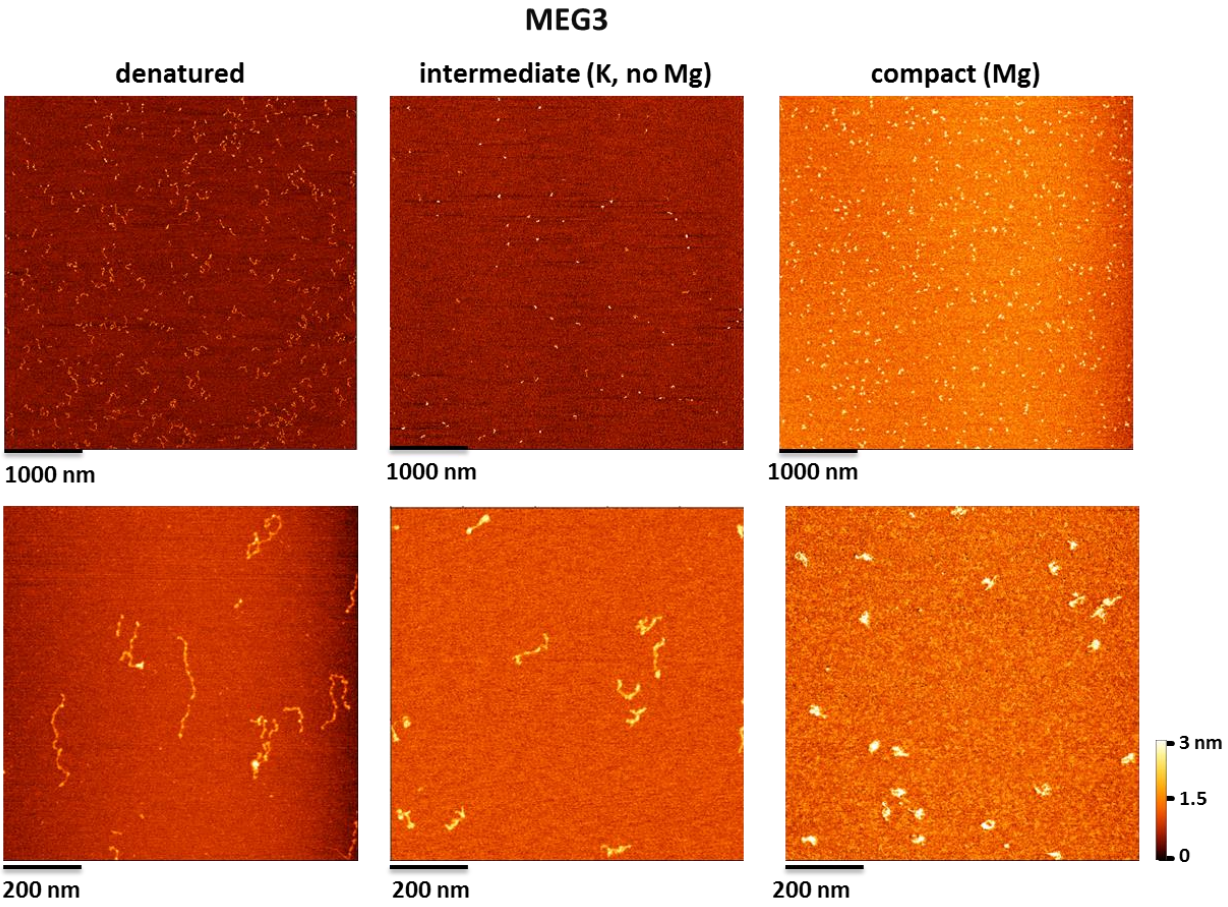
Motif class	Secondary structure motif	MEG3
Stem loop (N_{loop})	-3	3
	-4	7
	-5	3
	-6	10
	-7	2
	-8	5
	-9	1
	-11	1
	-15	1
	-18	1
	-21	1
	-35	1
Multiway junction	3WJ	12
	5WJ	3
	6WJ	1
Symmetric internal loop ($N_{5'}$,$N_{3'}$)	(1, 1)	9
	(2, 2)	5
5'-Assymetric internal loop ($N_{5'}$,$N_{3'}$)	(1, 0)	8
	(2, 0)	2
	(3, 0)	1
	(4, 0)	1
	(4, 2)	1
	(4, 3)	2
	(5, 2)	1
	(9, 0)	1
	(10, 8)	1
3'-Assymetric internal loop ($N_{5'}$,$N_{3'}$)	(0, 1)	1
	(0, 5)	2
	(1, 3)	1
	(3, 15)	1
	(4, 7)	1
	(5, 6)	1
	(10, 13)	1

Appendix Table 3: Comparison of structural motifs in MEG3v1 and other experimentally mapped lncRNAs.

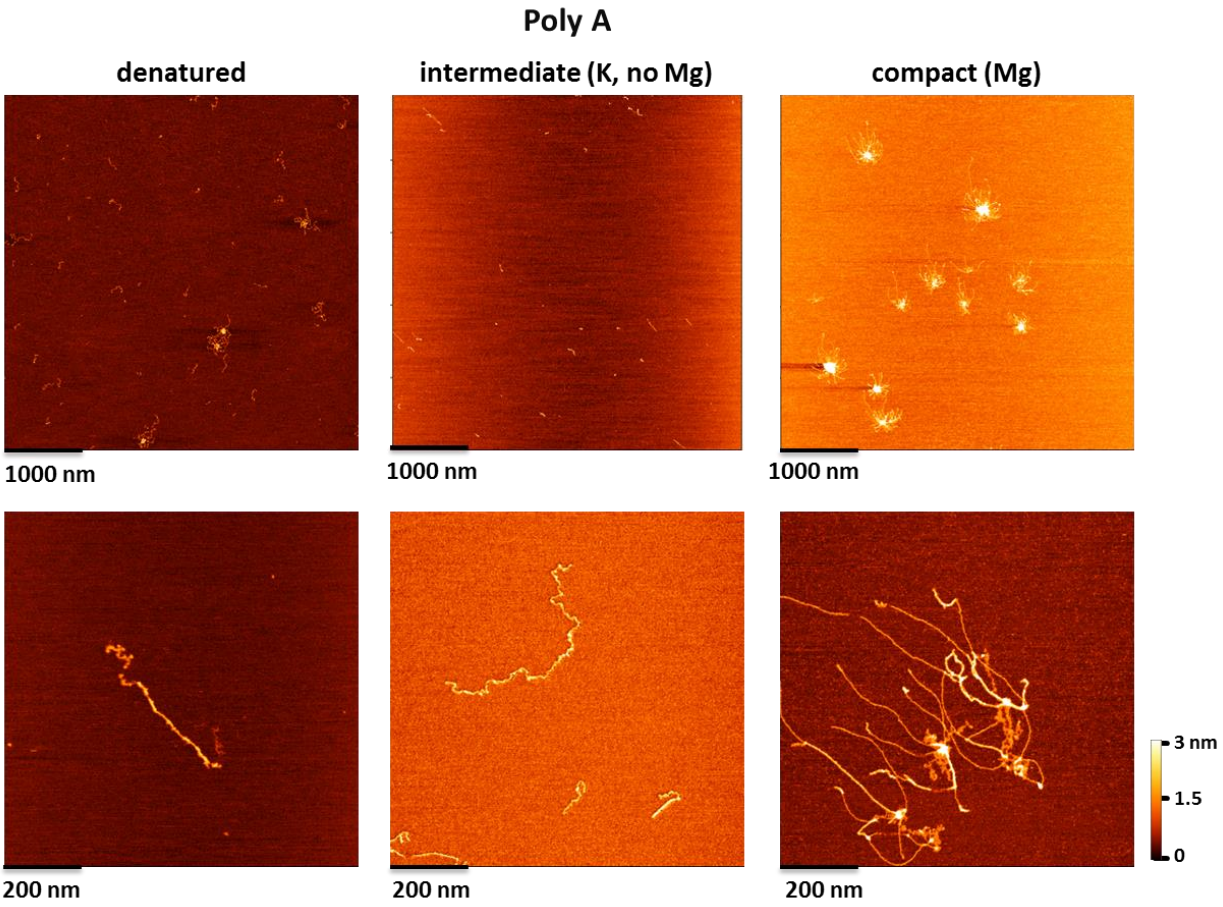
	MEG3	HOTAIR	SRA	COOLAIR
Size / nt	1595 nt	2148 nt	873 nt	658 nt
Stem loop (Nloop)	36	38	17	8
Multiway junction (3WJ)	12	9	4	1
Multiway junction (> 4WJ)	4	9	2	1
Symmetric internal loop	14	17	12	2
5'-Asymetric internal loop	18	28	8	10
3'-Asymetric internal loop	8	24	9	10



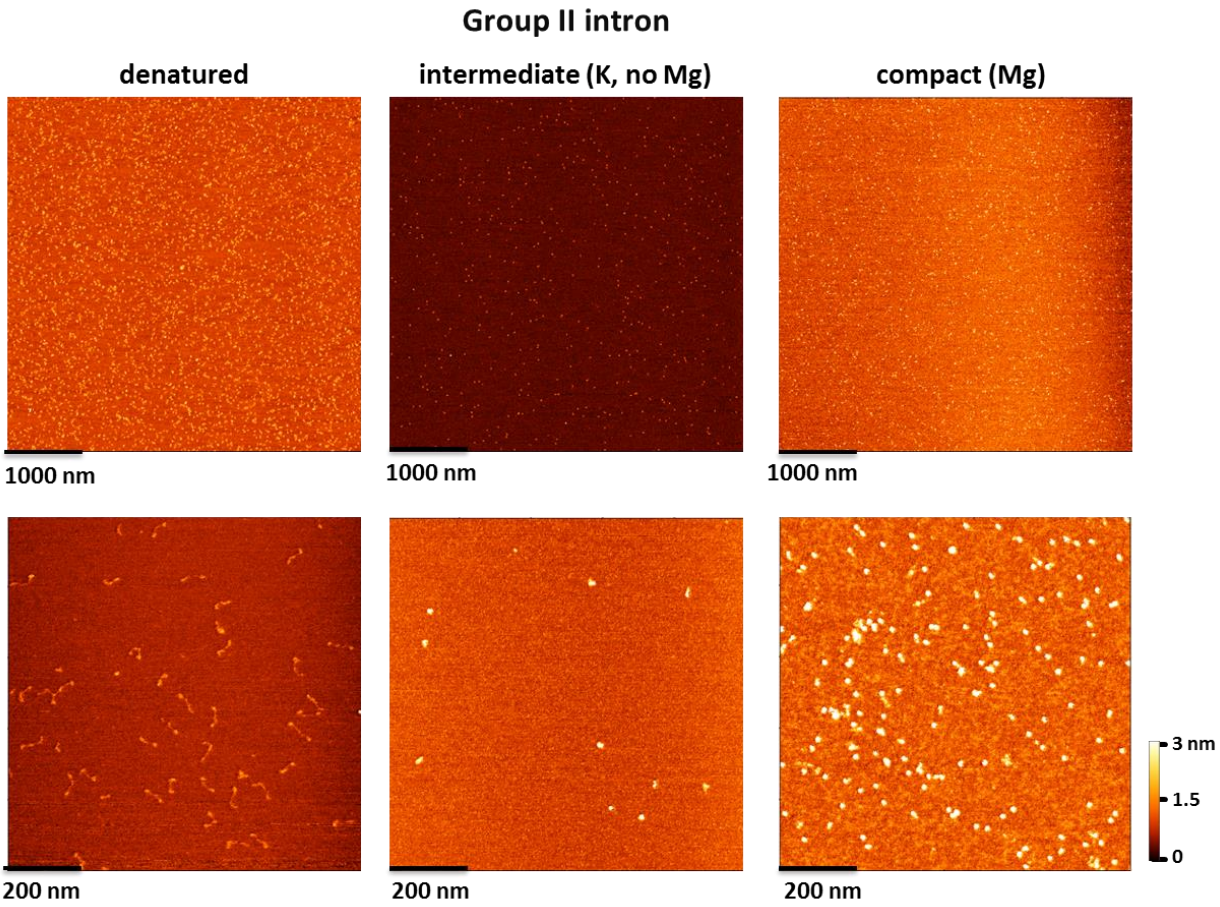
Appendix Figure 4: qRT-PCR levels of MEG3 expression in different cell lines.
Relative abundance of endogenous MEG3 in different cell lines, as determined by RT-PCR. Error bars indicate standard deviation from at least three independent experiments. *Figure was created by Isabel Chillon.



Appendix Figure 5: Representative AFM images from which particles were picked (MEG3). Related to figure 25. One representative AFM image, processed in Gwyddion, per state. States are indicated on top of the respective image.

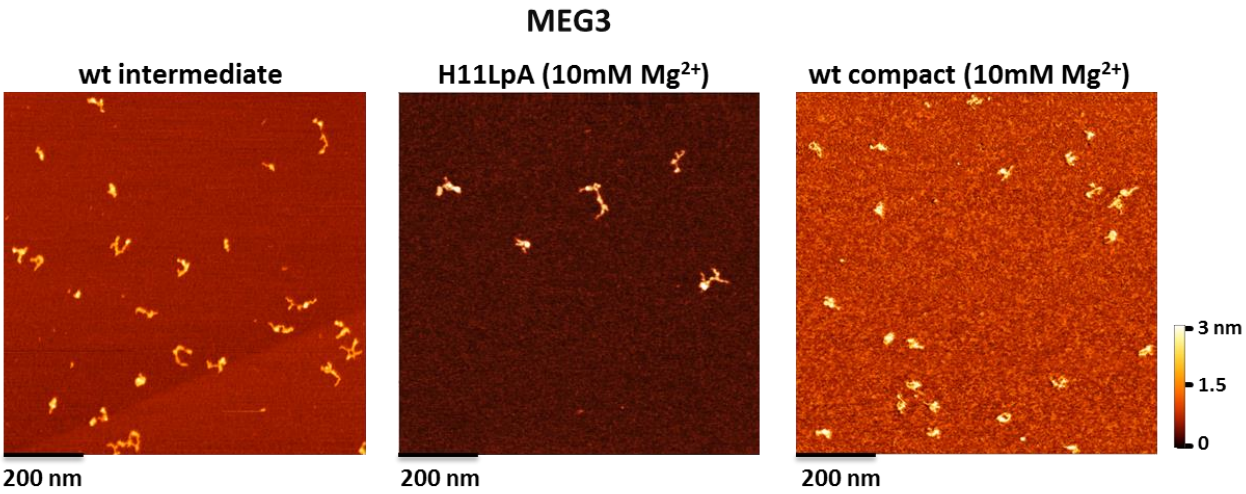


Appendix Figure 6: Representative AFM images from which particles were picked (Poly A). Related to figure 25. One representative AFM image, processed in Gwyddion, per state. States are indicated on top of the respective image.



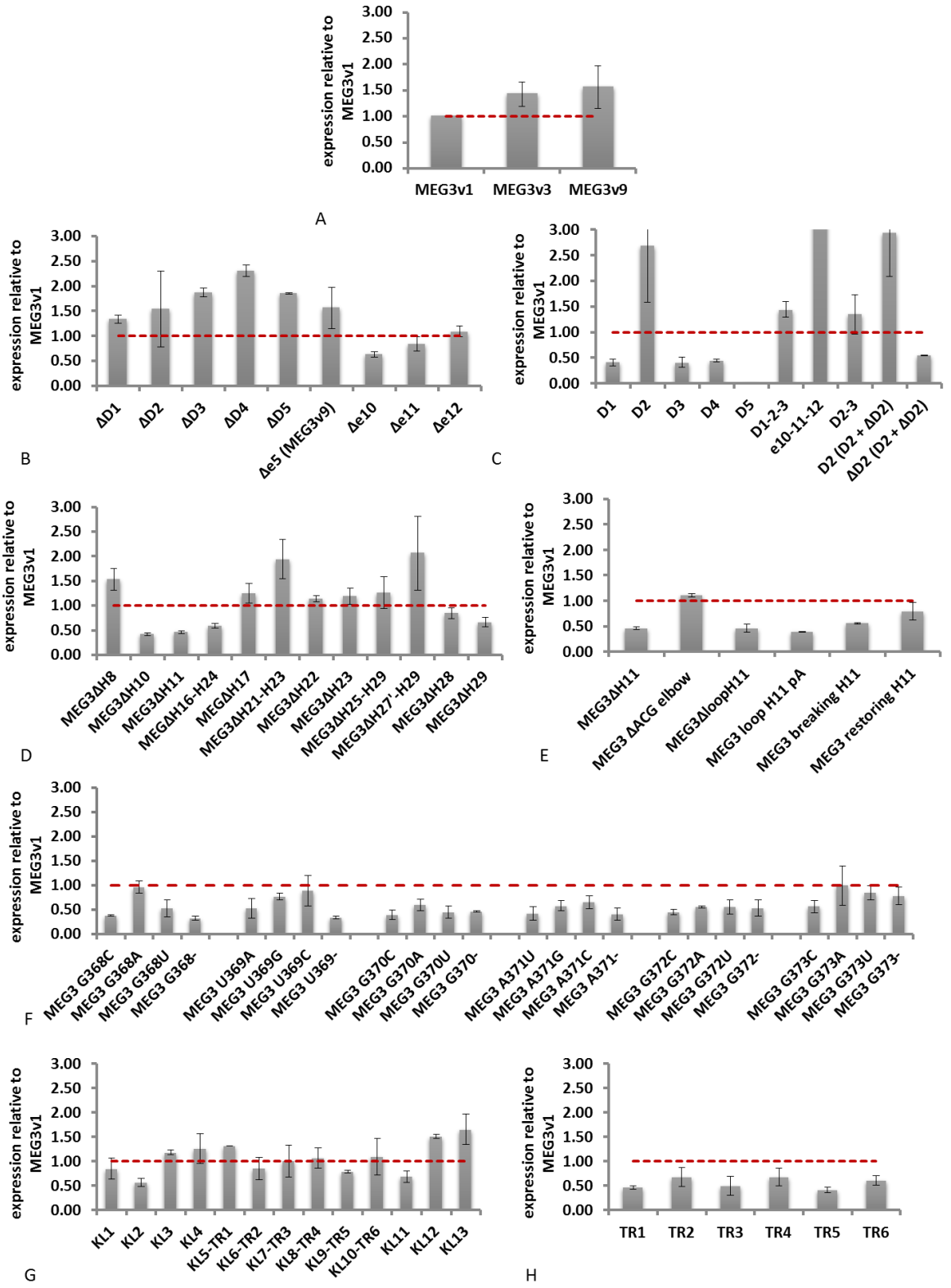
Appendix Figure 7: Representative AFM images from which particles were picked (group II intron).

Related to figure 25. One representative AFM image, processed in Gwyddion, per state. States are indicated on top of the respective image.



Appendix Figure 8: Representative AFM images from which particles were picked (MEG3wt vs H11LpA).

Related to figure 39. One representative AFM image, processed in Gwyddion, per state. States are indicated on top of the respective image.



Appendix Figure 9: qRT-PCR.

Abundance of all mutants in transfected HCT116 cells relative to actin and neomycin mRNAs. Values are expressed as fold changes with respect to the abundance of v1 (=1, dotted red line). All constructs were expressed under the same conditions used for functional assays. Error bars indicate standard error mean from n = 2 experiments. Construct D5 is > 10000 fold less expressed than v1.

Appendix Table 4: Evolutionary conservation of the MEG3 exons.

Order (subgroup)	Species Common name	Species Latin name	Chromosome	Transcript ID (NCBI reference number)	BLAT[1] D1 (E1-E2)	BLAT[1] D2-D3 (E3)	BLAT[1] D4-D5 (E10-E12)	Infernal D2-D3[2] (E3)	Infernal H11-J8 (putative KL)
Primates									
(Humans)	Human	*Homo sapiens*	14	NR_002766.2	E1-E2	E3	E10-E12	seed	yes
(Great apes)	Chimpanzee	*Pan troglodytes*	14	hypothetical	E1-E2	E3	E10-E12	seed	yes
	Bonobo	*Pan paniscus*	14	hypothetical	E1-E2	E3	E10-E12	final	yes
	Gorilla	*Gorilla gorilla*	14	hypothetical	E1-E2	E3	E10-E12	final	yes
	Orangutan	*Pongo abelii*	14	NR_037685.1	E1-E2	E3	E10-E12	final	yes
(Lesser apes)	Gibbon	*Nomascus leucogenys*	22a	hypothetical	E1	E3	E10-E12	final	yes
(Monkeys)	Rhesus macaque	*Macaca mulatta*	7	hypothetical	E1-E2	E3	E10-E12	seed	yes
	Baboon	*Baylor panu*	7	hypothetical	E1-E2	E3	E10-E12	final	yes
	Golden monkey	*Rhinopithecus roxellana*	KN299482v1	hypothetical	E1-E2	E3	E10-E12	final	yes
	Green monkey	*Chlorocebus sabeus*	24	hypothetical	E1-E2	E3	E10-E12	final	yes
	Proboscis monkey	*Nasalis larvatus*	ıd Un_JMHX01250125v1	hypothetical	E1-E2	E3	E10-E12	final	yes
	Marmoset	*Callithrix jacchus*	10	hypothetical	E1-E2	E3	E10-E11	final	yes
	Squirrel monkey	*Saimiri boliviensis*	JH378253	hypothetical	E1-E2	E3	E10-E12	final	yes
(Prosimians)	Bushbaby	*Otolemur garnettii*	GL873539	hypothetical	E1	E3	E12	final	yes
	Malayan lemur	*Galeopterus variegatus*	NW_007729566v1	hypothetical	E1-E2	E3	E10, E12	final	yes
	Mouse lemur	*Microcebus murinus*	KQ057910v1	hypothetical	N.I.	E3	E11-E12	final	yes
	Tarsier	*Tarsius syrichta*	2342v1 and KE935878v1	hypothetical	E2	E3	E12	final	yes
Rodentia									
	Mouse	*Mus musculus*	12	NR_003633.3	E1-E2	E3	E10-E12	seed	yes
	Rat	*Rattus norvegicus*	6	NR_131064.1	E1-E2	E3	E10-E12	final	yes
	Squirrel	*Spermophilus tridecemlineatus*	JH393414	hypothetical	E1-E2	E3	E10, E12	seed	yes
	Kangaroo rat	*Dipodomys ordii*	5495	hypothetical	N.I.	E3	E12	seed	yes
	Naked mole-rat	*Heterocephalus glaber*	JH602047	hypothetical	N.I.	E3	E12	final	yes
	Chinese hamster	*Cricetulus griseus*	KE377858	hypothetical	E1-E2	E3	E10, E12	final	yes
Eulipotyphla									
	Hedgehog	*Erinaceus europeus*	JH835375	hypothetical	N.I.	E3	E12	final	yes
	Shrew	*Sorex araneus*	JH798154	hypothetical	N.I.	E3	E11	final	
Chiroptera									
	Megabat	*Pteropus vampyrus*	9479	hypothetical	E1	E3	E12	seed	yes
	Microbat	*Myotis lucifugus*	)690 and AAPE02067072	hypothetical	E1-E2	E3	E12	seed	yes
Artiodactyla									
	Cow	*Bos taurus*	21	NR_037684.1	E1-E2	E3	E10-E12		
	Bison	*Bison bison*	KN265017v1	hypothetical	E1-E2	E3	E10-E12	final	
	Pig	*Sus scrofa*	7	NR_021488.1	E1-E2	E3	E10-E12	seed	yes
	Alpaca	*Vicugna pacos*	KB632524	hypothetical	E1-E2	E3	E12	final	yes
	Sheep	*Ovis aries*	18	hypothetical	E1-E2	E3	E10, E12	final	
	Minke whale	*Balaenoptera acutorostrata*	KI537651	hypothetical	E1-E2	E3	E10-E12	final	yes
	Dolphin	*Tursiops truncatus*	JH472808	hypothetical	E1-E2	E3	E10-E12	seed	yes
Perissodactyla									
	Horse	*Equus caballus*	24	hypothetical	E1-E2	E3	E12	seed	yes
	White rhinoceros	*Ceratotherium simum*	JH767771	hypothetical	E1-E2	E3	E10, E12	seed	yes
Pholidota									
	Chinese pangolin	*Manis pentadactyla*	KN007316	hypothetical	E1-E2	E3	E12	seed	yes
Carnivora									
	Dog	*Canis lupus familiaris*	8	hypothetical	E1-E2	E3	E10, E12	seed	yes
	Cat	*Felix catus*	B3	hypothetical	E1-E2	E3	E12	seed	yes
	Panda	*Ailuropoda melanoleuca*	GL192537.1	hypothetical	E1-E2	E3	E10, E12	seed	yes
	Ferret	*Mustela putorius furo*	GL896944	hypothetical	E1-E2	E3	E12	seed	yes
Xenarthra									
	Armadillo	*Dasypus novemcinctus*	JH578131	hypothetical	E1	E3	N.I.	final	yes
Afrotheria									
	Elephant	*Loxodonta africana*	9	hypothetical	E2	E3	E10, E12	seed	yes
	Manatee	*Trichechus manatus*	JH594651	hypothetical	E1-E2	E3	E10, E12	final	yes
	Tenrec	*Echinops telfairi*	JH980297	hypothetical	N.I.	E3	E12	final	yes
Marsupialia									
	Tasmanian devil	*Sarcophilus harrisii*	1_GL834715_random	hypothetical	N.I.	E3	N.I.		

Notes:

[1]: N.I. = not identified

[2]: seed = sequence included in the cmbuild seed alignment; final = sequence included by cmsearch in the final alignment

Appendix Table 5: List of vectors.

Plasmid name	gene	core
pTU1	MEG3 variant1	blue script core
pMM23	gIIi for cristalization (with deletations)	blue script core
pTU2	MEG3 without exon5 (variant 9)	blue script core
pTU3	MEG3 domain1 (2-196)	blue script core
pTU4	MEG3 domain2 (230-410)	blue script core
pTU5	MEG3 domain3 (471-902)	blue script core
pTU6	MEG3 domain4 (951-1113)	blue script core
pTU7	MEG3 domain5 (1116-1486)	blue script core
pTU8	MEG3 variant1	pCMS-d2EGFP-MEG3
pTU9	empty	pCMS-d2EGFP-MEG3
pcDNA3	empty	pcDNA3
pTU15	MEG3 variant 1	pcDNA3
pTU30	MEG3 without exon5 (variant 9)	pcDNA3
pcI-MEG3	MEG3 -first 24 nt -polyA	pcI
pTU31	MEG3 without exon5 (variant 9) -first 24 nt -polyA	pcI
pTU32	MEG3 variant 1	pcI
pTU33	MEG3 without exon5 (variant 9)	pcI
pTU34	MEG3 helix22/domain3(585-628) deleted	pcDNA3
pTU35	MEG3 helix23/domain3(630-678) deleted	pcDNA3
pTU36	MEG3 helix2/domain1(38-88) deleted	pcDNA3
pTU37	MEG3 helix11/domain2(351-390) deleted	pcDNA3
pTU38	MEG3 domain3(472-901) deleted	pcDNA3
pTU39	MEG3 domain1 (2-196)	pcDNA3
pTU40	MEG3 domain2 (230-410)	pcDNA3
pTU41	MEG3 domain3 (471-902)	pcDNA3
pTU42	MEG3 domain4 (951-1113)	pcDNA3
pTU43	MEG3 domain5 (1116-1486)	pcDNA3
pTU44	MEG3 domain 1-2-3 (1-902)	pcDNA3
pTU45	MEG3 exon 10-11-12	pcDNA3
pTU46	MEG3 deleted domain 1 (1-196)	pcDNA3
pTU47	MEG3 deleted domain 2 (230-410)	pcDNA3
pTU48	MEG3 deleted domain 4 (951-1113)	pcDNA3
pTU49	MEG3 deleted domain 5 (1116-1486)	pcDNA3
pTU50	MEG3 deleted helix8	pcDNA3
pTU51	MEG3 deleted junction 9 and helixes 28-29 (805-851)	pcDNA3
pTU52	MEG3 loop on top helix 11 poly A (365-8xA-374)	pcDNA3
pTU53	MEG3 deleted loop on top of helix11 (365-374)	pcDNA3
pTU54	MEG3 mutated one side of helix 11 (358-364 AGCATTG)	pcDNA3
pTU55	MEG3 mutated the other side of helix 11 to get complementary strand (375-	pcDNA3

	381)	
pTU56	MEG3 deleted ACG elbow (383-384-385)	pcDNA3
pTU57	MEG3 KL1 polyA 102-107	pcDNA3
pTU58	MEG3 KL2 polyA 213-218	pcDNA3
pTU59	MEG3 KL11 polyA 978-983	pcDNA3
pTU60	MEG3 KL13 polyA 1492-1497	pcDNA3
pTU61	MEG3 G368C	pcDNA3
pTU62	MEG3 U369A	pcDNA3
pTU63	MEG3 G370C	pcDNA3
pTU64	MEG3 A371U	pcDNA3
pTU65	MEG3 G372C	pcDNA3
pTU66	MEG3 G373C	pcDNA3
pTU67	MEG3 UA366AU	pcDNA3
pTU68	MEG3Δe10 (1050-1183)	pcDNA3
pTU69	MEG3Δe11 (1184-1259)	pcDNA3
pTU70	MEG3Δe12 (1260-END)	pcDNA3
pTU71	MEG3ΔH21 (569-691)	pcDNA3
pTU72	MEG3ΔH25 (748-891)	pcDNA3
pTU73	MEG3ΔH27' (792-862)	pcDNA3
pTU74	MEG3ΔH28 (809-822)	pcDNA3
pTU75	MEG3ΔH29 (828-846)	pcDNA3
pTU76	MEG3 exon 3	pcDNA3
pTU77	MEG3 helix11/domain2(351-390) deleted	blue script core
pTU79	MEG3 KL9(TR5) polyA 875-880	pcDNA3
pTU80	MEG3 KL7(TR3) polyA 866-871	pcDNA3
pTU81	MEG3ΔH16 (487-744)	pcDNA3
pTU83	MEG3 KL3 poly A (338-343)	pcDNA3
pTU84	MEG3 KL4 poly A (529-534)	pcDNA3
pTU85	MEG3 KL12 poly A (1483-1488)	pcDNA3
pTU93	MEG3 domain2 and 3 (230-902)	pcDNA3
pTU94	MEG3ΔH10 (307-340)	pcDNA3
pTU95	MEG3ΔH17 (498-517)	pcDNA3
pTU103	MEG3 G368A	pcDNA3
pTU104	MEG3 U369G	pcDNA3
pTU105	MEG3 G370A	pcDNA3
pTU106	MEG3 A371G	pcDNA3
pTU107	MEG3 G372A	pcDNA3
pTU108	MEG3 G373A	pcDNA3
pTU109	MEG3 G368U	pcDNA3
pTU110	MEG3 U369C	pcDNA3
pTU111	MEG3 G370U	pcDNA3
pTU112	MEG3 A371C	pcDNA3

pTU113	MEG3 G372U	pcDNA3
pTU114	MEG3 G373U	pcDNA3
pTU115	MEG3 G368-	pcDNA3
pTU116	MEG3 U369-	pcDNA3
pTU117	MEG3 G370-	pcDNA3
pTU118	MEG3 A371-	pcDNA3
pTU119	MEG3 G372-	pcDNA3
pTU120	MEG3 G373-	pcDNA3
pTU121	MEG3 KL10 (TR6) polyA 880-885	pcDNA3
pTU123	MEG3 loop on top helix 11 poly A (365-8xA-374)	blue script core
pTU133	MEG3 KL5 (TR1) polyA 857-862	pcDNA3
pTU134	MEG3 KL6 (TR2) polyA 862-866	pcDNA3
pTU135	MEG3 KL8 (TR4) polyA 871-875	pcDNA3
pTU142	MEG3 G370C U860G	pcDNA3
pTU143	MEG3 G370C U864G	pcDNA3
pTU144	MEG3 G370C U869G	pcDNA3
pTU145	MEG3 G370C U873G	pcDNA3
pTU146	MEG3 G370C U878G	pcDNA3
pTU147	MEG3 G370C U882G	pcDNA3
pTU148	MEG3 deleted domain3	blue script core
pTU149	MEG3 domain 2-3	blue script core
pTU150	MEG3 G370C U860, 864,869,873,878,882G	pcDNA3
pTU151	MEG3 deleted domain 2 (230-410)	blue script core
pTU158	MEG3 G370C U860, 864,869,873,878,882G A 784 794 C	pcDNA3
pTU159	MEG3 G370C U 864, 869, 878, 882 G	pcDNA3
pTU160	MEG3 G370C U 860, 869 G	pcDNA3
pTU161	MEG3 G370C U 864, 869 G	pcDNA3
pTU162	MEG3 G370C U 869, 873 G	pcDNA3
pTU163	MEG3 G370C U 869, 878 G	pcDNA3
pTU164	MEG3 G370C U 869, 882 G	pcDNA3
pTU165	MEG3 A371U U859A	pcDNA3
pTU166	MEG3 A371U U863A	pcDNA3
pTU167	MEG3 A371U U868A	pcDNA3
pTU168	MEG3 A371U U872A	pcDNA3
pTU169	MEG3 A371U U877A	pcDNA3
pTU170	MEG3 A371U U881A	pcDNA3
p53-luc	Luciferase reporter with 14 copies of the p53-binding consensus sequence	
PG13luc	Luciferase reporter with 13 copies of the p53-binding consensus sequence	pBluescript II SK + luc
pGL-p21	Luciferase reporter containing p53-binding consensus sequence for p21	
pGL-MDM2	Luciferase reporter containing p53-binding consensus sequence for MDM2	
pRL-TK Renilla	renilla luciferase (Luciferase Control Reporter Vectors)	
pcI-p14	p14	pcI

Appendix Table 6: List of primers (excluding cloning primers).

Name	sequence	description
Meg3RT1	CCTTGAAGACAAGGAGGTGG	primers for SHAPE, 5'amino C6 modified
Meg3RT22	CTGGCTGGTCAGTTCCGGTC	primers for SHAPE, 5'amino C6 modified
Meg3RT3	CCCAAAGGGATCCTTCCATTCAGG	primers for SHAPE, 5'amino C6 modified
Meg3RT42	CTATGCCAGATCCTGCCTGAGGC	primers for SHAPE, 5'amino C6 modified
Meg3RT52	CAGGCCTTTCAAGAAGCTTGGC	primers for SHAPE, 5'amino C6 modified
Meg3RT6	CCAGGAAGGAGACGAGAGGC	primers for SHAPE, 5'amino C6 modified
Meg3RT72	CGTCCATCAGTCAGAGGGCG	primers for SHAPE, 5'amino C6 modified
Meg3RT8	CAGGAAACACATTTATTGAGAGCACAG	primers for SHAPE, 5'amino C6 modified
1f_V_MEG3b	CGGAGAGCAGAGAGGG	forward primer for *in-vivo* SHAPE of MEG3 1(19-34)
1r_V_MEG3b	GGGTGATGACAGAGTCAGTC	revers primer for *in-vivo* SHAPE of MEG3 1(511-530)
2f_V_MEG3	CCTGACCTTTGCTATGCTC	forward primer for *in-vivo* SHAPE of MEG3 2 (392-410)
2r_V_MEG3b	CTGATGCAAGGAGAGCC	revers primer for *in-vivo* SHAPE of MEG3 2 (919-935)
3f_V_MEG3	CAGGATCTGGCATAGAGGAG	forward primer for *in-vivo* SHAPE of MEG3 3 (783-802)
3r_V_MEG3b	GAATAGGTGCAGGGTGTC	revers primer for *in-vivo* SHAPE of MEG3 3(1276-1293)
4f_V_MEG3	CCTCTCGTCTCCTTCCTG	forward primer for *in-vivo* SHAPE of MEG3 4 (1159-1176)
4r_V_MEG3	CAGGAAACACATTTATTGAGAGC	revers primer for *in-vivo* SHAPE of MEG3 4 (1561-1583)
13F_112_D1Ex2	GTCTCTCCTCAGGGATGAC	forvard primer for RT-PCR amplifying D1(Ex2)
14R_191_D1Ex2	TTGGCAGCAGCTCAGCA	reverse primer for RT-PCR amplifying D1(Ex2)
15F_230_D2Ex3	GAGCACGGTTTCCTGGAT	forvard primer for RT-PCR amplifying D2(Ex3)
Meg3RT22	CTGGCTGGTCAGTTCCGGTC	reverse primer for RT-PCR amplifying D2(Ex3)
01F_MEG3_Ex3	TCGATGAGAGCAACCTCCTA	forvard primer for RT-PCR amplifying D3(Ex3)
02R_MEG3_Ex3	TGCTGATCACCTCCTCTATG	reverse primer for RT-PCR amplifying D3(Ex3)
08F_951_Ex5	GGCCTGTCTACACTTGCTG	forvard primer for RT-PCR amplifying D4(Ex5)
05R_1029_D4	GGAGTAGAGCGAGTCAGGAA	reverse primer for RT-PCR amplifying D4(Ex5)
09F_1077_Ex10	GGCTGAAGAACTGCGGAT	forvard primer for RT-PCR amplifying D5(Ex10)
10R_1179_Ex10	AACCAGGAAGGAGACGAGAG	reverse primer for RT-PCR amplifying D5(Ex10)
pcDNA3-F-Neo	TGGATTGCACGCAGGTTCT	forvard primer for RT-PCR amplifying part of neomycine
pcDNA3-R1-Neo	GGACAGGTCGGTCTTGACA	reverse primer for RT-PCR amplifying part of neomycine
ACTB_Fwd	TTCCAGCAGATGTGGATCAG	forvard primer for RT-PCR amplifying part of actin
ACTB_Rev	GGTGTAACGCAACTAAGTCA	reverse primer for RT-PCR amplifying part of actin

References

Andrade, M.A., Perez-Iratxeta, C., and Ponting, C.P. (2001). Protein repeats: structures, functions, and evolution. Journal of structural biology *134*, 117-131.

Aw, J.G., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.L., Tapsin, S., Chan, Y.S., Tan, C.P., Sim, A.Y*., et al.* (2016). In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. Molecular cell *62*, 603-617.

Belyi, V.A., Ak, P., Markert, E., Wang, H., Hu, W., Puzio-Kuter, A., and Levine, A.J. (2010). The origins and evolution of the p53 family of genes. Cold Spring Harbor perspectives in biology *2*, a001198.

Benetatos, L., Vartholomatos, G., and Hatzimichael, E. (2011). MEG3 imprinted gene contribution in tumorigenesis. International journal of cancer Journal international du cancer *129*, 773-779.

Blythe, A.J., Fox, A.H., and Bond, C.S. (2016). The ins and outs of lncRNA structure: How, why and what comes next? Biochimica et biophysica acta *1859*, 46-58.

Bongyong Lee, A.S., John Marchica, Erwin Holzhauser, Xiaoli Chen,, Jian-Liang Li, T.S., Subramaniam Shyamala Govindarajan,Fatu Badiane Markey,, and Mona Batish, S.J.L., Shaojie Zhang, Animesh Ray, Ranjan J. Perera (2017). The long noncoding RNA SPRIGHTLY acts as an intranuclear organizing hub for pre-mRNA molecules. Science Advances *3: e1602505*.

Braconi, C., Kogure, T., Valeri, N., Huang, N., Nuovo, G., Costinean, S., Negrini, M., Miotto, E., Croce, C.M., and Patel, T. (2011). microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. Oncogene *30*, 4750-4756.

Brosnan, C.A., and Voinnet, O. (2009). The long and the short of noncoding RNAs. Current opinion in cell biology *21*, 416-425.

Brown, J.A., Bulkley, D., Wang, J., Valenstein, M.L., Yario, T.A., Steitz, T.A., and Steitz, J.A. (2014). Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nature structural & molecular biology *21*, 633-640.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. Genome biology *16*, 20.

Calo, A., Stoliar, P., Bystrenova, E., Valle, F., and Biscarini, F. (2009). Measurement of DNA morphological parameters at highly entangled regime on surfaces. The journal of physical chemistry B *113*, 4987-4990.

Cao, J. (2014). The functional role of long non-coding RNAs and epigenetics. Biological procedures online *16*, 11.

Carr, M.S., Yevtodiyenko, A., Schmidt, C.L., and Schmidt, J.V. (2007). Allele-specific histone modifications regulate expression of the Dlk1-Gtl2 imprinted domain. Genomics *89*, 280-290.

Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Proteins by RNA Interactome Capture. Methods in molecular biology *1358*, 131-139.

Chaudhary, R., and Lal, A. (2016). Long noncoding RNAs in the p53 network. Wiley interdisciplinary reviews RNA.

Chen, S.W., and Pellequer, J.L. (2011). DeStripe: frequency-based algorithm for removing stripe noises from AFM images. BMC structural biology *11*, 7.

Chen, Y., and Pollack, L. (2016). SAXS studies of RNA: structures, dynamics, and interactions with partners. Wiley interdisciplinary reviews RNA *7*, 512-526.

Cheunsuchon, P., Zhou, Y., Zhang, X., Lee, H., Chen, W., Nakayama, Y., Rice, K.A., Tessa Hedley-Whyte, E., Swearingen, B., and Klibanski, A. (2011). Silencing of the imprinted DLK1-MEG3 locus in human clinically nonfunctioning pituitary adenomas. The American journal of pathology *179*, 2120-2130.

Chillon, I., Marcia, M., Legiewicz, M., Liu, F., Somarowthu, S., and Pyle, A.M. (2015). Native Purification and Analysis of Long RNAs. Methods in enzymology *558*, 3-37.

Chillon, I., and Pyle, A.M. (2016). Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. Nucleic acids research *44*, 9462-9471.

Chunharojrith, P., Nakayama, Y., Jiang, X., Kery, R.E., Ma, J., De La Hoz Ulloa, C.S., Zhang, X., Zhou, Y., and Klibanski, A. (2015). Tumor suppression by MEG3 lncRNA in a human pituitary tumor derived cell line. Molecular and cellular endocrinology *416*, 27-35.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*, e00005.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics *25*, 1974-1975.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nature structural & molecular biology *20*, 1250-1257.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. Proceedings of the National Academy of Sciences of the United States of America *106*, 97-102.

Delageniere, S., Brenchereau, P., Launer, L., Ashton, A.W., Leal, R., Veyrier, S., Gabadinho, J., Gordon, E.J., Jones, S.D., Levik, K.E*., et al.* (2011). ISPyB: an information management system for synchrotron macromolecular crystallography. Bioinformatics *27*, 3186-3192.

Dhir, A., Buratti, E., van Santen, M.A., Luhrmann, R., and Baralle, F.E. (2010). The intronic splicing code: multiple factors involved in ATM pseudoexon definition. The EMBO journal *29*, 749-760.

Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. Trends in genetics : TIG *30*, 121-123.

Draper, D.E. (2004). A guide to ions and RNA structure. Rna *10*, 335-343.

Draper, D.E. (2013). Folding of RNA tertiary structure: Linkages between backbone phosphates, ions, and water. Biopolymers *99*, 1105-1113.

Duszczyk, M.M., Wutz, A., Rybin, V., and Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. Rna *17*, 1973-1982.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., Forman, D., and Bray, F. (2013). Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. European journal of cancer *49*, 1374-1403.

Giro, A., Bergia, A., Zuccheri, G., Bink, H.H., Pleij, C.W., and Samori, B. (2004). Single molecule studies of RNA secondary structure: AFM of TYMV viral RNA. Microscopy research and technique *65*, 235-245.

Gordon, F.E., Nutt, C.L., Cheunsuchon, P., Nakayama, Y., Provencher, K.A., Rice, K.A., Zhou, Y., Zhang, X., and Klibanski, A. (2010). Increased expression of angiogenic genes in the brains of mouse meg3-null embryos. Endocrinology *151*, 2443-2452.

Hakem, R., and Mak, T.W. (2001). Animal models of tumor-suppressor genes. Annual Review of Genetics, 2019-2241.

Hansma, H.G., Revenko, I., Kim, K., and Laney, D.E. (1996). Atomic force microscopy of long and short double-stranded, single-stranded and triple-stranded nucleic acids. Nucleic acids research *24*, 713-720.

Harris, M.E., and Christian, E.L. (2009). RNA crosslinking methods. Methods in enzymology *468*, 127-146.

Hawkes, E.J., Hennelly, S.P., Novikova, I.V., Irwin, J.A., Dean, C., and Sanbonmatsu, K.Y. (2016). COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. Cell reports *16*, 3087-3096.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nature protocols *2*, 1849-1861.

Higuchi, T. (1988). Approach to an irregular time-series on the basis of the fractal theory. Physica D *31*, 277-283.

Hudson, W.H., Pickard, M.R., de Vera, I.M., Kuiper, E.G., Mourtada-Maarabouni, M., Conn, G.L., Kojetin, D.J., Williams, G.T., and Ortlund, E.A. (2014). Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate. Nature communications *5*, 5395.

Huynen, M., Gutell, R., and Konings, D. (1997). Assessing the reliability of RNA folding using statistical mechanics. Journal of molecular biology *267*, 1104-1112.

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R*., et al.* (2013). Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. Molecular cell *51*, 156-173.

Jaeger, L., Michel, F., and Westhof, E. (1994). Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. Journal of molecular biology *236*, 1271-1276.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383.

Johnson, R. (2012). Long non-coding RNAs in Huntington's disease neurodegeneration. Neurobiology of disease *46*, 245-254.

Kameswaran, V., and Kaestner, K.H. (2014). The Missing lnc(RNA) between the pancreatic beta-cell and diabetes. Frontiers in genetics *5*, 200.

Kaneko, S., Bonasio, R., Saldana-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Molecular cell *53*, 290-300.

Karabiber, F., McGinnis, J.L., Favorov, O.V., and Weeks, K.M. (2013). QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. Rna *19*, 63-73.

Kaushik, K., Leonard, V.E., Kv, S., Lalwani, M.K., Jalali, S., Patowary, A., Joshi, A., Scaria, V., and Sivasubbu, S. (2013). Dynamic expression of long non-coding RNAs (lncRNAs) in adult zebrafish. PloS one *8*, e83616.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome research *12*, 656-664.

Leeper, T.C., and Varani, G. (2005). The structure of an enzyme-activating fragment of human telomerase RNA. Rna *11*, 394-403.

Li, M.Z., and Elledge, S.J. (2012). SLIC: a method for sequence- and ligation-independent cloning. Methods in molecular biology *852*, 51-59.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N., and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic acids research *43*, W580-584.

Lin, Y., Schmidt, B.F., Bruchez, M.P., and McManus, C.J. (2018). Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. Nucleic acids research *46*, 3742-3752.

Liu, F., Somarowthu, S., and Pyle, A.M. (2017). Visualizing the secondary and tertiary architectural domains of lncRNA RepA. Nature chemical biology *13*, 282-289.

Liu, S., Zhu, J., Jiang, T., Zhong, Y., Tie, Y., Wu, Y., Zheng, X., Jin, Y., and Fu, H. (2015). Identification of lncRNA MEG3 Binding Protein Using MS2-Tagged RNA Affinity Purification and Mass Spectrometry. Applied biochemistry and biotechnology.

Lu, K.H., Li, W., Liu, X.H., Sun, M., Zhang, M.L., Wu, W.Q., Xie, W.P., and Hou, Y.Y. (2013). Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. BMC cancer *13*, 461.

Lyubchenko, Y.L., Shlyakhtenko, L.S., and Ando, T. (2011). Imaging of nucleic acids with atomic force microscopy. Methods *54*, 274-283.

Marcia, M., and Pyle, A.M. (2012). Visualizing group II intron catalysis through the stages of splicing. Cell *151*, 497-507.

Mathews, D.H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. Rna *10*, 1178-1190.

Menendez, D., Inga, A., and Resnick, M.A. (2010). Estrogen receptor acting in cis enhances WT and mutant p53 transactivation at canonical and noncanonical p53 target sequences. Proceedings of the National Academy of Sciences of the United States of America *107*, 1500-1505.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nature reviews Genetics *10*, 155-159.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proceedings of the National Academy of Sciences of the United States of America *105*, 716-721.

Miyoshi, N., Wagatsuma, H., Wakana, S., Shiroishi, T., Nomura, M., Aisaka, K., Kohda, T., Surani, M.A., Kaneko-Ishino, T., and Ishino, F. (2000). Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. Genes to cells : devoted to molecular & cellular mechanisms *5*, 211-220.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E.*, et al.* (2015). MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. Nature communications *6*, 7743.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. Journal of immunological methods *65*, 55-63.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J.*, et al.* (2015). Rfam 12.0: updates to the RNA families database. Nucleic acids research *43*, D130-137.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933-2935.

Nečas, D., and Klapetek, P. (2012). Gwyddion: an open-source software for SPM data analysis. Central European Journal of Physics *10*, 181-188.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grutzner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635-640.

Nguyen, T.C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F.H., Sridhar, B., Huang, N., Zhang, K., and Zhong, S. (2016). Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. Nature communications *7*, 12023.

Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic acids research *40*, 5034-5051.

Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Frontiers in genetics *6*, 2.

Patel, T.R., Chojnowski, G., Astha, Koul, A., McKenna, S.A., and Bujnicki, J.M. (2016). Structural studies of RNA-protein complexes: A hybrid approach involving hydrodynamics, scattering, and computational methods. Methods.

PDBconsortium (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic acids research *47*, D520-D528.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. Journal of applied crystallography *45*, 342-350.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research *29*, e45.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nature structural & molecular biology *25*, 154-162.

Pyle, A.M. (2002). Metal ions in the structure and function of RNA. Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry *7*, 679-690.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC bioinformatics *11*, 129.

Rice, G.M., Busan, S., Karabiber, F., Favorov, O.V., and Weeks, K.M. (2014a). SHAPE analysis of small RNAs and riboswitches. Methods in enzymology *549*, 165-187.

Rice, G.M., Leonard, C.W., and Weeks, K.M. (2014b). RNA secondary structure modeling at consistent high accuracy using differential SHAPE. Rna *20*, 846-854.

Rivas, E., Clements, J., and Eddy, S.R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nature methods *14*, 45-48.

Ryschon, T.W., Rosenstein, D.L., Rubinow, D.R., Niemela, J.E., Elin, R.J., and Balaban, R.S. (1996). Relationship between skeletal muscle intracellular ionized magnesium and measurements of blood magnesium. J Lab Clin Med *127*, 207-213.

Salic, A., and Mitchison, T.J. (2008). A chemical method for fast and sensitive detection of DNA synthesis in vivo. Proceedings of the National Academy of Sciences of the United States of America *105*, 2415-2420.

Sanbonmatsu, K.Y. (2016). Towards structural classification of long non-coding RNAs. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1859*, 41-45.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M.*, et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife *2*, e01749.

Schon, P. (2016). Imaging and force probing RNA by atomic force microscopy. Methods *103*, 25-33.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophysical journal *78*, 1606-1619.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global Mapping of Human RNA-RNA Interactions. Molecular cell *62*, 618-626.

Sherpa, C., Rausch, J.W., and Le Grice, S.F.J. (2018). Structural characterization of maternally expressed gene 3 RNA reveals conserved motifs and potential sites of interaction with polycomb repressive complex 2. Nucleic acids research.

Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A., and Weeks, K.M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nature methods *11*, 959-965.

Smola, M.J., Calabrese, J.M., and Weeks, K.M. (2015a). Detection of RNA-Protein Interactions in Living Cells with SHAPE. Biochemistry *54*, 6867-6875.

Smola, M.J., Christy, T.W., Inoue, K., Nicholson, C.O., Friedersdorf, M., Keene, J.D., Lee, D.M., Calabrese, J.M., and Weeks, K.M. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. Proceedings of the National Academy of Sciences of the United States of America *113*, 10322-10327.

Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., and Weeks, K.M. (2015b). Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. Nature protocols *10*, 1643-1669.

Somarowthu, S., Legiewicz, M., Chillon, I., Marcia, M., Liu, F., and Pyle, A.M. (2015). HOTAIR forms an intricate and modular secondary structure. Molecular cell *58*, 353-361.

Spiniello, M., Knoener, R.A., Steinbrink, M.I., Yang, B., Cesnik, A.J., Buxton, K.E., Scalf, M., Jarrard, D.F., and Smith, L.M. (2018). HyPR-MS for Multiplexed Discovery of MALAT1, NEAT1, and NORAD lncRNA Protein Interactomes. Journal of proteome research *17*, 3022-3038.

Su, L.J., Brenowitz, M., and Pyle, A.M. (2003). An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. Journal of molecular biology *334*, 639-652.

Su, W., Xie, W., Shang, Q., and Su, B. (2015). The Long Noncoding RNA MEG3 Is Downregulated and Inversely Associated with VEGF Levels in Osteoarthritis. BioMed research international *2015*, 356893.

Takahashi, N., Okamoto, A., Kobayashi, R., Shirai, M., Obata, Y., Ogawa, H., Sotomaru, Y., and Kono, T. (2009). Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. Human molecular genetics *18*, 1879-1888.

Tarver, J.E., Dos Reis, M., Mirarab, S., Moran, R.J., Parker, S., O'Reilly, J.E., King, B.L., O'Connell, M.J., Asher, R.J., Warnow, T.*, et al.* (2016). The Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. Genome biology and evolution *8*, 330-344.

Tavares, R.C.A., Pyle, A.M., and Somarowthu, S. (2018). Covariation analysis with improved parameters reveals conservation in lncRNA structures. bioRxiv.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391-403 e319.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science *329*, 689-693.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26-46.

Volders, P.J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., Vandesompele, J., and Mestdagh, P. (2013). LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic acids research *41*, D246-251.

Wang, C., Yan, G., Zhang, Y., Jia, X., and Bu, P. (2015). Long non-coding RNA MEG3 suppresses migration and invasion of thyroid carcinoma by targeting of Rac1. Neoplasma *62*, 541-549.

Wapinski, O., and Chang, H.Y. (2011). Long noncoding RNAs and human disease. Trends in cell biology *21*, 354-361.

Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H., and Lucks, J.B. (2016). Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Methods *103*, 34-48.

Weinberg, R. (1991). Tumor suppressor genes. Science, 1138–1146.

Weinberg, Z., and Breaker, R.R. (2011). R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. BMC bioinformatics *12*, 3.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nature protocols *1*, 1610-1616.

Woodson, S.A. (2005). Metal ions and RNA folding: a highly charged topic with a dynamic future. Current opinion in chemical biology *9*, 104-109.

Xia, Y., He, Z., Liu, B., Wang, P., and Chen, Y. (2015). Downregulation of Meg3 enhances cisplatin resistance of lung cancer cells through activation of the WNT/beta-catenin signaling pathway. Molecular medicine reports *12*, 4530-4537.

Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M. (2015). RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. Nature *517*, 33-38.

Xue, Z., Hennelly, S., Doyle, B., Gulati, A.A., Novikova, I.V., Sanbonmatsu, K.Y., and Boyer, L.A. (2016). A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. Molecular cell *64*, 37-50.

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. Molecular cell *70*, 1038-1053 e1037.

Yan, J., Guo, X., Xia, J., Shan, T., Gu, C., Liang, Z., Zhao, W., and Jin, S. (2014). MiR-148a regulates MEG3 in gastric cancer by targeting DNA methyltransferase 1. Medical oncology *31*, 879.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L.*, et al.* (2016). Ensembl 2016. Nucleic acids research *44*, D710-716.

Zampetaki, A., Albrecht, A., and Steinhofel, K. (2018). Long Non-coding RNA Structure and Function: Is There a Link? Frontiers in physiology *9*, 1201.

Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y., and Klibanski, A. (2010a). Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. Endocrinology *151*, 939-947.

Zhang, X., Zhou, Y., and Klibanski, A. (2010b). Isolation and characterization of novel pituitary tumor related genes: a cDNA representational difference approach. Molecular and cellular endocrinology *326*, 40-47.

Zhang, X., Zhou, Y., Mehta, K.R., Danila, D.C., Scolavino, S., Johnson, S.R., and Klibanski, A. (2003). A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. The Journal of clinical endocrinology and metabolism *88*, 5119-5126.

Zhao, J., Zhang, X., Zhou, Y., Ansell, P.J., and Klibanski, A. (2006). Cyclic AMP stimulates MEG3 gene expression in cells through a cAMP-response element (CRE) in the MEG3 proximal promoter region. The international journal of biochemistry & cell biology *38*, 1808-1820.

Zhou, Y., Cheunsuchon, P., Nakayama, Y., Lawlor, M.W., Zhong, Y., Rice, K.A., Zhang, L., Zhang, X., Gordon, F.E., Lidov, H.G.*, et al.* (2010). Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. Development *137*, 2643-2652.

Zhou, Y., Zhang, X., and Klibanski, A. (2012). MEG3 noncoding RNA: a tumor suppressor. J Mol Endocrinol *48*, R45-53.

Zhou, Y., Zhong, Y., Wang, Y., Zhang, X., Batista, D.L., Gejman, R., Ansell, P.J., Zhao, J., Weng, C., and Klibanski, A. (2007). Activation of p53 by MEG3 non-coding RNA. The Journal of biological chemistry *282*, 24731-24742.

Zhu, J., Liu, S., Ye, F., Shen, Y., Tie, Y., Zhu, J., Wei, L., Jin, Y., Fu, H., Wu, Y.*, et al.* (2015). Long Noncoding RNA MEG3 Interacts with p53 Protein and Regulates Partial p53 Target Genes in Hepatoma Cells. PloS one *10*, e0139790.