Luis Filipe Santos Pereira Fernandes da Silva

Licenciado em Bioquimica

Shedding light on miRNA targeting through structure

Dissertação para obtenção do Grau de Mestre em

Bioquimica

Orientador: Doutora Katja Petzold, lider de grupo, Karolinska Institutet

Co-orientador: Lorenzo Baronti, PhD, Karolinska Institutet

Outubro 2017

Luis Filipe Santos Pereira Fernandes da Silva

Licenciado em Bioquimica

Shedding light on miRNA targeting through structure

Dissertação para obtenção do Grau de Mestre em

Bioquimica

Orientador: Doutora Katja Petzold, lider de grupo, Karolinska Institutet

Co-orientador: Lorenzo Baronti, PhD, Karolinska Institutet

Outubro 2017

Shedding light on miRNA targeting through structure

Acknowledgements

I dedicate this thesis to my parents, sisters, grandparents and Ai-Ling. Without you I would never be here and wouldn't be able to do this work. All I am is thanks to you, the achievements I have made happened because of your helping hand, time and time again. I will not forget. Thank you all.

Dr. Katja Petzold, my supervisor, thank you for the opportunities you gave me on your lab throughout the year, for all the space you gave me to let me grow and the trust you put on me. Specially thank you for the challenges and guidance.

Dr. Emma Andresson, for the extraordinary help on the collaborating project with the Petzold group and great brainstorming moments where ideas collide and are brought to life.

All my lab colleagues, an enormous thank you for all the support, great time spent together laughing and learning with each other. Emelie Steiner for the deep talks, the most out of the blue funny moments and patience with my NMR questions. Judith Schlagnitweit, for all the discussions about ideas, spending the time with me unveiling NMR and the great time at fika. Ileana Guzzetti, for the super intense yet super incredible and mega exciting time spent together brainstorming, having deep fundamental talks about the principles of our research (super exciting) and understanding personality. Hampus Karlsson, for being an amazing pratical teacher with an eye for detail, for being a calm mind ready to ensure things are done at their natural pace and the remarkable well placed dark humour jokes. Lorenzo Baronti, for being there when the work tasks are heavy on just ones shoulders, for helping me with everyday theoretical and practical questions, guiding me when lost and for the great moments spent on the pub. Luca Retattino, for being a great remarkable happy go jolly friend being always there on good and bad times, for the incredible mega fun personality which shines a great light on dark days and for all the help on the lab when I had practical questions. Hannes Freyrer, for the great support both outside and inside the lab, the great discussions about work practicality, discussions on the field of NMR and RNAs and all the good times spent with common mates. Noah Hopkins, for helping us all with programming questions. Andrea Coulthard, Maja Marušič and Sarah Sandoz, for the time spent discussing key aspects of work and sparking a good mood in the group.

To all of my friends, in Stockholm and Portugal, for being there at all times. Thank you Rafael Rippel, António Lopez, Diogo Gonçalves, Bruno Guerreiro, João Ramos, Mauro Duarte, Dário Valezim, Miguel Correia, João Velez, for being who you are, awesome friends, which I will never ever Forget. You are all in my memory for ever, to more moments together in the future, University wouldn't be the same without you all. Thank you Jonhy, Malfalda castro, Filipa Bouvança, Astronomer (Astro Boy), Daniela, Carolina, Nuno and Matina for the exciting adventure we had and will have in Stockholm, the deep conversations and pure sense of friendship, without you all Stockholm would have been a very different place.

Resumo

Nos miRNAs, sabe-se que o emparelhamento da “seed sequence” é um parâmetro importante para a seletividade, mas ainda existem conexões inexploradas sobre o fato de atingir um determinado mRNA e como a decisão entre degradação ou repressão da tradução é feita.

Neste trabalho, usamos espectroscopia de RMN de estado líquido, EMSA e experiências de medição de temperatura de fusão com espectroscopia UV para caracterizar biofisicamente o complexo de miRNA-34a com o mRNA alvo, CD44 e PNUTS. Aqui apresentamos a preparação da amostra de RMN, bem como os primeiros passos para a descrição biofísica dos complexos mRNA.miRNA.

Os dados de EMSA sugerem que ambos os complexos de miRNA.mRNA são transientes e a possibilidade de miRNA.gCD44 possui um Kd maior do que o miRNA.PNUTS. Através de RMN, o complexo miRNA.PNUTS foi considerado mais estruturalmente dinâmico, mas as medições de temperatura de fusão, sugerem um complexo com Tm mais elevado que miRNA.gCD44. Finalmente, realizamos várias simulações em vários alvos mRNA diferentes com o MCFold. Estes dados revelam uma tendência estrutural, apresentando uma dobra no mRNA, o que poderá ser uma necessidade para encaixar o duplex na proteína Ago.

Mais estudos devem ser realizados para resolver as estruturas de ambos os complexos, e uma maior população de estruturas simuladas deve ser coletada.

Palavras-chave: miRNA, mRNA, “seed sequence”, transiente, simulação, EMSA, RMN, Tm.

Abstract

In miRNAs, it is known that seed pairing is an important parameter for targeting, but there are still missing links on why it targets a certain mRNA and how the decision between degradation or translation repression is made.

In our work, we use liquid state NMR spectroscopy, EMSA and UV melting experiments to biophysically characterise the complex of miRNA-34a with the target mRNA of CD44 and PNUTS. Here we present the NMR sample preparation as well as the first steps for the biophysical description of the mRNA.miRNA complexes.

Our EMSA data suggests both miRNA.mRNA complexes are transient and miRNA.gCD44 possibility has a higher Kd then miRNA.PNUTS. Through NMR miRNA.PNUTS complex was deemed more structurally dynamic, yet interestingly with UV melting it had a higher Tm then miRNA.gCD44 complex. Finally, we also performed several simulations on several different mRNA targets with MCFold unveiling a kink on the mRNA as a structural trend which could be a necessity for docking of the duplex into Ago.

Still further studies must be conducted in order to solve the structures of both duplexes in study and a bigger population of simulated structures must be collected.

Keywords: miRNA, mRNA, “seed sequence”, transient, simulation, EMSA, RMN, Tm.

Table of contents

List of figures

List of tables

List of figures and tables

List of abbreviations

RNA – Ribonucleic Acid

A – Adenosine

U – Uridine

G – Guanidine

C – Cytosine

DNA – Deoxyribonucleic Acid

WC – Watson Crick

Wb – Wooble

Mm – Mismatch

H-bond – Hydrogen bond

UV – Ultra-Violet

NMR – Nuclear Magnetic Resonance

miRNA – microRNA

mRNA – messenger RNA

RISC – RNA-induced Silencing Complex

Ago – Argonaute

ORF – Open Reading Frame

UTR – Untranslated Region

NCM – Nucleotide cyclic motif

GMP – Guanosine Monophosphate

GTP- Guanosine Triphosphate

DMSO - Dimethyl Sulfoxide

ATP – Adenosine Triphosphate

CTP – Cytosine Triphosphate

UTP – Uridine Triphosphate

GTP – Guanosine Triphosphate

HPLC – High Performance Liquid Chromatography

Nt – Nucleotide

IE – Ion Exchange

RP – Reverse Phase

Abs – Absorbance

Sp – Spermidine

DTT – Dithiothreitol

DDG - ΔΔG

1) Introduction

1-1) RNA structure

Ribonucleic Acid (RNA) is a polymer of ribonucleotide mono-phosphate units. A nucleotide consists of a pentose sugar in its β-furanose form, a nitrogenous base and one phosphate group. The base can be a purine: adenine (A) or guanine (G); or a pyrimidine: uridine (U) or cytosine (C). The base, at its N9 or N1 position is covalently bonded to the C-1' of the sugar, depending if it's a purine or pyrimidine respectively (fig. 1 a&b, adapted from IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) Definitive rules for nomenclature of organic chemistry, J. Am. Chem. Soc. 82, 5545-5574 (1960)). The backbone consists of ribose sugars linked by mono-phosphate groups, via a phosphodiester bond. This bond is mediated by the phosphate bonded with the O-5' of the sugar at the 5' end of the helix direction, and the O-3' of the sugar at the 3' end (Stryer, L. et al., 2000, pp. 271-276).

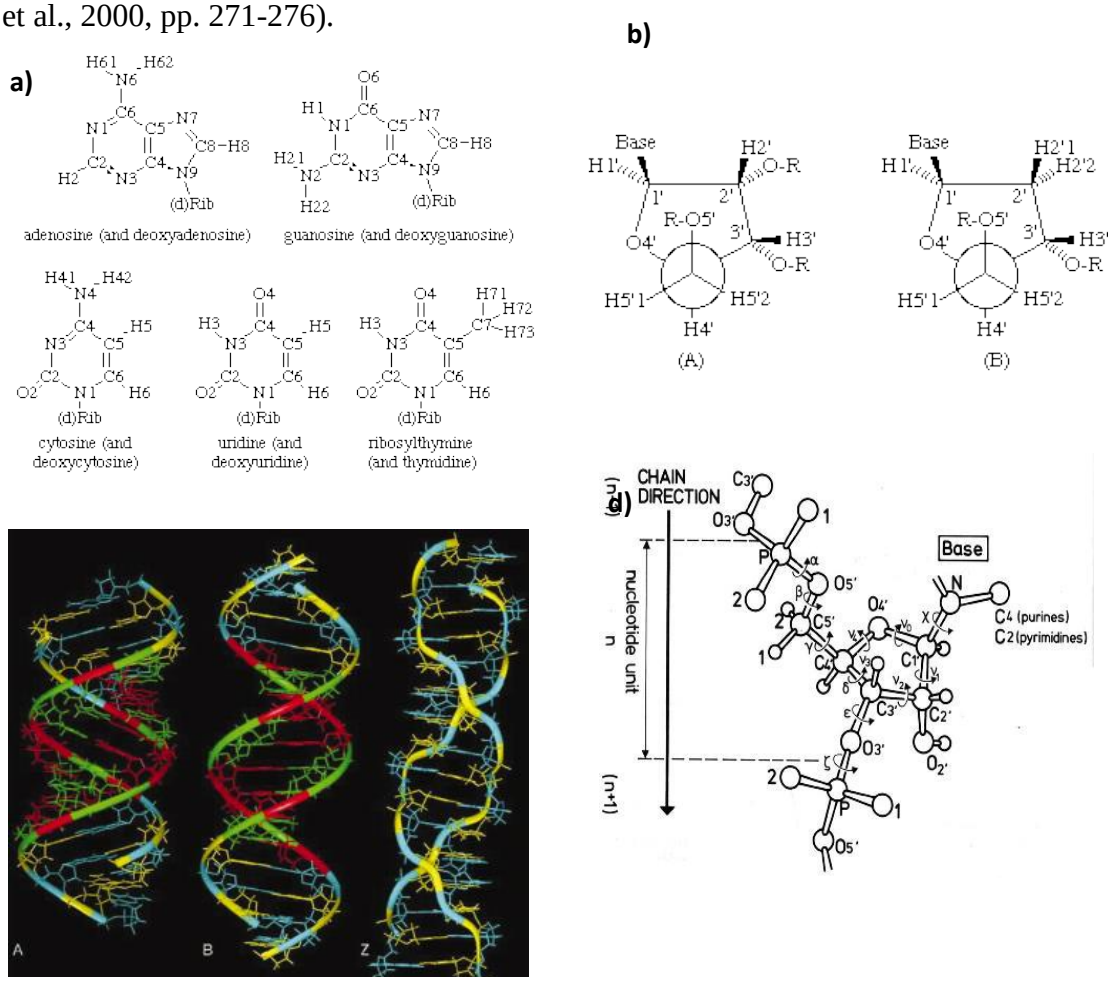


Figure 1.a) Representation of the possible bases in DNA and RNA; b) representation of the sugar bonded to the bases (adapted from IUPAC Commission on the Nomenclature of Organic Chemistry (CNOC) Definitive rules for nomenclature of organic chemistry, J. Am. Chem. Soc. 82, 5545-5574 (1960); c) Representation of the many possible conformations of nucleic acids using DNA as a model to present the most common conformations: A,B and Z-helix (adapted from Ghosh, A., & Bansal, M., 2003); d) Representation of the possible torsion angles on a monomer (adapted from Ghosh, A., & Bansal, M., 2003).

A nucleic acid can assume many conformations, fig. 1 c) (adapted from Ghosh, A., & Bansal, M., 2003), depending on its sequence and the conformation of its sugar pucker. Using DNA as an example, the three most common structures are illustrated above. Although the research conducted was on RNA, DNA structures illustrate a greater variety available in comparison to its RNA counterparts. The reason why will be explained on the paragraph to come.

If the C-3' of the pentose is out of the plane (named $C_{3'}$-endo), gives rise to an A-helix (fig. 1 c) helix A), but if C-2' is out of plane (named $C_{2'}$-endo) it generates a B-helix (fig. 1 c) helix B). A-helix is originated by varying the torsion angle v1 and B-helix, torsion angles v3(fig. 1 d) (adapted from Ghosh, A., & Bansal, M., 2003). Interestingly, another form of helix, named Z-helix, was discovered to be left-handed, instead of right-handed like helixes A and B (fig. 1 c) helix Z). Another notorious difference was the fact that, to hold such conformation, the sequence had to be a continuous alteration between purines and pyrimidines and high salt concentrations to minimize electrostatic repulsion between the phosphates. All three types of helixes exist for DNA but for RNA, only A-helix so far has been observed. This is due to the hydroxyl group (2'-OH) present on the RNA. This 2'-OH prevents the ribose in RNA to adopt a $C_{2'}$-endo conformation in the helical conformation because of steric hindrance. Structural differences between helixes A and B come from the fact that one is favoured in case of relative dehydration of the solvent from helix-A to B, since helix-A is more compact, fewer water molecules bind to phosphate groups. Helix-Z has the sugar-phosphate backbone perform a zigzag thanks to the repeating alternation of purine pyrimidine (Stryer, L. et al., 2000, pp. 789-791).

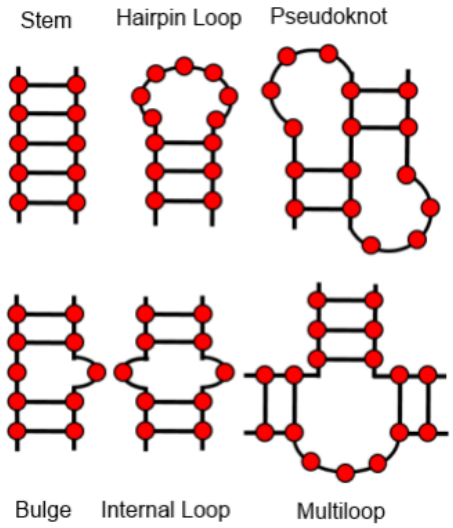


Figure 2. Representation of the most common RNA secondary structures in a 2D plain (adapted from Chheda, N., & Gupta, M. K., 2014).

Regardless of sugar puckering, all helices can have different conformations depending on the context of the sequence, which as a consequence promotes variations of the torsion angles presented in fig. 1 d) (Nelson, D. L., 2008, pp. 280).

When a RNA single strand of a certain sequence folds, secondary structures can arise depending on base pairing pattern (Nelson, D. L., 2008, pp. 284-285). Such secondary structures can be represented in 2D as can be seen in fig 2 exemplified some of the most

common (adapted from Chheda, N., & Gupta, M. K., 2014). A stem is a helical segment where all nucleotides are base paired. A single stranded RNA can fold to form a stem thanks to a hairpin loop. Other secondary structural elements depicted in figure 2 can be found in single stranded RNA as well as in double stranded RNA (duplex) (Eric Westof & Auffinger, P., & Westhof, E., 1998). This high abundance of secondary structure elements confers to RNA molecules the ability to fold into complex tertiary structures to perform their function. In this work we focus on characterising secondary structure in duplex RNA. It should be noted that both the bulge and internal loop secondary structures could also be classified as "internal asymmetrical loop" and "Internal symmetrical loop" respectively. This comes from the fact that an internal loop requires the same amount of nucleotides to be unpaired parallel to each other, while on a bulge the amount is unequal (Westhof, E., & Auffinger, P., 2000).

1-2) RNA biophysics, forces involved in RNA folding and annealing

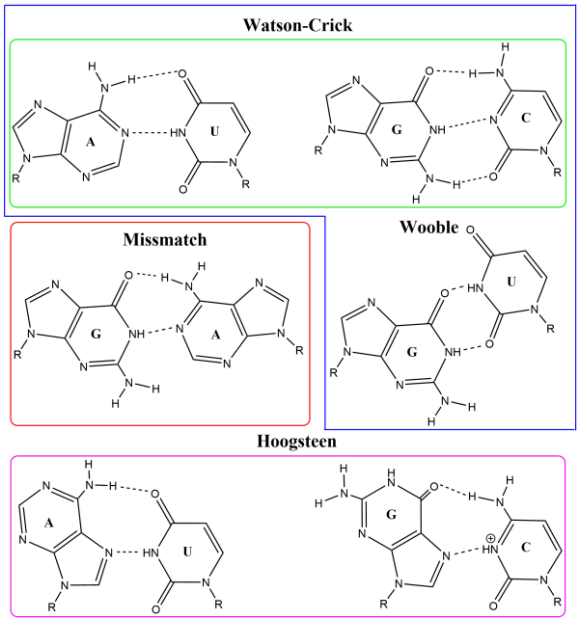


Figure 3. Representation of canonical base pairs (Blue box), containing WC (green box) and Wb pairs, an example of a Mm (red box) and a Hg (pink box) resulting from purine rotation around the glycoside bond and base flipping.

"As exposed previously, RNAs regulate a vast landscape of distinct cellular functions, with each RNA type performing a very specific role. In order for this role to be enacted, certain secondary/tertiary structure features and dynamics must be present, which play a fundamental role in functional complexity" (Mustoe, A. M. et al., 2014). In this sub-chapter, I will review possible base-pairs, base stacking, base pair position structural relevance, secondary structures, solvent exposure, role of different salts.

1-2.2) Possible base-pairs, their structural influences and odd cases

One of the main factors that form RNA structure is the formation of base-pairs, which are currently defined by the following names: Watson-Crick (WC), Hoogsteen (Hg), wobble (Wb) and mismatches (Mm). Watson-Crick and wooble base-pairs can be grouped on the definition of "canonical" (blue box, fig.3) and mismatches "non-

canonical" (red box fig.3) (Roy, A. et al., 2008). WC and Hg refer to AU and GC pairing, being the only difference the H-bond pattern, which happens as a consequence of a purine rotation around the glycosidic bond and base flipping, as portrayed on fig.3 (left).

Mms correspond to all possible combinations of pairing excluding GU pairs and WC. However, GU pairs, or wobbles, have very similar thermodynamic stability to WC pairs in general, hence why they are both grouped as "canonical" and Mms as "non-canonical", being the latter more unstable (Auffinger, P., & Westhof, E., 1998; Roy, A. et al., 2008). Structurally speaking, Wbs can promote sequence dependent and sequence independent changes. In any of the four nucleotides involved in WC base pairs, the glycosidic bond angle between the C1' of the sugar and the base is approximately 54º in A-helix RNAs, independent of sequence. On Gus this angle is approximately 40º in Us and 65º in Gs, indicating that, if a WC base pair would be substituted by either a GU or UG pair, it would have implications on the bending of the RNA (Varani, G., & McClain, W. H., 2000). On the other hand, a context dependent structural feature, would be the under/over twisting that GUs can implicate on the RNA helix, which is highly dependent on base stacking. Besides twisting it could also induce kinks on the structure. Interestingly, both of these sequence dependent features, have been shown to be usefull for docking point on RNA recognizing proteins (Masquida, B., & Westhof, E., 2000).

Mms, are a quite complicated matter structurally. It is known that if a base pair only has two H-bonds or less between bases, the system has a need to stabilize it with the surrounding solvent in order to compensate for energy loss. This is due to hydrogen bond donors and acceptors are exposed (Auffinger, P., & Westhof, E., 1998; Pan, B., & Sundaralingam, M., 1999). In 1999, Baocheng Pan analysed 23 different X-ray resolved structures of RNAs and extrapolated various possible H-bond patterns between Mms. The study came to the conclusion that CA, AA, and UU have two bonds between them, while CUs only have one H-bond while GG and GAs have the possibility of having both one and two by changing conformation. Depending on the near neighbours Mms may increase major groove width and in some other case promote kinks on the structure which can be an accessibility point for RNA recognizing proteins (Pan, B., & Sundaralingam, M., 1999). According to the state of the art as of now, we know by fact that this data is very generalized as it depends on many other factors (to be presented on the next sub-chapter). It must be taken into account, that the latter thermodynamic data was taken from several RNA crystallography structures averaging thermodynamic data to reach a general conclusion. In reality, pairing stability is context dependent, subject to sequence and its

surrounding solvent. If we think of all possible RNA structures, we will have an array of all possible arrangements. But on certain conditions, certain conformations are more thermodynamically probable to be present, stabilizing them.

In order for the well-known RNA double helix to be formed, the molecule must be in a polar environment. On this situation, it is promoted the interaction between the backbone phosphates with the solvent, the formation of a hydrophobic centre along the RNA where the bases are protected from the solvent, stacking on each other. Hydrogen bonds between bases, interaction between sugars are also promoted. This play of forces in concert leave the RNA mostly exposed to the solvent with the exception of the hydrogen bonded atoms, which can be observed in water (Westhof, E., & Beveridge, D. L., 1990). By now, it should be evident that in reality when talking about base pair and structure stability, a series of forces in play from a specific environment must be taken into account.

1-2.3) Base stacking

Base stacking is a major driving force in RNA structure folding. Stacking can be described, by approximation, as a combination of London dispersion attraction, short-range repulsion, electrostatic interactions and interactions with the bulk solvent. Even though frequently used terms as π-π base stacking are used to describe these interactions, quantum chemical research points out the absence of specific delocalized π-electrons participating on stacking energies. However, a complete physic-chemical description of what is base stacking remains to be reached, since it is context specific and environment dependent (Šponer, J. et al., 2013). Understanding base stacking interactions and their contributions to A-helix stability is necessary to understand the basis of RNA structural and dynamic behaviour as designing reliable secondary structure prediction software.

Throughout the years many studies have been conducted on RNA oligos of different sizes and sequences. Most of these experiments that aimed at quantifying RNA-RNA interactions were taking advantage of the hyper-chromic shift in absorption at 260 nm that characterise double strand to single strand transition (melting transitions) in nucleic acids. A great advance on the state of the art in understanding stability of base stacking can be attributed to Martin Serra, Douglas Turner and Znosko's groups. Similar methodology was provided by all authors using UV-melting and NMR in some cases, with different experimental validated secondary structure RNA duplexes through data base searching. In 1989 Carl Longfellow and colleagues published the first article to

tackle thermodynamics of bulges in duplex stability depending on the adjacent base pairs on its 3' and 5' sides, accounting for base stacking interactions. Through melting experiments of several duplexes with a single nucleotide bulge, a model for predicting the energy contribution of bulges was attempted to be used to fit the data. The authors exposed the fact that probably constructing a model considering only near-neighbour base pairs of the bulge is not adequate. This is supported by several points observed: in some cases, bigger bulges are more destabilizing, in other not; the same bulge, flanked by the same directly adjacent base pairs at 3' and 5' but with different base pair sequences beyond displayed different stabilities (Carl E. Longfellow et al., 1990). Later in 2007 Joshua Blose et al probed the effects of near and non-near neighbours base pairs on bulge and duplex stabilities. In similarity to Znosco B. et al in 2004 (Znosko, B. M. et al., 2004), in general, pyrimidine single bulges tend to be more stable then purine ones. The author exposed a relationship between duplex stability and bulge position based on non-near neighbour interactions. Having a bulge closer to the stem end increases its stability in comparison to bulges near the middle of the stem. This could be because thigh packing and electrostatic constraints would impose a higher energy cost for insertion of a Mms on the middle rather than at the end of the stem (Cate J. et al, 1999). Based only on near-neighbour interactions a general ranking was presumed for bulge stability: $\frac{5'-GBG-3'}{3'-U\ \ U-5'} \approx \frac{5'-UBU-3'}{3'-G\ \ G-5'} > \frac{5'-UBX-3'}{3'-G\ \ Y-5'} > \frac{5'-GBX-3'}{3'-U\ \ Y-5'} \approx \frac{5'-XBU-3'}{3'-Y\ \ G-5'}$ (Blose, J. M. et al., 2007) being XY any WC base pair and B a bulged nucleotide. In 1999 UV melting and NMR studies paved the works for establishing a hierarchy for Mms influence in duplex stability. Ryszar Kierzek and co-workers concluded that depending on Mm identity and its position the destabilizing effect changed, giving insight into non-near-neighbour effects on Mms. As with bulges, Mms like UU or AA seem to destabilize the duplex less at the end of the stem rather than at the middle, except for the case of GG Mms which seemed to be insensitive to position. The authors probed for near-neighbour patterns of single Mm stability near WC base pairs: $\frac{5'-G\boldsymbol{U}G-3'}{3'-C\boldsymbol{U}C-5'} > \frac{5'-A\boldsymbol{U}A-3'}{3'-U\boldsymbol{U}U-5'}$; but also theorised in conjunction, the justification for the findings of Santalucia in 1991 ranking: $\frac{5'-C\boldsymbol{GA}G-3'}{3'-G\boldsymbol{AG}C-5'} > \frac{5'-C\boldsymbol{GG}G-3'}{3'-G\boldsymbol{GG}C-5'}$ for tandem Mms, and $\frac{5'-C\boldsymbol{G}G-3'}{3'-G\boldsymbol{G}C-5'} > \frac{5'-C\boldsymbol{G}G-3'}{3'-G\boldsymbol{G}C-5'}$ for single Mms (SantaLucia Jr, J. et al., 1991), being the Mms in bold. As in 2007 Joshua Blose and co-workers, Ryszar Kierzek and co-workers based the justification for this hierarchy on environment constriction (Cate J. et al, 1999, Science). Hence tandem GA Mms would have more flexibility then a single GA

Mm, therefore being able to accommodate pairing between nucleotides and vice versa for the case of GG. This was confirmed by NMR observing H-bonds for single GG Mms as based on imino proton NMR spectra (Peyret, N. et al., 1999). Already the idea of importance of context in which a certain base pair is installed arises, clarifying that the identity of a single Mm or any other base pair alone cannot dictate how stable it is. Clearly its stability depends on its neighbours and its distant neighbours. In 2008 Amber R. Davis & Bret M. Znosko, through UV-melting experiments of several RNA duplexes, presented a general hierarchy of Mm stability disregarding which adjacent base pairs were present: GG > UU > AC > CU > AA ≈ AG ≈ CC. This hierarchy however, was presented in ranges of Kcal/mol, which already indicates that identity alone couldn't possibly define stability (Davis, A. R., & Znosko, B. M., 2008). Later studies by Znosko's group would clarify even further the identity problem, comparing the later hierarchy considering adjacent GU Wbs (Davis, A. R., & Znosko, B. M., 2008), all possible combination of tandem Mms and their closing adjacent base pairs (Christiansen, M. E., & Znosko, B. M., 2008) and stability of tandem Mms in different sequence contexts (Christiansen, M. E., & Znosko, B. M., 2009). In all previous presented cases, the hierarchy changed. In 2008 Martin Serra's group highlighted the importance of dangling single and double nucleotides at 3' ends of the stem, adding yet another parameter to the equation. Through melting studies, it was observed the stabilizing capacity of this nucleotides in a sequence dependent manner which depend on the closing base pair. This could be achieved by stacking of the dangling nucleotides with the closing nucleotides at the stem end (Miller, S. et al., 2008). Helix stabilization by 3' dangling ends is related to strengthening of H-bonds in the terminal base pair due to shielding from the solvent (Isaksson, J., & Chattopadhyaya, J., 2005). More recently, in 2016, Martins group developed melting studies on RNAs with two bulges, investigating their influence in duplexes of different sizes, distance and orientation between them. In general, the were deemed to have a destabilizing effect if inserted in stems with no bulges, having a smaller destabilizing effect on smaller stems, probably to the lack of non-near neighbour interactions. No correlation was found between distance and orientation of the bulges, but the authors do point out that the model for energy contribution of the bulges to duplex stability was incomplete. This can be justified by observing 3D structures where in some cases, the bulges are separated by one base pair and the free bulged nucleotides can stack, or interact with a second RNA molecule (Crowther, C. V. et al, 2017).

It is safe to say there are an enormous amount of parameters involved in base pair stability and stacking.

1-2.4) Cations on structure stabilization

Having into account that RNA is a polyanion, cations play an obvious role at charge stabilization. A barrier yet to surpass is the understanding of population distribution of certain cations, which is a consequence of the highly variable RNA structure. The frequency of studies on Mg^{2+} stabilization of RNA on the literature is a consequence of how it can easily stabilize RNA at only mM concentrations, while other monovalent cations can only reach mediocre stability at much higher concentrations. (Draper, D. E., 2004). Simulations have shown that size of the ions matters. Smaller ions are more effective the bigger ones on structure stability. Through observation of mobility of folded RNA with different ions at the same concentration in an 8% non-denaturing polyacrylamide gel, it was deduced that divalent ions with decreased charged density stabilized RNA the most. The hierarchy is as: $Mg^{2+} > Ca^{2+} > Sr^{2+}$ (Koculi, E. et al., 2007).

1-3) System of study: miRNA-34a

1-3.1) MiRNA biosynthesis and RISC complex formation

The biosynthesis of Human miRNAs can be divided on the following steps on set order: pri-miRNA formation; pre-miRNA formation; cytoplasm translocation; miRNA duplex loading. This steps will be explained here in conjunction with a small background to the relevant proteins involved in this process whilst elucidating what is a miRNA.

1-3.1.1) *From Pri to Pre-miRNA*

MiRNAs genes are transcribed by Polymerase II to a large pri-miRNA with a 5'-Cap and a 3'-Polyadeninylation modification (Lee, Y. et al., 2004). Pri-miRNA is then processed by the enzyme Drosha on the nucleus, which executes a double endo-nucleolytic cleavage. The result is a ~70 nt pre-miRNA hairpin precursor, to be transported by Exportin5-RAN-GTP complex to the cytoplasm (Yi, R. et al., 2003). Once on the cytoplasm, Dicer-TRBP protein complex, preforms an endo-nucleolytic cleavage, originating an asymmetric RNA duplex containing the 22-nt guide (miRNA) and a partially complementary passenger strand (miRNA*) (Ha, M., & Kim, V. N., 2014; Winter et al., 2009). Thanks to partial complementarity between strands, non-canonical base pairs, mismatches and bulges can be present. However, other structural features

arisen from Dicers activity, which will prove to be more essential for miRNA targeting, being a 5' phosphorus group on both ends and two 2 nts overhanging (Nykänen, A. et al., 2001). The final step comprises loading of the complex into the Argonaute protein (Ago 1-4), forming the RNA-Induced Silencing complex (RISC), a multiprotein complex. To proceed to this last step, first, we require a small introduction to Argonaute and the RISC complex.

1-3.1.2) *Ago and RISC*

Argonaut proteins participate in the gene regulation pathways (RNAi) on eukaryotes and defence against exogenous DNA material in prokaryotes (Makarova, K.S. et al., 2009; Olovnikov, I. et al., 2013). They are composed of three main domains: PAZ, MID and PIWI. PAZ is thought to participate in RNA binding, MID 5' nt recognition of the guide strand, while PIWI in endoribonuclease activity and MID (Sontheimer, E. J., 2005; Faehnle, C. R., & Joshua-Tor, L., 2010). There are eight isoforms of Agos in humans, four of which largely expressed in many tissues and studied in greater details (hAgo1, hAgo2, hAgo3 and hAgo4) (Peters, L. et al, 2007). eAgos (eukaryotic Ago) are often classified according to their ability to enzymatically cleave a scissile phosphodiester bond of the sugar-phosphate backbone of the target in a RNAi type of activity. Of these isoforms, hAgo2 is the only one able to slice the targets, whereas hAgo1, hAgo3 and hAgo4 are not (Meister, G. et al., 2004). According to screening in-vivo studies, hago2 and 3 presented a buffet of miRNA partners (Azuma-Mukai et al., 2008), suggesting that slicing could be a major player on mRNA pos-transcriptional control. As previously said, Ago participates on the formation of the RISC complex, a multiprotein complex, in conjunction with its loaded miRNA. Unfortunately, a single coherent view of RISCs actual full composition is still not solid. This arises from several different recorded complex sizes across publications. Still, one common trend seems to emerge from the noise, pointing that in all case Ago is present (Sontheimer, E. J., 2005).

1-3.1.3) *miRNA loading*

The final miRNA maturation involves loading of the miRNA/miRNA* complex into Ago. During loading the guide miRNA is preferentially selected over miRNA. This is due to thermodynamic stability of the 5' ends of the duplex (Khvorova, A. et al., 2003). The strand that exhibits less stable base pair at the 5' end is selectively loaded, requiring ATP-dependent Hsc70/Hsp90 chaperone machinery, whereas the other strand is discarded (Iwasaki, A., & Medzhitov, R., 2010; Miyoshi et al., 2011), by a process not well known.

The binding initiates at the 5' nucleotide (Iwasaki, A., & Medzhitov, R., 2010) that is first recognised by MID preferring Us and As over Gs and Cs in mammalian Ago (Frank et al., 2010). Afterwards, a structural rearrangement brings together the MID and PIWI domain to form a composite 5' binding pocket and allow the rest of the RNA duplex to dock into the nucleic acid channel of Ago (Nakanishi, K., 2016). Lastly, removal of the passenger strand, miRNA*, can be achieved by either by a step-wise conformational change of the protein that expels the thermodynamically unstable RNA strand (Ago1, Ago3 and Ago4 isoforms) or through the intrinsic endo-nucleolytic activity of Ago2 isoform. The end product would be the assembled protein complex, Ago plus the rest of the RISC complex, and a single stranded miRNA, ready to recognize one of its many mRNA partners. This presents the miRNA forming a kink on the backbone contacting with the Ago protein (Nakanishi, K., 2016).

1-3.2) miRNA-mRNA interactions:

MiRNAs, unlike siRNAs, don't have a full complementary sequence to the mRNA target region. Instead, they have a "seed region", located at the 5' of the miRNA, which there is full complementarity between nucleotide 2-7 of the miRNA. There other nucleotides however, have imperfect base pairing, including Mms, Wbs, bulges and loops (Bartel, D.P., 2009). To be taken into heavy consideration, is the several paths that miRNAs can regulate gene expression, in between them so far observed: translational repression, mRNA cleavage, and deadenylation (Maroof, H. et al. 2014).

David Bartel group throughout the years has been doing consistent work on the study of miRNA binding motifs with an array of experimental technics coupled computational prediction (fig. 4).

The four most common seed region types are: 6-mer, a perfect match from nucleotides 2-7 of miRNA; 7mer-8mer, seed match augmented by an additional match at the 8^{th} nucleotide; 7mer-A1, seed match augmented by a WC base pair with an A from the mRNA with the 1^{st} nucleotide of the miRNA; 8-mer, seed match flanked by both a

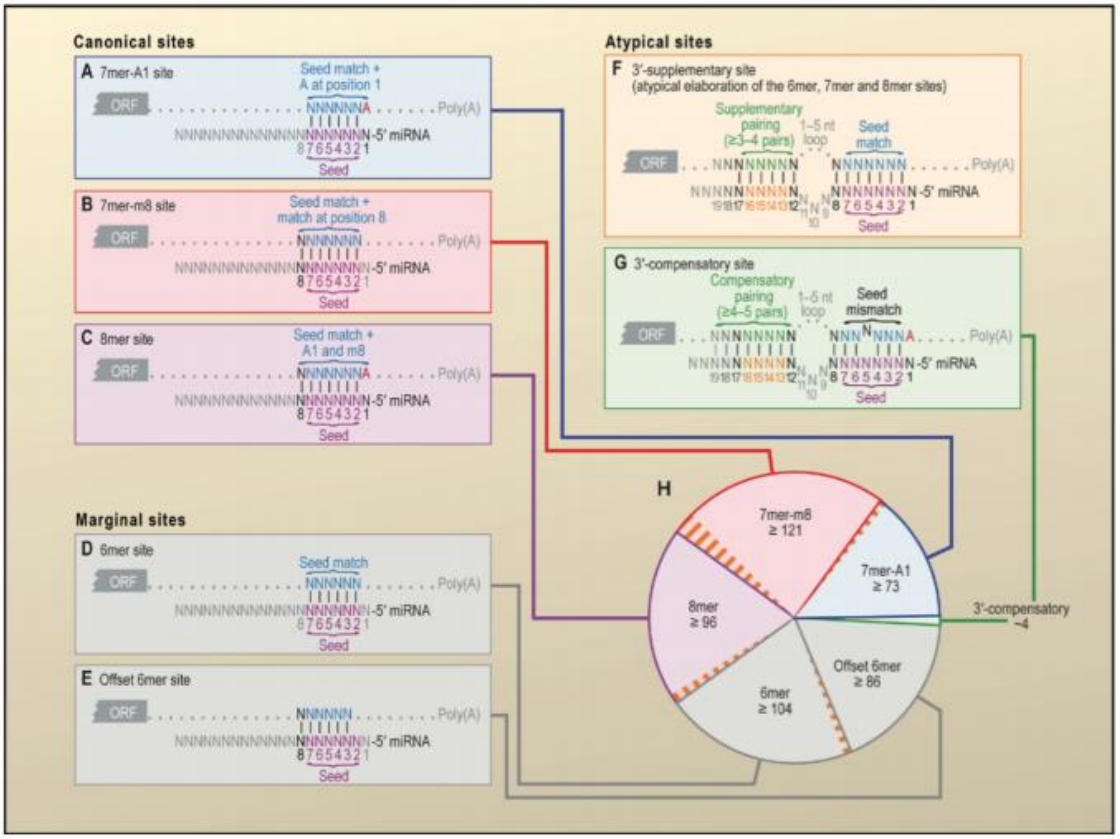


Figure 4. (Reproduced from Bartel 2009) - Types of miRNA target sites - (A,B and C) Canonical sites consist of WC base paired nucleotides from position to 2-7. Base pair can extend to position 8 where induce improved efficacy. Adenine opposite of position 1 of the guide has also proven to increase silencing efficacy. (D and E) Marginal sites where only the seed is base paired. (F and G) Atypical sites are considered when a supplementary 3' base pair is observed centred around positions 13-16 of the guide. (H) Number of preferentially conserved mammalian sites matching a typical highly conserved miRNA.

WC base pair at nucleotide number 8 and 1 (Grimson, A. et al., 2007). Interestingly, it is not only in RNA biophysics that context matters, also in targeting the same abstract concept applies to UTR context. MiRNAs can target mRNA either on the ORF, 5'-UTR or 3'UTR, but it has been reported that the most frequent and efficient targeting happens at the 3'-UTR (Bartel, D.P., 2009). In comparison, normally 7-8mer sites in general seem so be efficient sites, but in certain UTR contexts that can vary to high extents. Identity alone, does not directly define repression capability. This capability is also affected by other miRNA bonding sites close by, promoting cooperation. However, distance between sites dictates this cooperation, having them too far away has the same effect as if they would be alone, too close and it could have negative effects on repression, but at a distance interval they can augment each other, if not closer the eight nucleitides (Grimson, A. et al., 2007; Bartel, D.P., 2009). Distance from the stop codon and the Poly-(A) tail is also of importance since there is a correlation of efficiency between optimal distance from both. If target site is placed on the middle it tends to be less efficient then if closer to stop codons or Poly-(A) tail (Grimson, A. et al., 2007). The reason for this is theorized to be the same to why such sites are conservatively flanked by AU rich areas: highly unfolded

strands. This could possibly grand easier access to the RISC complex, hence augmenting inhibition efficiency (Bartel, D.P., 2009). Conservation of WC base pair positioning on the miRNA from nucleotides 13-16, was mined through use of a computational rubric for identifying productive 3' pairing in miRNAs. This site is associated with enhanced repression (Grimson, A. et al., 2007).

Away from the regular cases, approximately one percent of the sites found until 2009, has been denoted as "3' compensatory". This interesting case arises from a mismatched seed region which in order to compensate for loss in complementary on the 5' end of the miRNA, compensates with additional WC pairing at the 3' end (Bartel, D.P., 2009).

1-3.3) MiRNA-34 family - p53 and the case of miRNA-34a:

The miRNA-34 family consists of three members so far recorded present in humans: miRNA-34a, miRNA-34b and miRNA-34c. MiRNA-34b and 34c both have 23 nucleotides, while miRNA-34a has 22. These three non-coding tumour suppressor RNAs when deregulated, consequently affect both genetic and epigenetic pathways in a vast array of cancers: brain, pancreas, skin, lung, lymphoid system, colon, oesophagus, liver, cervix and prostate. On the basic mode of action of cancer cells, this family of miRNAs influence: proliferation, invasion, metastasis, apoptosis, cell survival, cell cycle progression, senescence, migration, angiogenesis and promoter methylation (Maroof, H. et al. 2014). Also important roles such as neuronal development (Corney, D. C. et al., 2010) and liver regeneration (Wang, X. P. et al., 2016) are controlled by this family.

The transcription factor p53 is post-transcriptionally activated when in case of oxidative stress, damaging of DNA and activation of a potential oncogene. The miRNA-34 family is a confirmed target of p53. It is prevalent in most studies that miRNA-34 family are considerably induced then other miRNAs, with miRNA-34a in distinction. Working in concert, miRNA-34a is required for p53 tumour suppression regulation to take action, underlining even more the essential role of this miRNA (Hermeking, H., 2007).

1-3.4) Targets in study for miRNA-34a:

1-3.4.1) CD44:

CD44 is a cell surface glycoprotein (Spring, F. A., 1988), which has affinity for a variety of ligands, of which hyaluronic acid (HA) is one of the most important and well studied. CD44-HA complex plays a role in cell-cell and cell-matrix interactions and as a

consequence, CD44 dictates cell migration and if misregulated, can participate in tumor growth (Vikesaa, J. et al., 2006).

In 2011 mRNA CD44 was identified and validated as a target of miRNA-34a, showing the ability to inhibit metastasis of prostate cancer cells (Liu, C. et al., 2011).

1-3.4.2) *PNUTS*:

PNUTS is serine/threonine-specific protein phosphatase which participates mainly on the regulation of apoptosis in cancer cell lines (De Leon, G., 2008), DNA damage repair (Landsverk, H. B. et al., 2010) and telomere maintenance (Kim H. et al., 2009).

Recently in 2013 mRNA PNUTS was validated as a target for miRNA-34a (Boon, R. A. et al., 2013). Reinier demonstrates, through the inhibitory effects of miRNA-34a on mPNUTS, that this protein has an in important role in aging. As the organism gets older, miR-34a expression is highly induced, inhibiting PNUTS, hence promoting DNA damage and apoptosis, which normally leads to heart failure. Not surprisingly, knockdown of PNUTS in cancer cell lines seems to cause reduced viability, inducing apoptosis (De Leon, G., 2008).

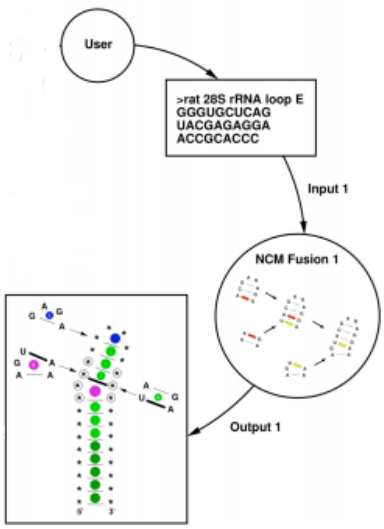


Figure 5. Display of the logic MCfold software uses to build the 2D structure representation of the RNA secondary structure. Using NCMs (nucleotide cyclic motifs), the software tries piecing several pieces together depending on the input sequence. The decision on a hierarchy of stable structures as an output is then decided by an energy minimization algorithm (adapted from Parisien, M., & Major, F., 2008).

1-3.4.3) *WNT1*:

WNT1 is part of the wnt family, which consists of secreted glycoproteins that participate in intercellular signalling during development (Komiya, Y., & Habas, R., 2008; Nusse, R. et al., 1990). This family has been characterized over the years as a crucial player in tissue development and formation, where WNT1 takes part mostly in neural development, such as central neural system of mouse fetus (McMahon, A. P., & Bradley, A., 1990), human monocyte dendritic cell maturation (Zhan, Y., & Wu, L., 2012) and dopaminergic neurogenesis (Castelo-Branco G. et al, 2013), but also bone formation and remodelling (Keupp, K. et al., 2013). WNT1, is also a proto-oncogene, which overexpression can lead to

mammary tumours (T. Zhan et al 2017 Oncogene), gastric and pancreatic cancer (Katoh, M., & Katoh, M., 2003).

In 2009 mRNA WNT1 was identified as a target for miRNA-34a (Hashimi, S. T. et al., 2009), providing also insight into the influence of this miRNA in monocyte-derived dendritic cell differentiation, through inhibition of WNT1.

1-4) Method introduction:

1-4.1) MC-Fold:

MC-Fold is a RNA secondary structure simulation software based on sequence as an input. The output consists of both a dot bracket notation of the secondary structure and a pdf file with a 2D figure representation of set structure, both accompanied by its calculated free energy. McFold progressed RNA prediction by taking into account all types of base pairs, from WC, to Wb and Mms. In order to build structure from sequence, the authors presumed that it was enough to take into account nucleotide cyclic motifs (NCM) as a first order object. From this small NMCs, big RNA structures can be built joining various small NMCs together using an energy minimization algorithm to create a ranking of least to most stable structures, as demonstrated in fig. 5 (adapted from Parisien, M., & Major, F., 2008). A database of all possible NCMs was generated creating a series of possible loops, stems and bulges. In this way, all possible structure conformations are generated from a sequence. For each structure, each small NCM or secondary structure features energy, both experimental and predicted, are added to give the final free energy of the RNA (Parisien, M., & Major, F., 2008).

Although a very good tool, limitations exist. The limited amount of experimental data on Mms limits the algorithm (Zuker, M., 2003). Dangling nucleotides as stabilizing agents and non-near neighbour interactions are disregarded (Parisien, M., & Major, F., 2008).

1-4.2) T7 polymerase, the RNA producing system:

In order to produce RNA sample, *in vitro* transcription using T7 polymerase was applied, since it is one of the most efficient ways of obtaining sample of intermediate size (20-150 nucleotides) (Milligan Groebe 1987). A drawback of this technic, is the fact that several side products of different lengths can be produced by T7 polymerase. Gladly there are ways to circumvent this issues. A common technic is to add 2 Gs at the 5’-end of the

DNA template (if not already naturally present), which can increase the yield of desired product (Kuzmine, I. et al., 2003). If 2 Gs are on the 5'-end, GMP can be added, since it has been described that it has higher affinity to T7 then GTP (Martin, C. T., & Coleman, J. E., 1989). To diminish the addition of nucleotides at the 3'-end we used C2´-oxy-methilated at the two last 5'-end bases of the DNA template (Kao, C. et al., 1999), which in combination with DMSO provides significant changes in yield (Chen, Z. & Zhang, Y., 2005).

For further optimization of the reaction mixture other reagents must also be used and varied: Tris HCl for buffer capacity; $MgCl_2$ is an essential cofactor of T7 forming a complex with the triphosphates, compensating for negative charges (Yin, Y., & Carter Jr, C. W., 1996); DL-Dithiothreitol (DTT) to prevent oxidative damage to the protein; Spermidine (SP) as an enhancer of T7 transcription reaction (Iwata, M. et al., 2000).

Each nucleotide (ATP, CTP, GTP, UTP) along with the annealed DNA is also introduced and varied if necessary, along with the T7 polymerase.

1-4.3) HPLC in RNA purification:

In biophysics in order to study a particular molecular system, it is of high priority that said system is solvated in only buffer containing no impurities in order to obtain trustworthy data. In our case, such impurities qualify as: T7 polymerase, DNA template and RNA side products from T7 transcription. To purify HPLC is employed.

In focus two types of columns are introduced: 1) Reverse-phase Ion-pairing and 2) Ion-exchange. 1) Stationary phase is non-ionic and hydrophobic whilst the RNA is the opposite situation due to the presence of the phosphate groups. To promote interaction with the stationary phase an organic cation is used to interact with RNA and serve as the interacting bridge between it and the stationary phase. Once the RNA interacts with the column through the organic cation, an organic solvent, such as acetonitrile is use to compete with the cation for the interaction with the stationary phase. 2) The anion-exchange has different retention periods of the RNA depending on the molecular size and sequence. The stationary phase is a matrix of positively charged functional groups ready to interact with the phosphate backbone of the RNA and other ions on the liquid phase. Separation is achieved by competition of the several anions of the elution buffer and the RNA for the stationary phase, increasing concentration of salts like $NaCl_2$, promotes such gradual separation (Gjerde, D. T. et al., 2009).

1-4.4) UV melting of RNAs:

In the RNA field, to infer both melting temperature and concentration, it is commonly used UV spectroscopy. This technic takes advantage of the interaction of UV radiation and the π-electrons delocalized on the aromatic rings of the nucleotide aromatic bases. Measuring absorbance and extrapolating concentration was preformed using Beer-Lambert Law. The absorbance monitored is at 260 nm (Dumas, P., 2014).

RNA melting studied by UV spectroscopy is due to increase in temperature, inducing base pair opening. Duplex unfolding lessens the near-neighbour effects on each base resulting in hyperchromicity or in other words, increase in Abs. Hence duplex formation results in hypochromicity or in other words, decrease in Abs (Klostermeier, D., & Hammann, C., 2013). Melting of a RNA duplex monitored by an UV spectrophotometer, results in a sigmoidal curve which could then be used to extrapolate its Tm.

1-4.5) NMR:

NMR, like other spectroscopy technics, utilizes the interaction between matter and electromagnetic radiation to extrapolate information of the system. The sample containing the system of interest is positioned in a magnetic field and exposed to electromagnetic radiation. On this technic, nuclei absorb and then re-emit the radiation at a certain frequency on the Mhz range. This frequency depends on three factors: identity of nucleus, chemical environment on the vicinity of the nucleus and strength of the field.

Chemical environment differences provide specially with a range of different frequencies between nucleus of the same identity. With this in mind, in practice, we could then determine in which environment a certain nucleus could be surrounded. But the problem arises if we measure that same nucleus in different field strengths, which would output a different frequency. For practicality then, chemical shift is used, being:

$$\delta = \frac{\upsilon_{sample} - \upsilon_{ref}}{\upsilon_{ref}} . 10^6$$

This formulation for chemical shift (δ), provides that the value is matches in different machines with different field strengths. In this case, the reference used is normally TMS (tetramethylsilane). The value of the chemical shift is reported in parts per million (ppm) being TMS ppm=0. 1D NMR experiments for ^1H, can return information of every ppm of each single ^1H present in the sample.

Unfortunately, overlap is a considerable issue in NMR, especially in 1D experiments. To overcome such issues, 2D NMR experiments come of aid. Such technics used in this work explore interaction between nucleus through bond or space. Through bond interactions are defined as ^{n}J, being "n" the number of bonds separating the two coupled nucleus and "J" defines the type of coupling through bond interaction. It should be noted that, bigger "n" corresponds to weaker interaction. This gives rise to ^{1}H-^{15}N/^{13}C HSQCs (Heteronuclear Single Quantum Coherence) for RNAs, where we have ^{2}J. Through space interactions can be shorted to NOE interactions (Nuclear Overhauser Effect). The NOE effect is dependent of distance, where the bigger the distance between two nuclei the weaker the signal. It provides us with valuable structural information constructing the surroundings of each nucleus until we can describe the whole sample. Advantage of such interactions gives rise to the 2D experiment, NOESY ^{1}H-^{1}H experiments (Nuclear Overhauser Enhancement Spectroscopy) (Crichton, R. R., & Louro, R. O., 2012).

2) Materials and methods

2-1) Materials:

2-1.1) Reagents:

2-1.1.1) *Transcription reaction*:

DNA template, with T7 promoter sequence, 5'-TATAGTGAGTCGTATTAA-3' at the 3´-end and C2´-methoxy groups at the two last nucleotides at 5´-end, HPLC purified or desalted from e.g. IDT; Ultrapure type 1 water, 18.2 MΩ.cm@25°C; Tris base (Fisher Scientific, no. 10103203); MgCl2 (Sigma-Aldrich Corporation no. 63068); DL-Dithiotreitol (Sigma-Aldrich Corporation no. 43815); Spermidine Trihydrochloride Bioextra (SigmaAldrich Corporation no. 85578); DMSO (Sigma-Aldrich Corporation no. D8418); GMP (Sigma-Aldrich Corporation no. 650684); UMP (Sigma-Aldrich Corporation no. 651370); ATP (Sigma-Aldrich Corporation no. A2383); UTP (Sigma-Aldrich Corporation no. U6625); GTP (Sigma-Aldrich Corporation no. G8377); CTP (Sigma-Aldrich Corporation no. C1506); NTPs 13C, 15N-labeled (Silantes, product no. 121606200); T7 RNA polymerase; Inorganic Pyrophosphatase; RNAseOUT (ThermoFisher catalog no. 10777019).

2-1.1.2) *HPLC Buffers*:

Acetonitrile (Sigma-Aldrich Corporation no. 34851); NH4OAc (Fisher Scientific no.0598410); NaOAc (Sigma-Aldrich Corporation no. S2889); NaClO4 (Sigma-Aldrich Corporation no. 310514); Acetic Acid (Sigma-Aldrich Corporation no. 49199); Hydrochloric Acid (Sigma-Aldrich Corporation no. H1758); Sodium Hydroxide (Sigma-Aldrich Corporation no. 72068); Tetrabutylammonium hydrogen sulphate (Fisher Scientific no. 0095650).

2-1.1.3) *NMR Buffer*:

Sodium phosphate, monobasic (Sigma-Aldrich Corporation no. 57907); Sodium phosphate, dibasic (Sigma-Aldrich Corporation no. 57907); NaCl (Sigma-Aldrich Corporation no. S3014); EDTA (Sigma-Aldrich Corporation no. E6758).

2-1.1.4) *Gel electrophoresis*:

40% Acrylamide/Bis Solution (Bio-Rad no. 161-0144); Invitrogen UltraPure 10X TBE-buffer (SigmaAldrich Corporation no. T4415); Ethidium bromide (Merck no. 1.11608.0030). Urea 8M (Sigma-Aldrich Corporation no. 101822332).

2-1.2) Equipment:

2-1.2.1) *Material concentration/ solvent exchange*:

Micro tube, 1.5 ml (Sarstedt no. 72.690.001); Centrifuge tube, 15 ml VWR Collection; (VWR no. 525-0400); Centrifuge tube, 50 ml VWR Collection (VWR no. 525-0402); Centrifuge tube, 15 ml VWR Collection (VWR no. 525-0400); Centrifugal filter units 3 kDa (or larger depending on RNA), 15, 3 and 2 ml; Refrigerated centrifuge (Eppendorf 5804 R).

2-1.2.2) *HPLC system*:

Dionex Ultimate 3000 UHPLC system, Thermo Scientific (Chromeleon 7 operating software); AFC-3000, HPLC Fraction collector (Thermo Scientific no. 5702.1000); DNAPac PA200 9x250 Semi-Prep column (Thermo Scientific, product no. 063421); DNAPac PA200 9x50 guard column (Thermo Scientific, product no. 063419); DNAPac PA200 22x250 Semi-Prep column (Thermo Scientific, product no. SP6734); DNAPac PA200 22x50 guard column (Thermo Scientific, product no. SP6731); Hypersil Gold 250×10 Reverse phase Semi prep column (Thermo Scientific, product no. 5005-259070); Hypersil Gold 10×10 guard column (Thermo Scientific, product no. 25005-019023).

2-1.2.3) *Gel electrophoresis*:

Amersham Biosciences Electrophoresis Power Supply 3501 XL; Bio-Rad ROTEAN® II electrophoresis equipment.

2-1.2.4) *UV spectrophotometer*:

Evolution™ 260 Bio UV-Visible Spectrophotometer; ThermoFisher®. Quartz SUPRASIL High precision cell, Hellma Analytics, art. No. 105-201-85-40.

2-1.2.5) *NMR Spectrometer*:

Bruker Avance III spectrometer operating at 600 MHz (1H resonance frequency), with a 5 mm QCI cryoprobe (1H/13C/15N/31P). NMR spectra were acquired and processed using Bruker Topspin 3.2 software and Sparky used for assignment.

2-2) Methods:

2-2.1) MCFold structure generation:

With MCFold (Parisien, M., & Major, F., 2008) RNA structures were simulated from sequence for both the systems to be studied biophysically and other mRNA targets of miRNA-34a. Structures for the duplexes miRNA-34a.gCD44, miRNA-34a.PNUTS, and the hairpins of gCD44, PNUTS and miRNA-34a were simulated for comparison with biophysical data. A total of 34 other mRNAs were simulated in complex with miRNA-34a with the goal of extrapolating possible general hints structural features required for miRNA targeting. All of this 34 mRNAs were validated as targets by luciferase assays and were marked as validated by Tarbase v7.0 (I. S. Vlachos et al, 2014, Nucl. Acids Res.).

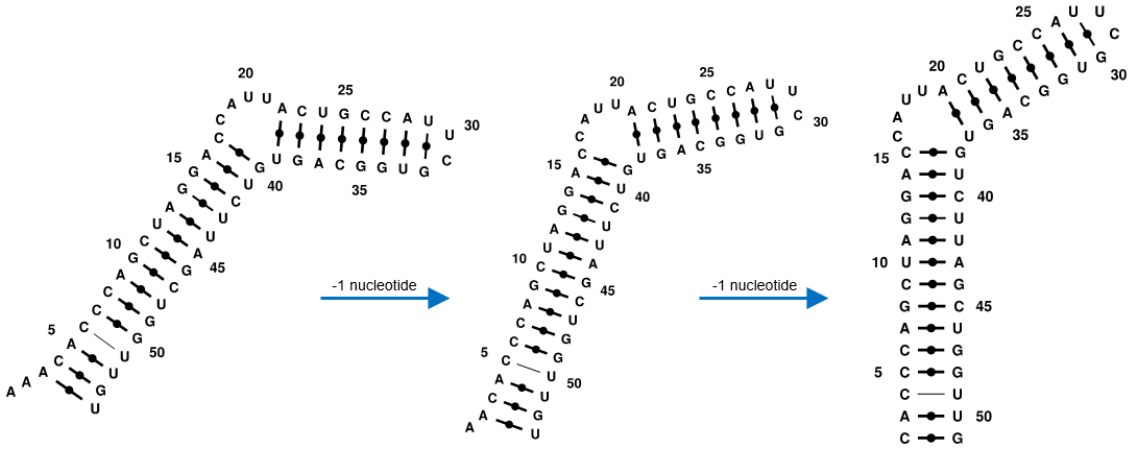


figure 6. Simulation of a duplex between an mRNA (strand on the left) and miR-34a (strand on the right) as an example of the compromise between mRNA length considering NMR limitations while retaining secondary structure representative of the mRNA.miRNA interaction.

For each sequence to be studied biophysically, fifty conformations were simulated using MCFold 2.32, but for the bulk 34 mRNAs, only twenty were used. For duplexes, the two sequences of mRNA and miRNA were linked with a 5'-UUCG-3' sequence, which would generate a loop between the two. This loop has been demonstrated to be stable (Cheong, C. et al, 1990).

For each mRNA target, the sequence length to be used to simulate the duplex with miRNA-34a, was varied between 22-30 nucleotides. Each length studied always included the full seed sequence. The objective here is to use a mRNA length within the NMR size limitations which is representative of the miRNA.mRNA interaction. In order to choose a length to study, we want to keep the mRNA as short as possible but retaining the secondary structure involved in the interaction fig 6. A set of rules was then designed

when choosing a mRNA length: dangling nucleotides on the mRNA side are disregarded, selecting smaller lengths, unless they provide a 5´-GG-3' sequence; MFE (minimum free energy) structures free energy was compared between different lengths, choosing the lowest; preferably choose a length which presents the highest number of conformations present also on lengths beneath.

2-2.2) T7 transcription reaction:

For all reactions, annealing of the desired DNA template to the T7 promoter region (3'-ATATCACTCAGCATAATT-5') was done at 95 ºC for 5 minutes and then put on ice for 20 minutes, having [T7 promotor] and [DNA template] at 25 mM and [$MgCl_2$] at 2mM.

For miRNA-34a the transcription reaction used is as: 80 mM Tris, 20 mM $MgCl_2$, 60 mM DTT, 2 mM Spermidine, 0,5 U/µL RNAse Out, 20 % (v/v) DMSO, 24 mM UMP, 1 mM ATP, 3 mM of GTP/CTP/UTP, 2 µM of annealed DNA, 3% (v/v) T7 polymerase.

For gCD44 the transcription reaction optimized is as: 20 mM Tris, 12 mM $MgCl_2$, 30 mM DTT, 2 mM Spermidine, 16 % (v/v) DMSO, 32 mM GMP, 3 mM of ATP/GTP/CTP/UTP, 2 µM of annealed DNA, 2% (v/v) T7 polymerase.

For WNT1 the transcription reaction optimized is as: 140 mM Tris, 18 mM $MgCl_2$, 30 mM DTT, 2 mM Spermidine, 16 % (v/v) DMSO, 8 mM GMP, 3 mM of ATP/GTP/CTP/UTP, 2 µM of annealed DNA, 2% (v/v) T7 polymerase.

2-2.3) Analytical and qualitative denaturing polyacrylamide gel electrophoresis:

In order to verify the yield of transcription reactions, sample degradation and assess which fractions of the HPLC purification contain the desired product denaturing polyacrylamide gel electrophoresis is used. All gels used contain 8M Urea, 20% (v/v) polyacrylamide and 1X TBE. Gels are ran at 950V, 66 mA and 13W if only one gel is running in the gel case or 22W if two are running at the same time.

2-2.4) HPLC purification:

HPLC purification employed follows a 2 step separation, first using the Ion-Pairing Reverse Phase (RP-IP) column followed by an Ion-exchange (IE) column. The first step serves only for elimination of lower deletion products of the crude transcription reaction and the second step to provide near single nucleotide separation. It should be noted, that in certain cases, only IE seems sufficient.

In RP-IP in order to bind the RNA to the hydrophobic stationary phase, an amphiphilic ion pairing agent, Tetrabutylammonium Hydrogen Sulfate (THS). In order to elute the RNA the following buffers are used: Buffer A, 100 mM Ammonium Acetate, 2mM of THS, at pH 6,5; Buffer B, 4mM Ammonium Acetate, 0,4 mM THS, 80% (v/v) Acetonitrile, at pH 6,5. Total elution happens when increasing gradient of Buffer B is used, containing acetonitrile as organic solvent (Murray, J.B. et al., 1994). For miRNA-34a, flow was kept at 3 ml/min; collection period 20 seconds; delay volume of 1752 µL; percentage of buffer B was varied as: 7% from [0-15mins], 25-38,1% from [17-57mins], 100% from [62-72mins], 7% from [77-90mins].

In IE the negatively charged backbone of the RNA interacts with the cationic stationary phase through the phosphate groups. To promote elution, the following buffers are used: Buffer A, 20mM Sodium Perchlorate, 20mM Sodium Acetate, 10% (v/v) Acetonitrile, at pH 6,5; Buffer B, 600mM Sodium Perchlorate, 20mM Sodium acetate, 10% (v/v) Acetonitrile, at pH 6,5. Thanks to a high concentration of Sodium Perchlorate on Buffer B, separation is achieved (Oliver R. W. A. 1997). Heating the column to 75 ºC is used to improve separation to single nucleotide resolution, thanks to the unfolding of RNA. For miRNA-34a, flow was kept at 5,5 ml/min; collection period 16 seconds; delay volume of 1752 µL; percentage of buffer B was varied as: 0% from [0-15mins], 20-36% from [16-46mins], 100% from [47-62mins], 0% from [63-73mins]. For CD44, flow was kept at 5 ml/min; collection period 15 seconds; delay volume of 1752 µL; percentage of buffer B was varied as: 0% from [0-15mins], 20-31,5% from [16-43,6mins], 100% from [44,6-56,6mins], 0% from [57,6-73,6mins].

2-2.5) Analytical native polyacrylamide gel electrophoresis:

To asses if miRNA-34a and other target mRNAs formed a complex and its Kd a native gel was used with 10% (v/v) polyacrylamide and 1X TBE. Experiments were done while the gel was cooled through water circulation. Duplex formation was achieved by

heating the sample at 95ºC for 5 minutes followed by cooling at room temperature for 20 minutes.

2-2.6) UV-melting/concentration measurements:

For duplex measurements, annealing was done but heating the sample to 95ºC for 5 minutes and cooled down at room temperature for 20 minutes. For measurements of single mRNAs or miRNA-34a, the sample was heated to 95ºC for 5 minutes and cooled down in ice for 20 minutes.

For concentration measurements, temperature of the cuvette was kept at 25 ºC, having the sample scanned from wavelength of 220 to 350 nm, bandwidth 1 nm, integration time 0,08 seconds, data interval 0,5 nm and scan speed 375 nm/min. Blank was measure with nMR buffer: 15 mM NaP, 25 mM NaCl, 0,1 mM EDTA, pH 6,42.

For UV-melting experiences, the sample was analysed at the fixed wavelength of 260 nm. Temperature scan was from 10 to 75 ºC. Before each measurement at a certain point, the temperature was equilibrated for 1 minute. All melting data was fitted using Origin.

2-2.7) NMR methods:

All samples were prepared having the mRNA strand unlabelled and the miRNA strand with 15N and 13C labelled when both strands are annealed having the mix heated at 95 ºC and cooled at room temperature. When only mRNA is analysed, it is always unlabelled.

^{1}H and ^{13}C characteristic RNA chemical shifts regions can be found represented in fig.7 a) in a ^{13}C-^{1}H HSQC. The ^{1}H-^{15}N HSQC in fig 7 b), shows ^{1}H-^{5}N correlations for G and U iminos with characteristic chemical shift regions highlighted. The imino nitrogen chemical shift of a G can be found between 152-144 and for a U 163-155 ppms (Fürtig, B. et al., 2003; Wijmenga, S. S., & van Buuren, B. N., 1998). Imino correlations can be clustered in four quadrants as ^{15}N chemical shifts are nucleotide specific and ^{1}H and are reporters of the base-pair type. This results in G residues in the two upper quadrants and divided in canonical Watson-Crick base paired (left) and non-canonical or wobble base-paired (right). By analogy U residues are located on the lower two quadrands and dived in canonical (left) and non-canonical or wobble (right) base-paired. (Fürtig, B. et al., 2003). Counting the number of peaks and assessing how broaden they are on both HSQCs, can hint at how dynamic the duplex is. Counting the ^{1}H-^{15}N HSQC number of

peaks we assess how many base pairs are present at a certain temperature without the overlap problem on a simple ^{1}H 1D.

Proton chemical shift ranges can vary depending on the base pair being canonical or non-canonical. For a G, chemical shift can range between 12-13.5 pm and for a U between 13-15 ppms being considered highly probable at being a canonical WC base pair. If the chemical shifts are shifted to higher fields then 12ppm for G and 13ppm for U, the probability of it being a non-canonical WC base pair increases (Fürtig, B. et al., 2003). Imino protons close in space by a radius of ~5 Å can be determined by ^{1}H-^{1}H NOESY through crosspeaks (off diagonal peaks) which obtain chemical shift information of both nucleus. Crosspeaks on this spectrum can only be observed if the imino protons are involved in base-pairs and protected from exchange with the bulk solvent. In fig. 7 c), an "imino walk" representation is depicted, where a certain chemical shift can be associated to a certain imino ^{1}H on the sequence. The proton is associated to a certain base with the aid of a ^{1}H-^{15}N HSQC. The NOESY was used to assess the connectivity among imino crosspeaks and in this way, via an iterative process based on secondary structure predictions, secondary structure of the complexes studied was proposed (Fürtig, B. et al., 2003). This method of analysis was used for both miRNA-34a-gCD44 and miRNA-34a.PNUTS complexes.

All NMR samples are in NMR buffer with 10% (v/v) deuterated water.

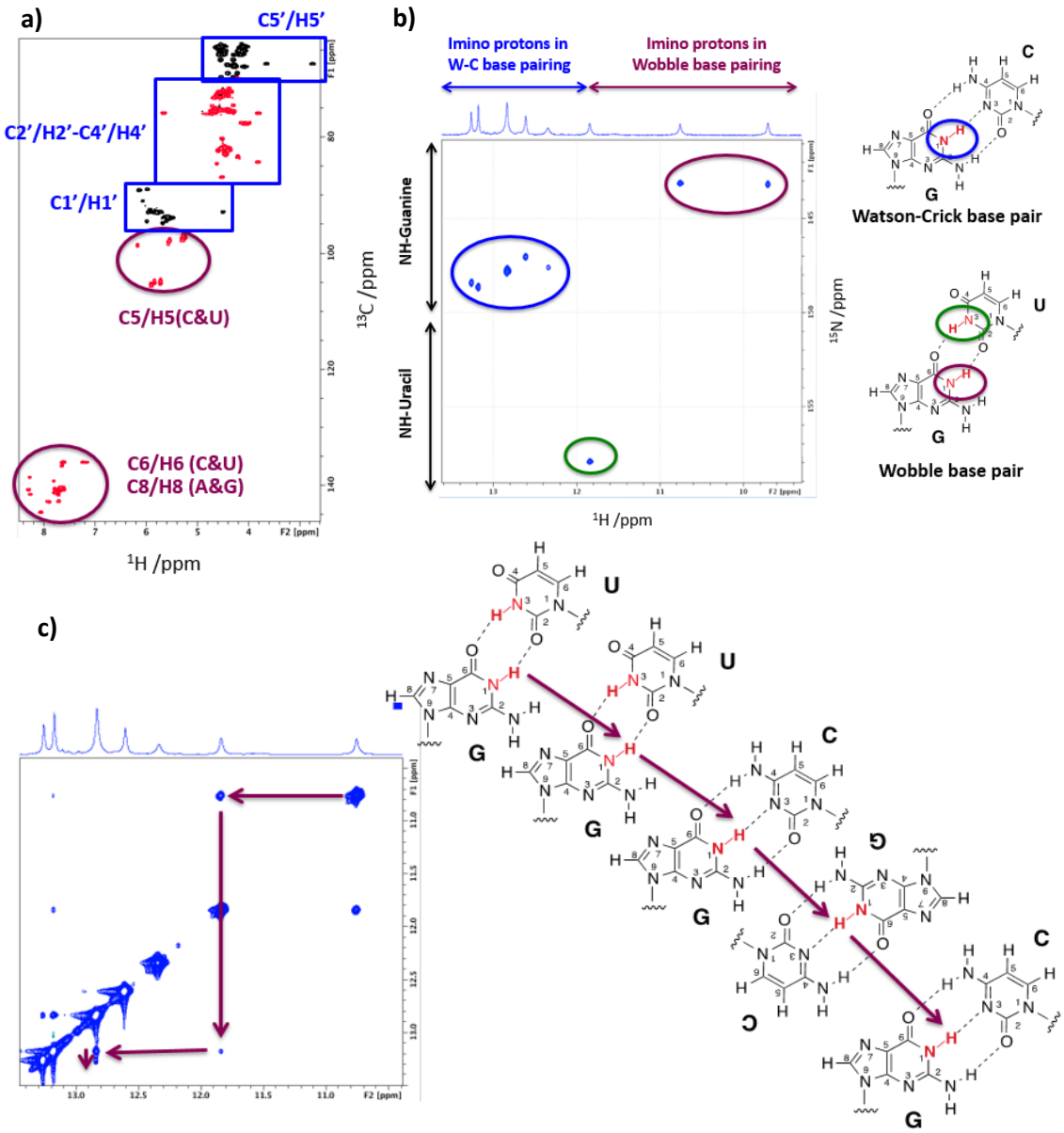


figure 7. a) ^{1}H-^{13}C HSQC demonstrations the several regions where it is more likely to find the chemical shift of: C5'H5', C2'H2', C4'/H4', C1'/H1', C5/H5, C6/H6, C8/H8; b) ^{1}H-^{15}N HSQC demonstrating the four quadrants in which we can divide this 2D experiment. Dividing Horizontally in half, on the top we have only imino protons of Gs while on the bottom of Us, whilst dividing in half vertically we have on the left WC base pairs and on the right Wb. At the right of the spectrum, we have a representation of two base pairs each having the imino protons colour coded according to the circled areas on the spectrum, blue for WC and green and purple for Wb imino protons; c) Sequential walk through the imino protons on a ^{1}H-^{1}H NOESY spectra (left), following connectivities through crosspeaks of protons close in space (right), which can either be base paired, in tandem or diagonally located.

3) Results and Discussion:

3-1) Transcription Optimization

3-1.1) WNT1

In optimization of WNT1 the same behaviour of reagent interdependence as observed in CD44 and gCD44 optimization. In both fig. 8 (left and right), $MgCl_2$ concentration is 24 mM and [Tris] is increased. A pattern can be observed that with the increase of Tris a hyperbolic depended of intensity is apparent (in light blue). Same can be seen for variation of Tris at [$MgCl_2$] of 18 mM. Difference is, at the same concentration of Tris at different $MgCl_2$ concentrations, for example, the third well of fig. 8 (left) and the fourth of fig. 8 (right), have very different intensity band patterns. This highlights once again, the interdependence problem. This could also translate into cases where DMSO and GMP, are at start point conditions which yield almost no main band product (fig. 8 right). A total of 68 reactions were tested as presented in supplementary material fig. 8. The optimized reaction as presented on methods, corresponds to the well with a red arrow in fig. 8 (right). All gels done for this RNA transcription reaction optimization are present in the supplementary material on fig. 21 and the concentrations used for each reagent on table 2 of the supplementary material.

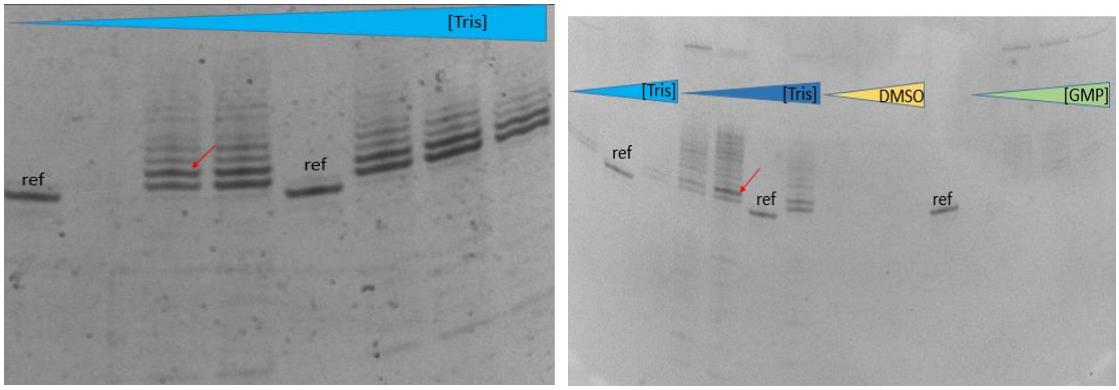


figure 8. 20% polyacrylamide denaturing gels of WNT1 optimization reactions as presented on table 2 on supplementary material. From right to left the reactions are: Gel on left) reactions 39-44; gel on right) 45-56. References presented are of 21 nts. On both gels we can clearly see the interdependence between all reagents. The reaction optimized as presented on methods, is pointed out by a red arrow.

3-1.2) CD44 and gCD44

Initially optimization of transcription of a small mRNA CD44 portion with the sequence "5'-GCCACUAUGUGUUGUUACUGCCA-3'", was attempted. No optimal condition could be found, as either low total yield or too many side products where found at any combination of to be optimized reactants. Hence we modified the construct. In fig.

9, it can be seen variation of DMSO and $MgCl_2$ concentrations at different starting point conditions, by which we mean, different concentrations of the other reagents

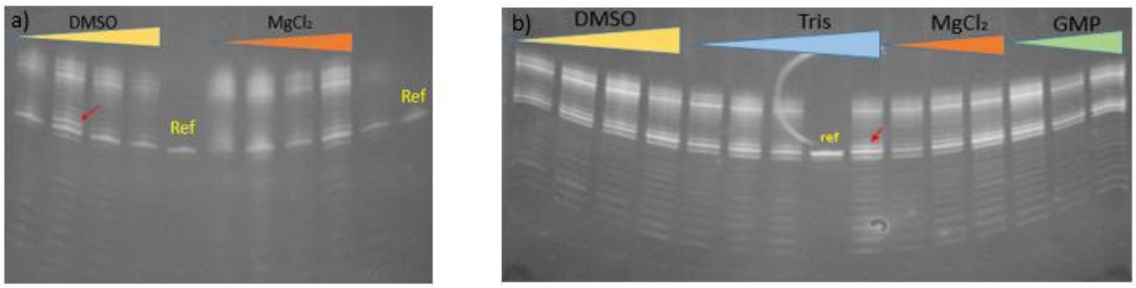


figure 9. 20% polyacrylamide denaturing gel of transcription products of CD44, having on a) reactions 22-30 and b) 8-21, as set on table 2 on the supplementary materials. The band corresponding to the RNA length we are trying to achieve is pointed out by a red arrow on both gels. On both we see that the -1 product is prevalent and a reagent interdependence is once again present. Reference is of 22 nts.

(concentrations pointed out in picture). It is clear to see that with different starting points, the effect of a certain reagent, at a certain concentration, can be quite different in terms of final product. A good example is the forth reaction from the left on fig. 9 a) and the first reaction in fig. 9 b). In both cases we have 14% v/v of DMSO, but [$MgCl_2$] is twice the one in b), which has a drastic effect on main band depletion, pointed out with a red arrow. Many more reactions were tested (Supplementary material fig. 19), but no significant optimization was reached in a reasonable time. In most cases for this sequence, either variation of a certain reagent did not have any apparent pattern on product influence at a certain start point (as shown in fig. 9 b) for Tris variation), or its variation had an incredible inhibitory effect on the main product (as seen in fig. 9 a) for $MgCl_2$ variation). In almost all cases, it was noticed that the -1 product was vastly produced.

In order to circumvent this nuisance, a G was added to the 5' of CD44 sequence, denoted as gCD44 for distinction, to augment transcription efficiency. In this case, it was obvious that an improvement has been reached as can be seen in figure 10, when varying [Tris] at different [MgCl2], as seen in light blue for low magnesium concentration and dark blue for higher, the -1 product is no longer preeminent, instead, both the main and the -1 are strikingly present. The amount of product of interest is much higher than in CD44 also. A total of 69 reactions were tested, which can be found on the supplementary material fig. 20.

In this chapter so far, it was made clear that there is interdependency between reagent concentration, but there is also sequence dependence. For reactions with the same reagent concentrations, WNT1 and gCD44 presented different band patterns, as can be seen for

reaction 46 of WNT1 (fig. 21 and Table 2 of supplementary materials) and reaction 3 of gCD44 (fig. 20 and table 2 of supplementary materials).

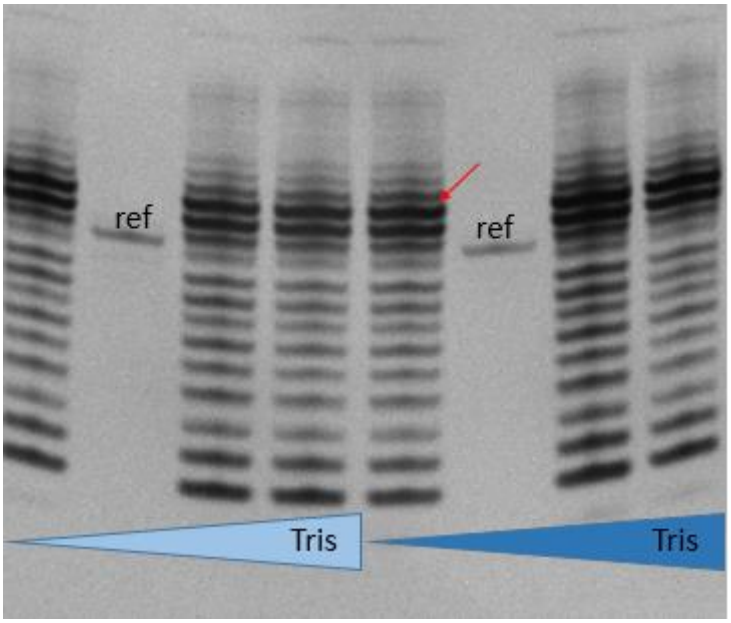
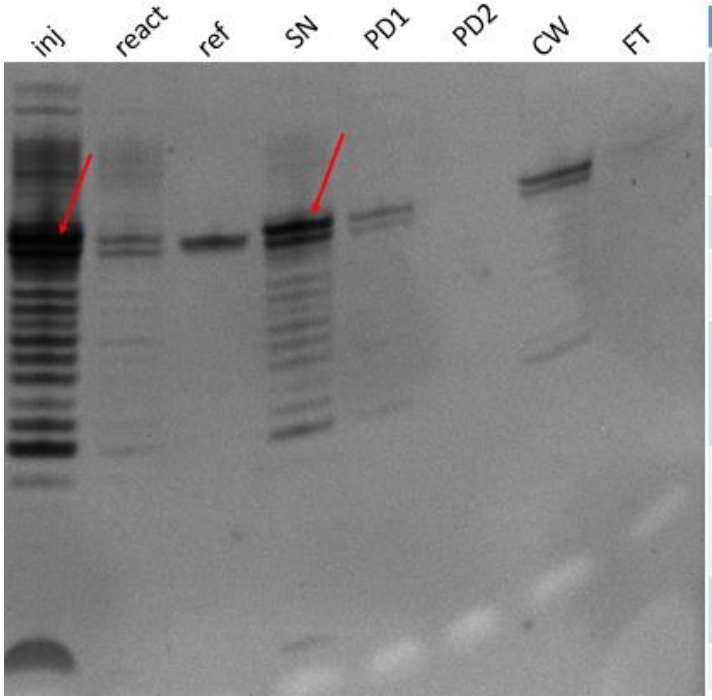


Figure 10. 20% polyacrylamide denaturing gel of transcription products of gCD44. Transcription reactions are from 11-16 (left to right) as described on table 2 in the supplementary materials. By varying [Tris] at at low [$MgCl_2$] (light blue) and high (dark blue), we observe that the -1 product is no longer the prevalent one and in both cases the amount of product of interest (pointed by red arrow) generated is much higher than CD44. Once again, reagent concentration dependence is apparent. Reference is of 22 nts.

3-2) Large scale transcription of miRNA-34a and gCD44

With set reaction, scaling up the volume is necessary for production of enough material for NMR, EMSA and UV melting studies. In fig.&Table 1, it can be seen the band pattern in a gel, from left to right, how a 10 mL reaction after being treated as discussed on the methods before the first injection on the HPLC. A comparison between a 50 uL and a concentrated 10mL reaction down to 500 µL can be seen on the same figure.



Gel name	Description
inj	Concentrated material for HPLC injection
react	50 uL transcription
ref	25 Nucleotides
SN	Supernatant after obtaining the pellet
PD1	Supernatant after dissolving the pellet with NMR buffer with 50 mM EDTA
PD2	Supernatant after dissolving the pellet with NMR buffer with 50 mM EDTA
CW	Concentrator wash after concentrating
FT	Flow through

Figure & Table 1. 20% polyacrylamide denaturing gel (left) displaying the products of each step for sample preparation before the first injection on the HPLC column. Red arrows point to the product of interest (gCD44). Comparing the concentrated material of a 10mL reaction to a 50 uL reaction we can confirm the scale up in product and relative band intensity which assures diminished amounts of pipetting errors. We can observe that a considerable amount of material of interest can be found in the pellet and membrane of the concentrator.

We can observe that when concentrated, the other side products are produced besides the main band and the -1 product, but at lower quantities. It's also worthy to point out the major improvements of this reaction vs the previously shown here (fig 10) and in the supplementary material (fig. 20). The optimized reaction is the only one that displays a stronger band of the product of interest relative to the -1 product band. When the transcription reaction occurs, pyrophosphate is produced, and when spinned down on the centrifuge, it forms a white pellet. In fig.&Table 1, is displayed how important it is to dissolve the pyrophosphate pellet and wash the membrane of the concentrator after the concentration step, since some material could be trapped on either the pellet or the concentrator membrane.

To purify gCD44, only IE column was used, as displayed on fig. 11 A, the chromatogram of a single injection of the concentrated 10 mL reaction. A near one nucleotide resolution was achieved as can be seen on fig. 11 B, where a very small contribution from the -1 product to the product of interest can be observed between 35,5-35,75 minutes. Beyond 35,75 minutes, only pure product seems to be present. Pooling fractions 81-87 and concentrating down to NMR volume, 220 uL, we obtained 0,57 mg with a concentration of 360,06 uM, delivering an yield of 6,34%. 0,5 uL of the final purified material was ran in a denaturing gel to assess the level of purification. In fig. 11 C a -1 product is visible after purification and concentration, in conjunction with an addition product which was not visible in fig 11 B, maybe due to dilution of each fraction. On the NMR sample point of view, the purification is considered acceptable to work with, since the main band is at a considerable high ratio of intensity relative to the two others.

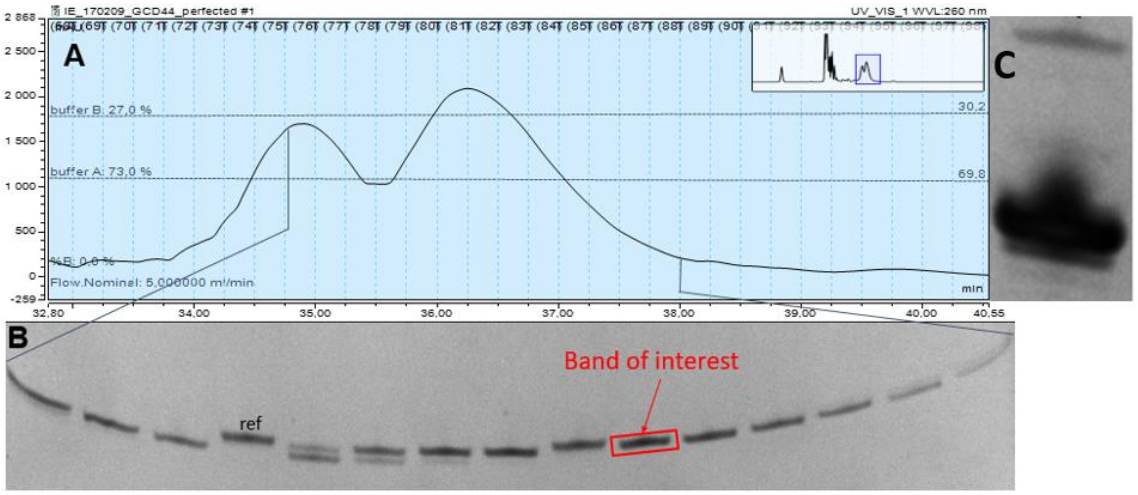
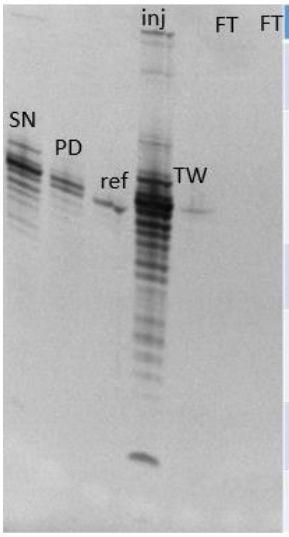


*figure 11. **A)** Chromatogram of purification of gCD44 in Ion Exchange HPLC, using the method desbribed on Materials and methods; **B)** 20% Denaturing polyacrylamide gel of the selected fractions on image **A**, displaying a near single nucleotide resolution, using a reference of 25 nts. **C)** Pooling the fractions in-between the 5th and the 12th band in the gel in B), after concentrating the sample to NMR volume (230uL), we can see an overload of main product as a consequence of high concentration. A -1 product and a nt addition product can be seen. But since the ratio between main band vs side products is high, it is viable for NMR studies.*

On another note, miRNA-34a requires purification with both RP and IE columns to achieve acceptable separation. In order to have good NMR signal, miRNA-34a is ^{15}N and ^{13}C labelled on every nucleotide, hence the importance on this case to retrieve as much product as possible. As described in the methods, in this specific case, the reaction is split into several Eppendorf tubes and then each tube is washed three times with water which in turn is then pooled into the concentrator before RP purification. In Fig.&Table 2 it is stressed again the importance of pellet dissolving but also demonstrated the product that

is retrieved from washing the Eppendorf tubes, since some RNA could be attached to the wall of the tube.



Gel name	Description
SN	Supernatant after obtaining the pellet
PD	Supernatant after dissolving the pellet with NMR buffer+EDTA 50 mM
ref	21 nt reference
inj	Concentrated material for injecting on IP_RP
TW	Material in one tube after 1st H20 wash
FT	Flow through

Figure & Table 2. 20% polyacrylamide denaturing gel (left) displaying the products of each step for sample preparation before the first injection on the HPLC column for miRNA.34a. Material of interest was recovered from the pellet by dissolving it. Here it is specially stressed how important is to wash the Eppendorf tubes with water, there can always be material on the walls of the Eppendorf.

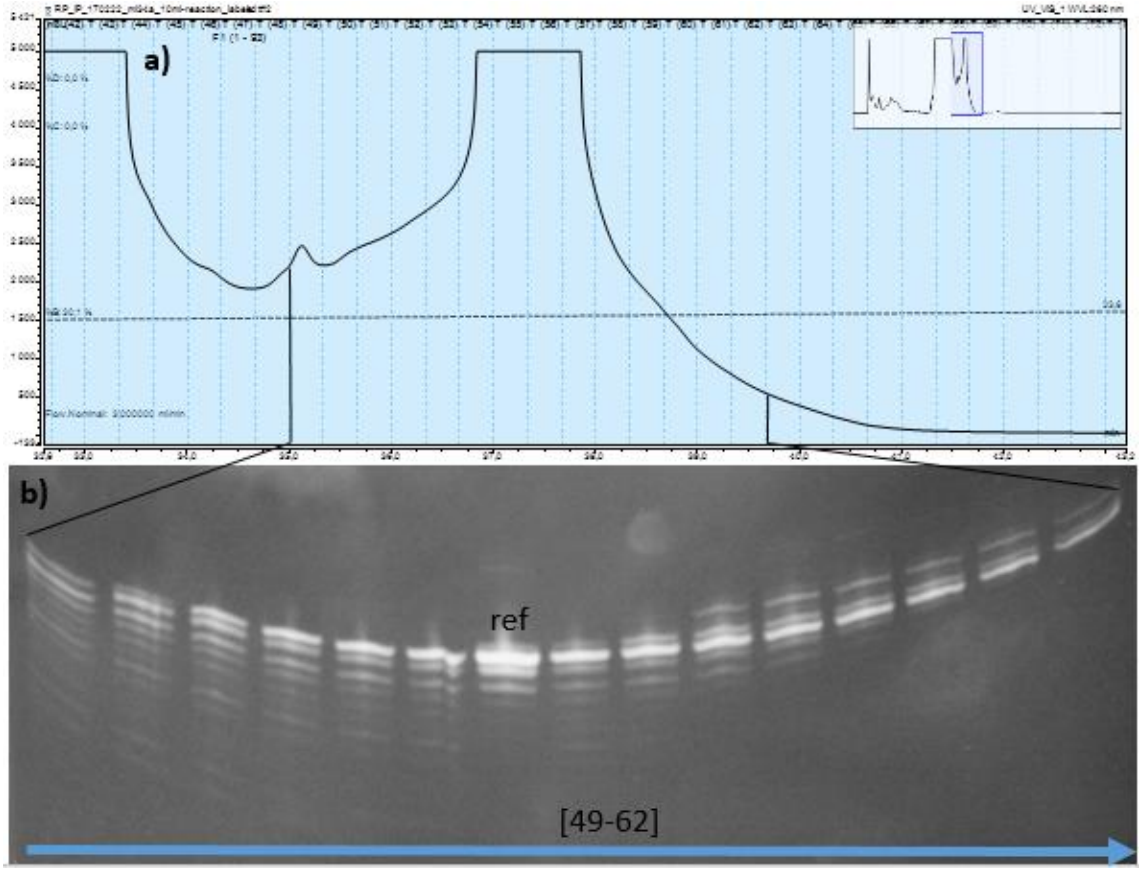


*Figure 12. **a)** Chromatogram of purification of miRNA-34a in a RP column, using the method described on Materials and methods; **b)** 20% Denaturing polyacrylamide gel of the selected fractions on image a), displaying a gradual ladder type separation. Separation of smaller sider products is achieved but not bigger side products. Reference used on the gel is 22 nts long.*

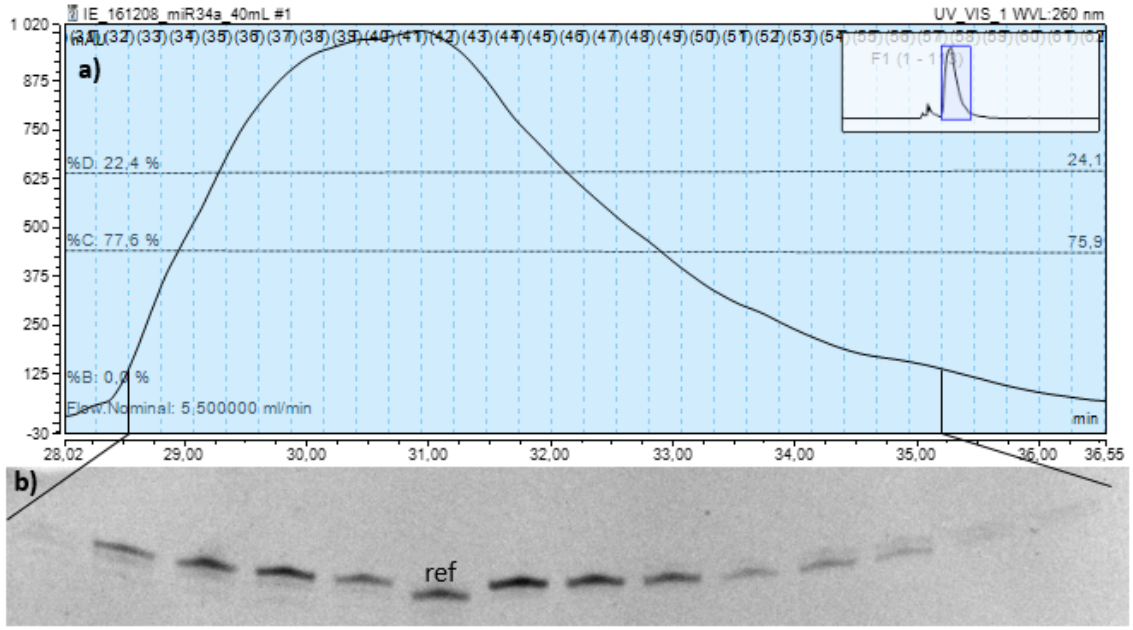


*Figura 12. **a)** Chromatogram of purification miRNA-34a in Ion Exchange HPLC, using the method described on Materials and methods; **b)** 20% Denaturing polyacrylamide gel of the selected fractions on image **a)**, displaying a near single nucleotide resolution, using a reference of 22 nts. Here separation of the remainder longer side products is achieved and other smaller side products.*

Fig. 12 a) displays the peak of interest on RP for a 10mL large scale of miR-34a, for an injection equivalent to 5 mL. The separation seen in b) is worse in comparison to gCD44 IE purification, but crucial to discard of lower products. Distinctly from gCD44 purification, this first purification step for miRNA-34a shows a step wise separation where the band of interest seems to be dragged out through the peak, while side products fall a pattern of separation according to its length. A final purification step requires pooling of fractions 51-62, concentration and separation with IE column. Only after IE purification preceding RP separation, we can obtain a near one nucleotide separation comparable to gCD44 purification method (fig. 13 b)). It is interesting to note that even though separation is comparable, the chromatogram peaks are not. On gCD44 we see two clear peaks for each band, -1 and main band, while for miRNA-34a just one broadened band, which from the gel we can confirm that it contains both the -1 and the main product. This dissimilar event happening on the same column, which requires a different slope for buffer B flushing during separation, could only hint that, even though the conditions are similar, the separation is RNA sequence dependent. Otherwise the same slope for Buffer B flushing could be used for all cases at the same conditions, which is not the case. It is of value to suggest that, in the context of IE HPLC separation, the temperature is also an important factor. This is due to different RNA sequences originating different structures with different melting temperatures, hence different stabilities. Following this line of logic, probably a more stable hairpin loop would either require a higher temperature for the same slope, or a steeper slope for the same temperature. In the future, a simpler separation for miRNA-34a would be useful time wise. Hence it would be interesting to try different gradients for IE only separation on miRNA-34a at different temperatures, or better yet, use a more aggressive denaturing agent like formamide or urea. This process has a yield of 2,5% producing 0,25 mg of pure material.

3-3) EMSA and UV-melting studies on miRNA-34a.gCD44 and miRNA-34a.PNUTS complex

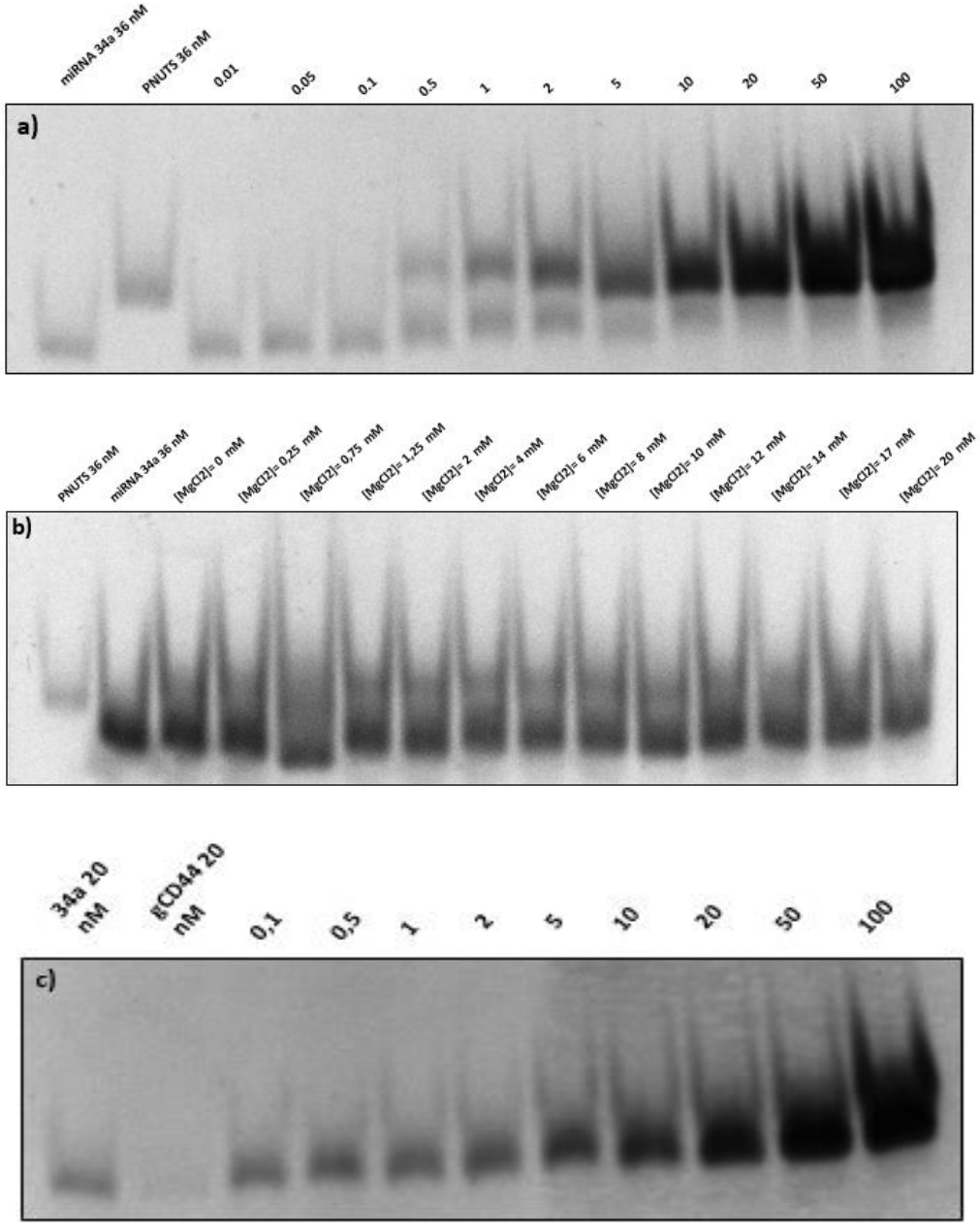
Biophysical studies were conducted on two mRNAs: gCD44 and PNUTS, free and in complex with miRNA-34a (it is of importance to point out that in this work, transcription optimization, purification and production of PNUTS, NMR titration and a single EMSA, were done by a previous student, Karen Scheriever). In my studies, experiments such as EMSA, UV melting and NMR 2D and 1D were applied. The order of experiments should be maintained because of the possibility that the molar ratio

necessary to for formation of a helical duplex between the mRNA and the miRNA, being different then 1:1. EMSA is used to assess such questions, while UV melting duplex stability.

EMSA experiments on miR-34a.PNUTS complex done by Karen displayed the possibility of a highly unstable complex, where Kd couldn't be measured. We tried to stabilize the duplex by the use of $MgCl_2$, which is a key factor for duplex formation in this case. Fixing [$MgCl_2$] physiological concentrations, 5 mM, we varied PNUTS concentration maintaining [miR-34a] constant (figure 14 a)). For the first time, it was noticeable a band with a slightly slower run time then PNUTS, hinting at possibly being the complex. A step wise upwards shift through different ratios of the free miRNA-34a band were apparent, which could be explained if the complex would be transient, where the two strands would be constantly binding and separating. Unfortunately, Kd could not be measured since there was still not a good distinction between bands. A final EMSA was also done on this complex, fixing both miRNA and mRNA concentrations at 36 nM and varying $MgCl_2$ concentration, as showed in fig. 14 b). No obvious variation in band run time was noticed for this 1:1 ratio experiment, besides a slight upwards shift on the bands with concentration higher than 0,75 mM, possibly pointing out the minimum magnesium required for duplex stabilization. This can already hint that this duplex will not have many imino crosspeaks present on the ^{1}H-^{1}H NOESY since it shows to be very unstable.

MiRNA-34a.gCD44 EMSA essays exhibited yet another transient duplex, which the Kd could not be measured due to the same events present in miR-34a.PNUTS essays, except even more attenuated. Two EMSAs were done for this complex at different fixed miRNA concentrations, 20 nM in fig. 14 c) and 20 uM in fig. 14 d). In both EMSAs we can see a slight upwards shift on the single band observed for every well in which titration occurs. In c) however, there seems to be a better visualization of this shift then in d), which could be due to less overload. EMSA d) on the other hand shows us at ration 1:50 and 1:100 what it seems to be either overload, or a distinct duplex band. Interestingly, no free miRNA band is observed, which suggests that this duplex, even though being transient, is more stable then miR-34a.PNUTS duplex.

UV melting data presented in fig.14 e) allowed us to clearly compare stability of the hairpin loops of miRNA-34a, PNUTS and gCD44, proposing the following stability order from more stable to less: miRNA-34a > PNUTS > gCD44. Tm goes in the following order respectfully: 36,1±0,1; 31,1±0,1; 30,6±0,1 ºC.



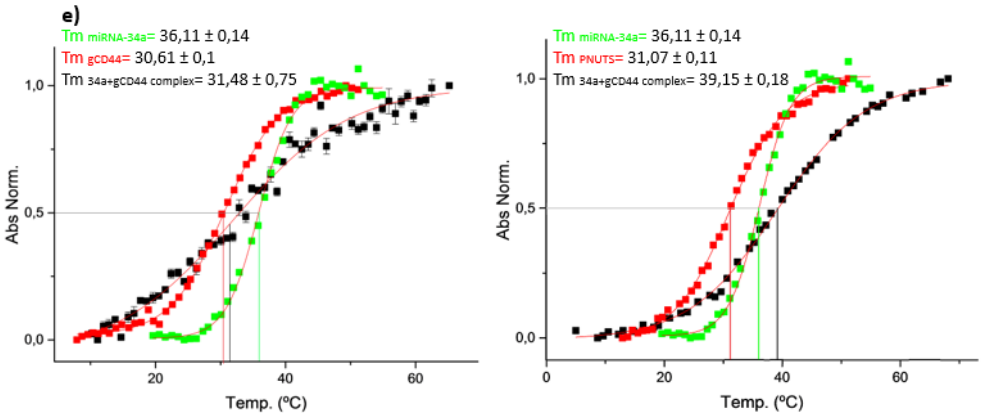


Figure 13. Native 10% (v/v) polyacrylamide gels for both miR-34a binding to PNUTS a) & b) and gCD44 c&d), where in the latter the concentration of both RNAs is lower in ***c)*** *compared to* ***d).*** *References, are always positioned in the second well from the left. Molar ratios of mRNA are represented for each well, above gel* ***a), b)*** *and* ***c)****. In* ***a)****, a fixed concentration of [$MgCl_2$] = 5 mM was maintained while titrating miRNA with PNUTS. In* ***b)*** *a titration of $MgCl_2$ was done while maintaining the ratio of miRNA:PNUTS at 1:1. No conclusions could be made in* ***b)****, but in* ***a)*** *we observe a slight movement of the miRNA-34a band upwards along the titration hinting at the duplex being transient.* ***c&d)*** *The same can be said for miRNA-34a.gCD44 complex, a shift in the band can also be seen upwards hinting at also being transient, except in case of* ***b)*** *in order to have visible complex formation $MgCl_2$ was needed;* ***e)*** *Melting profiles for free gCD44 (red), miRNA-34a.gCD44 complex (black) are presented on the left and for free PNUTS (red) and the complex miRNA-34a.PNUTS on black on the right, while free miRNA-34a (green) is present on both. Both duplex experiments were done at a ratio of 1:1 (miRNA:mRNA). The hierarchy of stability assessed through Tm for free systems is: miRNA-34a > PNUTS > gCD44 and for complexes: miR-34a.PNUTS > miR-34a.gCD44.*

Comparing gCD44 and miRNA-34a Tms, supports the fact that gCD44 can be easily purified with just IE thanks to column heating at 75 ºC. But it also suggests that miRNA-34a could be too, hence it's just a matter of testing the right method with the present buffers.

Comparing the specific duplex stabilities however is quite complicated. From the EMSAs we know that the ratio needed for complex formation is higher than 1:1. Contributions from both free mRNA and miRNA are then present on the measurements which means that the values of Tm for miR-34a.PNUTS (Tm= 39,2±0,2) > miR-34a.gCD44 (Tm=31,5±0,8) don't uniquely represent the events we specifically want to probe. Besides, two other factors must be considered: PNUTS and miRNA-34a are the most stable hairpin loops out of the three which could elevate the Tm more than the duplex really has; miR-34a.gCD44 duplex melting curve is not complete on the lower and upper base lines, because of UV machine limitations influencing the fit negatively.

3-4) NMR biophysical studies on miR-34a.PNUTS & miR-34a.gCD44 duplexes

Titration of labelled miRNA-34a with unlabelled gCD44 was done while monitoring the disappearance of free miRNA-34a H8/H6 peaks on an aromatic HSQC and the appearance of a new set of peaks of the duplex at 298K. This experiment was done in NMR buffer (see materials), which allowed us to probe the real ratio: 1:8,2 miRNA-34a/gCD44 (fig. 15 a), blue is free miRNA and red miR-34a.gCD44 fully formed duplex).

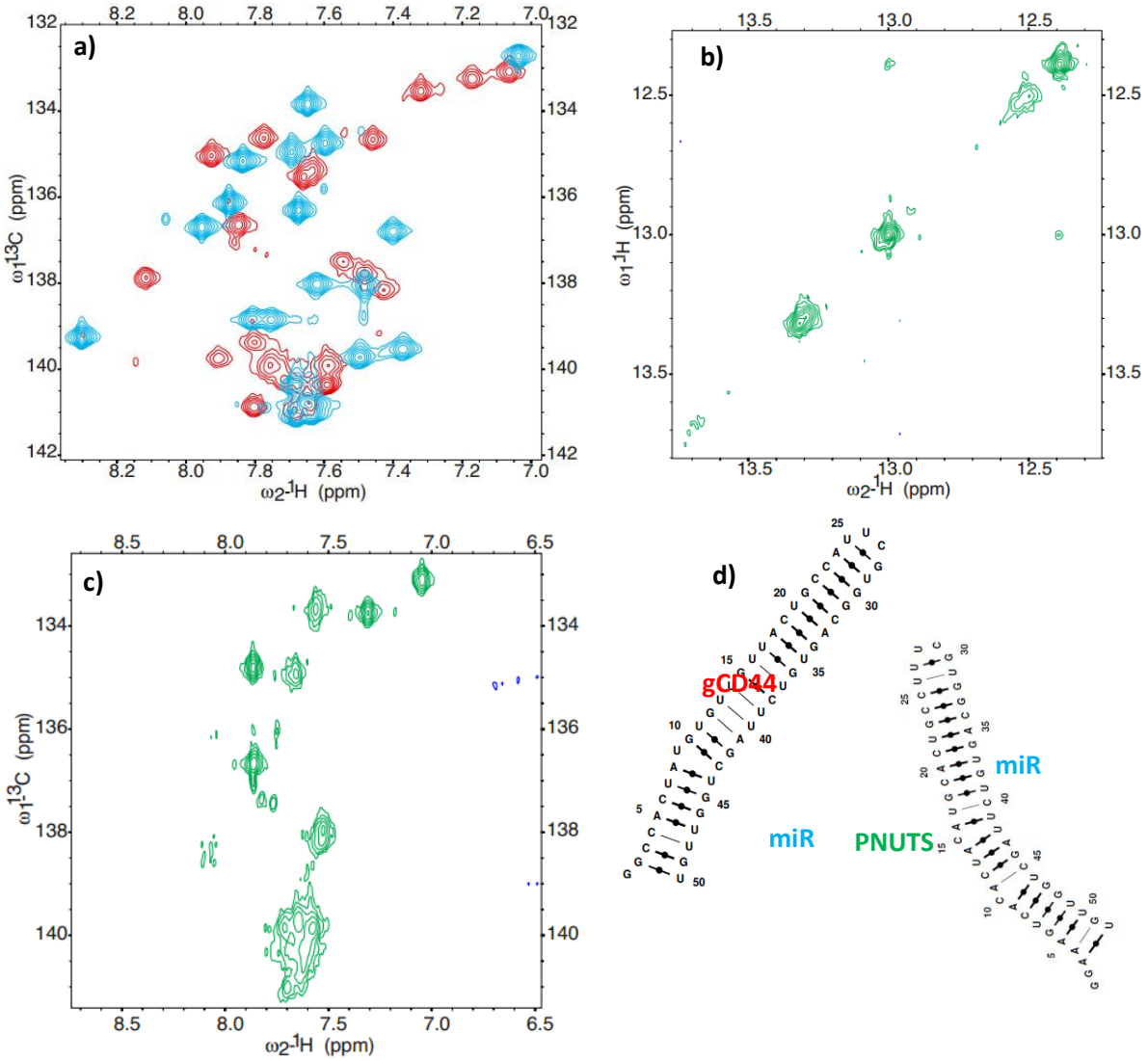


*Figure 14. a) Presented here are ^{1}H-^{13}C8/6 HSQCs of free miRNA-34a and **miRNA-34a bound to CD44** (molar ratio of 1:8.8 of gCD44 relative to miRNA-34a). Addition of mCD44 leads to a complete change of structure of miRNA-34a. This can be observed by a complete new set of peaks; b) ^{1}H-^{1}HNOESY 175ms mixing time. As can be seen, only a very faint crosspeak, close to noise, on the imino region is visible, miRNA-34a-mPNUTS complex hints at being dynamic, since this area has a set of broadened peaks. The molar ratio of miRNA-34a:PNUTS is of 1:2; **c)** ^{1}H-^{13}C 8/6/5 HSQC HSQC of the complex miRNA-34a-PNUTS, displaying the structural instability of the duplex, not only for having a reduced amount of total peaks but also an extensive area of broaden overlapped peaks; d) Simulated structures for both duplexes with MCfold (Parisien, M., & Major, F., 2008). Each strand is properly noted and color coded.*

A ratio of 1:2 of miR-34a/PNUTS for full duplex formation was observed. Regarding its NOESY 175 ms at 298K only one very weak imino cross peak is detected close to the noise, fig. 15 b). Besides that, its H8/H6 HSQC, fig. 15 c) has not only less peaks then miR-34a.gCD44s HSQC, fig. 15 a) red spectra, but much more broaden out. This tells us that miR-34a.PNUTS complex is likely very dynamic and probably most of its iminos are exchanging with water. From the MCfold predicted structures, fig. 15 d), we could count the expected number of peaks for each duplex (22 each). For miR-34a.gCD44 19 are countable and for miR-34a.PNUTS only 7 are countable excluding the big superposition of broaden peaks. These spectra were recorded with the following concentrations: [Free miR-34a]=342 uM; [miR-34a.gCD44]= [miR-34a] + [gCD44], [miR-34a]= 149,33 uM, [gCD44]= 555,98 uM; [miR-34a.PNUTS]=[miR-34a]+[PNNUTS], [miR-34a]=131,3 uM [PNUTS]=262,6 uM.

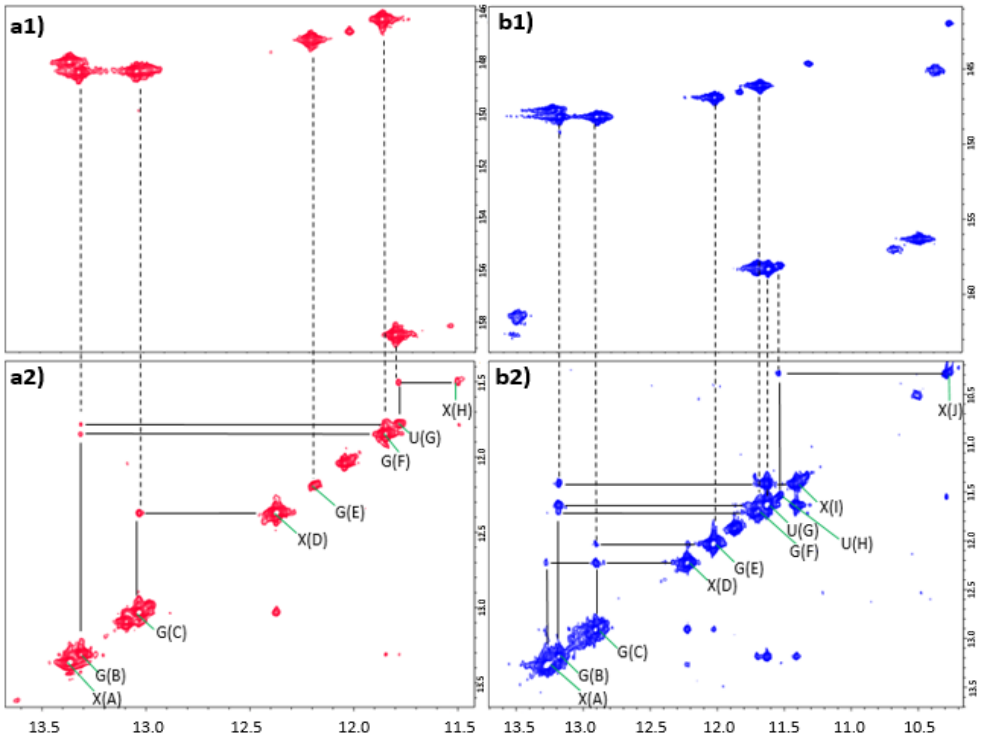


Figure 15. A ^{1}H-^{1}H NOESY (a2 & b2) and ^{1}H-^{15}N HSQC (a1 & b1) were recorded at two temperatures, 298K (red) and 284K (blue) for miRNA-34a.gCD44 complex. At both temperatures connectivities were assigned on the NOESYs (full lines), while assessing which proton frequencies correspond to the labelled strand (miRNA-34a), by checking the HSQC (dashed lines). Depending on which quadrant the corresponding peak is on the HSQC, it was assigned on the NOESY its identity, either a G or U. If no corresponding peak, then "X" was assigned. As expected in b1) & b2) there are more peaks then in a1) & a2).

From just the MCfold simulation we could assume wrongly that miR-34a.PNUTS is more stable just from observing the free energy. We have to take into account here that the predicted free energy by MCfold and the high number of WC base pairs on the

simulated structure, are both formulated with a certain number of assumptions. Even though the number of WC base pairs hints at the stability it is not a rule set in stone as discussed previously. This software does not take into account the positional influence of bulges. As mentioned before, loops in the middle of the duplex are destabilizing, especially if there are two. This could be an explanation to why the duplex is very unstable regardless of ratio between RNAs required for its formation.

Starting from the principle that the predicted structure for miR-34a.PNUTS is real, then it would make sense according to previous studies (Hermann, T., & Patel, D. J., 2000) that $MgCl_2$ would possibly stabilize these loops, hence the duplex. Also, mcfold doesn't consider the stabilizing effect of overhang nucleotides. It would be quite important if the bulges would be stabilized by magnesium, because then the contribution of the overhangs to duplex stability would probably be more significant (Miller, S. et al., 2008). This comes into accord with the EMSA displayed on gig.X a).

Assignment of the imino protons for miR-34a.gCD44 complex was achieved using both a NOESY with 175 ms mixing time and an ^{15}N-^{1}H HSQC recorded on the imino region. This was achieved at both at 298K and 284K, using the latter temperature to achieve better peak separation in case of peak overlap at 298K. Measuring at lower temperature not only enables us to have a more stable duplex but unfortunately increases tumbling of the molecule, hence line broadening can happen. In figure 16 an imino walk is presented for 298K (in red) and 284K (in blue). Since the duplex is more stable at 284K, assignment was attempted at that temperature first. ^{15}N-^{1}H HSQC, confirmed which iminos were from miR-34a, and its nitrogen chemical shift confirmed to which nucleotide it belonged to, either a G or a U, being a G between 152-144 and a U 163-155 (Fürtig, B. et al., 2003; Wijmenga, S. S., & van Buuren, B. N., 1998). Through its proton chemical shift, we can also extrapolate which type of base pair it is, for a G, if the chemical shift is between 12-13.5, there is a very high probability of being a WC base pair. The same can be said for a U, if it is between 13-15. If the chemical shift of said nucleotides are shifted to higher fields then 12ppm for G and 13ppm for U, the probability of it being a non-canonical base pair increases (Fürtig, B. et al., 2003), however a certain overlap around 12-13ppm can be expected, e.g. for WC base pairs next to bulges.

By counting peaks on a1) and b1), we can already confirm that at 284K the duplex is more stable since we can count nine peaks in a1) and seventeen in b1), being five WC and three non-WC in a1) and seven WC and ten non-WC in b1). In fig. 16, on both NOESYs a2) and b2), iminos that belonged to the unlabelled strand (gCD44) were denoted by an X(Y), being X an unknown nucleotide identify that bears said imino and Y being a certain letter given to the peak in order for easy distinction from others. For iminos connected to a labelled N (miR-34a strand), X in X(Y) was attributed to either G or U, thanks to recognition of its identity with the ^{15}N-^{1}H HSQCs a1) and b1). We should note, we only attempt to give a X(Y) identity to those peaks that had visible cross peaks in the NOESY. After identifying labelled iminos and all possible connections at 284K, the same was attempted at 298K. In order to do so, several 1Ds were recorded from 284K back to 298K to follow the peak shifts. In fig. 17 a), b) and d), each at different ^{1}H ranges on the imino reagion, we can see represented three 1Ds at different temperatures: blue 284K, purple 292K, red 298K. It is clear not only the peak shift, but separation as we lower the temperature, or overlap as we increase it. It is the case of overlap at 298K for peaks: X(A) and G(F). We can also observe new peaks arising, such is the case in a), peak at lowest field and d), the peak at lowest field and an extra peak which become apparent at 292K and starts to broaden and merged with its neighbour, denoted as X(I) through analysis of b1&2). It was observed in a1) of fig. 16 that a single weak cross peak was present between two unlabelled iminos. In fig. 17 c), overlapping the NOESY at 298K with a 1D of only gCD44 at 1,2 mM, in purple, we can see that cross peak corresponds to two iminos close in space of the free gCD44 that is unbound to miR-34a. That peak can also be followed back to 284K through observation of peak shift along the temperature gradient in fig. 17 a) and b), indicated in bold. It is interesting to point out that, only at 284K, certain NOE connectivity's were captured, such as: X(D) to X(A); X(D) to G(C) and X(J) to U(H). On

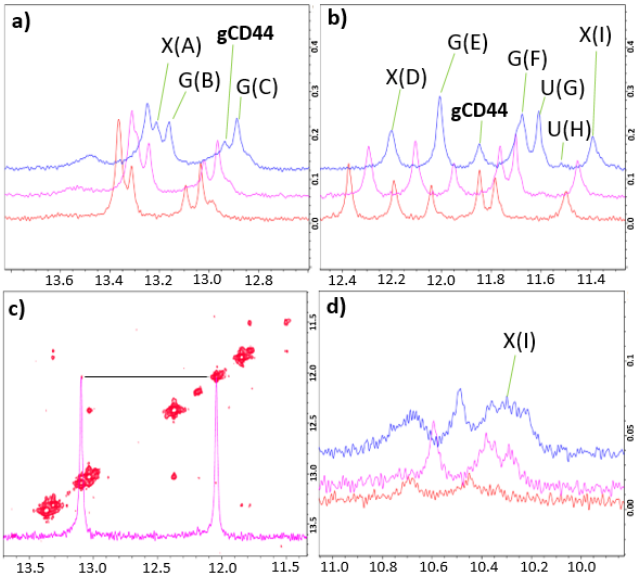


*Figure 16. **a,b,d)** Several ^{1}H 1Ds were recorded at 284K (blue), 292K (purple), red (298K) for miR-34a.gCD44. From the peaks assigned in figure 16 a2) & b2) at 284K, the shift of the peaks from 284K to 298K was followed. **C)** ^{1}H-^{1}H NOESY of miR-34a.gCD44 at 298K (red), was superimposed with a ^{1}H 1D from free gCD44, unveiling a crosspeak on the NOESY corresponding to free gCD44 and not to the complex.*

the same note, only at lower temperature was U(H) imino peak present on the ^{15}N-^{1}H HSQC.

At all times we should keep in mind that imino to imino NOE connections can only be captured if the iminos are from base paired nucleotides; if they are in tandem, on the same strand; or diagonally placed on the close neighbouring base pair on the opposite strand.

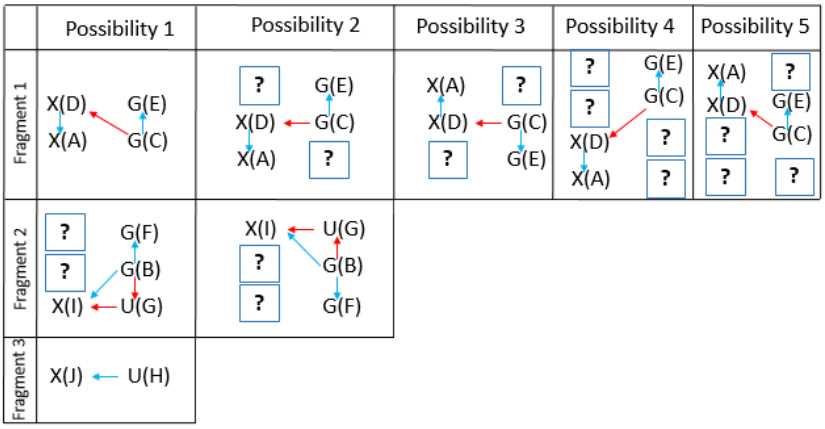


Table 1. For each set of connectivities captured on the NOESY at both 298K and 284K (fragment), a set of possibilities is presented. Red arroes indicate a crosspeak captured at both 298K and 284K and blue arrows, crosspeaks only captured at 284K. On the left side of the fragment is the gCD44 strand and the right miRNA-34a.

This places some constraints on the possible base pair patterns that can be extrapolated from the NOE connectivity's. In table 1, all possible combinations are presented for each systematic string of connections. At all times, on the right represented the unlabelled strand (gCD44) and on the right the labelled one (miR-34a). Arrows in blue represent NOE connections captured at 284K only and in red at both 284K and 298K, and the boxes in blue with a question mark represent any possible nucleotide. Fortunately, the sequence of both strands in the duplex are known, so connections between labelled nucleotides with revealed identities with the ^{15}N-^{1}H HSQC, can be matched to certain parts of the sequence. Using mcfold, a total of hundred structures were simulated, with only seven not being similar structurally between each other. In fig. 18 a), all seven structures are presented. Each of the seven structures was scanned for matches for all fragments and all its possibilities. Matches were highlighted with coloured boxed on the structures, being red fragment 1, blue fragment 2 and green fragment 3. In all structures, it seems obvious the most probable possibility for fragment 1 is number 4, which can be found in the seed sequence. For fragment 2 the number of possibilities was diminished to two by assuming, with great certainty, that X(I) (identified as X(H) at 298K) is base paired with U(G). This is due to the fact of the extremely intense NOE cross peak observed for

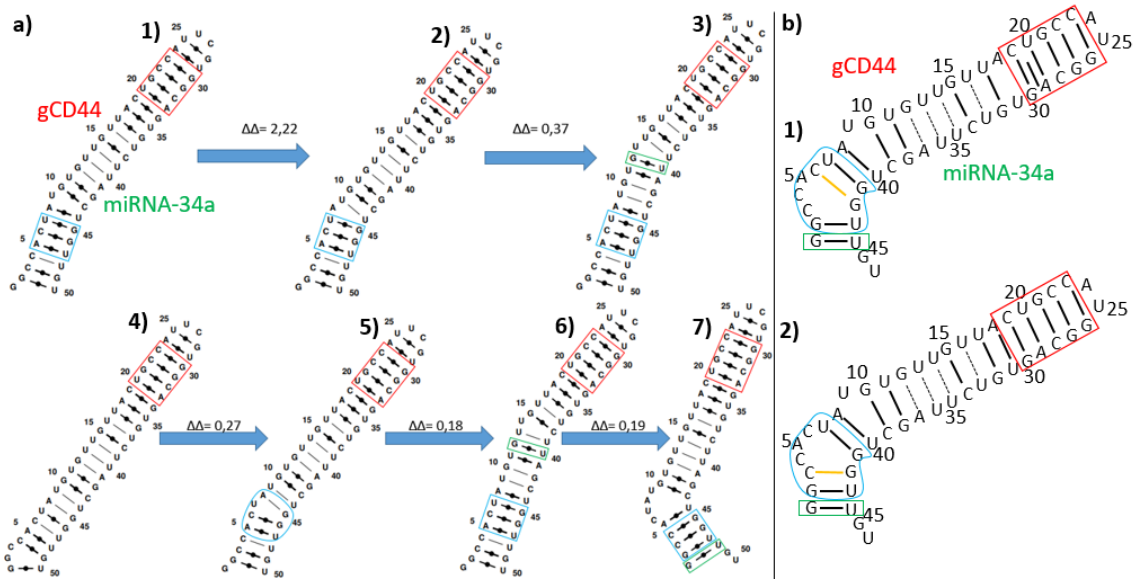


figure 17. ***a)*** *Exploring all possibilities for each fragment constructed (Table 1), we try to fit those fragments into the MCfold simulated structures (Parisien, M., & Major, F., 2008). For this task, 100 structures were explored and only 7 were dissimilar. The difference in free energy between structures is represented. Fragment 1 (red) had matches in all 7 structures, fragment 2 (blue) had matches in all except structure 4 and fragment 3 (green) had matches only on structures 3, 6 and 7;* ***b)*** *Proposed structures 1 and 2, not captured on MCfold simulations, have matches with all fragments. Difference between 1 and 2 is a shift of base pairing from G42 between C3 and C6 (base pair highlighted in orange on both 1 and 2).*

the connection between both. Unlike fragment 1 however, not all fragment 2 chemical shifts for the imino protons are within the designated high certainty WC range. Due to this, uncertainty if G(F), U(G) and X(I)/X(H) are iminos within WC or non-WC base pairs arises, leading to an array of possible matches within the simulated structures. In all structures except a4), there was a possible match. In all of the rest however, the match is only possible at the base of the stem, where there is a sequence of 3'-GGU-5' on the labelled strand, making possibility 1 on table X the only match, since 3'-UGG-5' does not exist. Finally, we have fragment 3, with only one possibility. The chemical shift range in which U(H) is present is very likely to be a non-WC base pair, either a Wb or a Mm. Since Wb is in general more stable considering only its identity, we can presume that, U(H) belongs to such pairing, being only present in a3), a6) and a7). It could be argued that U(H) would belong to a Mm, especially tandem UU Mms in a1), a2), a4), a5) and a7), since tandem UU Mms are in general the most stable between possible combinations of tandem Mms (Martha E. Christian & Brent M. Znosko, 2008, Biochemistry). Shankar and co-workers in 2007 captured imino chemical shifts for a UU base paired Mm, between 12-10 ppm range, where both U(H) and X(J) chemical shift falls. With no other imino cross peak connections present, this could also be taken into account, but for now, one the simplest and most probable solution is explored. Further connectivity's should be probed in order to prove a more unlikely hypothesis. Unfortunately for the simulations

ran with mcfold, not all possibilities were explored. In fig. 18 b1&2), we propose an additional possibility for secondary structure, similar to a7), but with an additional bulge. This comes to the fact that in none of the simulated structures, there was a match where U(G), G(F) and X(I)/X(H), all participated in non-WC base pairs. Even though easily assumed that this structure is unlikely the MFE from lack of representation in the simulated structures, in reality it is not the case. Mcfold does not take into account the contribution 3' or 5' dangling nucleotides to stability, position of the bulges, the distance between each other, salt interactions with bulge nucleotides and interactions between bulges such as base stacking of free bases (Hermann, T., & Patel, D. J., 2000; Clair V. Crowther et al, 2016, RNA Society; Parisien, M., & Major, F., 2008). It is worthy to point out the fact, it is more probable that X(J)-U(H), if considered a Wb, would be more stable, hence a visible cross peak between iminos, on the cases of a7) and b1&2). This is due to the 3' dangling nucleotides which have been shown to be able to stabilize the nucleotide base paired at the end of the stem (Clair V. Crowther et al, 2016, RNA Society). On other cases, X(J)-U(H), is present in the middle of the duplex where a high density of Mms are present and of a single nucleotide bulge a3&6). This cases have been proven to be more destabilizing to the duplex then Mms or bulges at the end of the stem (Clair V. Crowther et al, 2016, RNA Society; Blose, J. M. et al., 2007; Kierzek, R. et al., 1999; Davis, A. R., & Znosko, B. M., 2008).

Regardless of speculation, more NMR data may provide a more definitive idea about the structure and dynamics of this duplexes.

3-4) Simulations of several miR-34a.mRNA complexes:

A total of 36 miR-34a.mRNA duplex MFE structures were simulated using MCfold (Parisien, M., & Major, F., 2008), on which we vary the target mRNA. All structures were validated through Lucifrase essays (Hermeking, H. 2007). All this structures are presented in the supplementary material, fig.22. This experiments were ran in order to assess if preliminarily, there are any apparent general structural features for miRNA-34a targeting.

A common trend can be observed, displaying a certain preference in general for kink on the miRNA side. At the light of miR.mRNA duplex binding to Ago, it has been shown by x-ray crystallography, a curved like binding of miRNA to the protein, hence stimulating the idea that it is possibly necessary that the duplex provides such bending, for binding to occur. Other structures that did not display this bending commonly were

attributed to: 1) mRNAs with exactly 22 nucleotides length chosen; 2) duplexes with to bulges, one on each strand; 3) duplexes with a single bulge on the miRNA strand. Case 1) can still present a kink in reality even though not shown by the mcfold output representation. However, cases 2) and 3) could only present such feature if maybe there would be a low populated conformation with a bend on the miRNA side.

4) Conclusions and future prospects:

Purification of gCD44 was achieved with only an IE HPLC column to approximately one nucleotide resolution. Through conjoint analysis of EMSA experiments, we can conclude that both miR-34a.PNUTS and miR-34a.gCD44 are transient complexes, in which probably miR-34a.gCD44 has higher Kd. Ratio necessary for miRNA saturation with mRNA in NMR seems to be quite different from the expected through EMSA, in this case being lower. This might be due to different chemical environments required for each experiment. UV melting hints at miR-34a.PNUTS complex being more stable, but due to the presence of free mRNA, this results can be misleading. NMR date provides insight into a highly dynamics miR-34a.PNUTS duplex, which could explain the lower Kd values comparing EMSAs of both duplexes.

Preliminary results on bulk predictions of several miR-34a.mRNA structures with mcfold hint at a necessity of a kink on the miRNA side, facilitating binding with Ago. For a more statistical significant conclusion, a bigger population of duplexes with different mRNA targets must be generated (in progress). Other features must also be probed in order to understand which are the most relevant and what dependences they could have with others. Such features can be: bulge size, bulge content, distance of the bulge to seed sequence, distance between bulges, nucleotide content in the duplex, position of Wbs and Mms. Such tasks are on-going collaborating with the Andresson lab at Karolinska, which could bring better insight into how these structural features play a role in miRNA targeting.

In the future, further assignment on miR-34a.gCD44 duplex must be done in order to lift doubts regarding structure organization. Assignment of H1'-H8/H6/NH, NH-H2 and NH_2-NH/H6/H5 NOE connectivity's must be carried out at 298K and 284K, if necessary maybe, lower than 284K. Increasing concentration of mRNA and miRNA would be crucial for better signal to noise racio in the future. Taking into account the low stability of both duplexes, it should be considered $MgCl_2$ titration in order to assess stabilization and obtain not only a structure with conditions close to physiological solvent but also a more stable complex.

Because of the stoichiometry necessary to saturate miRNA with Mrna in both cases, EMSA does not seem to be a suitable technic to measure Kd and UV melting to evaluate stability. MST (Microscale thermophoresis), is a new emergent technic which can measure Kd independently of which solvent (water, NMR buffer, cell lysate) and

stoichiometry. It can measure Kd from pico to milimolar ranges, being the latter the case in study. It can also infer stoichiometry, binding modes, free energy, enthalpy and entropy. Finally, low sample consumption and little measurement time is required. Such measurements have been done on DNA, RNA, proteins and other small molecules (Jerabek-Willemsen, M. et al., 2014; Gaffarogullari, E. C. et al., 2013).

5) Bibliography

Auffinger, P., & Westhof, E. (1998). Hydration of RNA base pairs. *Journal of Biomolecular Structure and Dynamics*, *16*(3), 693-707.

Azuma-Mukai, A., Oguri, H., Mituyama, T., Qian, Z. R., Asai, K., Siomi, H., & Siomi, M. C. (2008). Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proceedings of the National Academy of Sciences*, *105*(23), 7964-7969.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, *136*(2), 215-233.

Blose, J. M., Manni, M. L., Klapec, K. A., Stranger-Jones, Y., Zyra, A. C., Sim, V., Griffith, C. A., Long, J. D. & Serra, M. J. (2007). Non-Nearest-Neighbor Dependence of Stability for RNA Bulge Loops Based on the Complete Set of Group I Single Nucleotide Bulge Loops. *Biochemistry*, *46*(51), 15123.

Boon, R. A., Iekushi, K., Lechner, S., Seeger, T., Fischer, A., Heydt, S., Kaluza, D., Tréguer, K., Carmona, G., Bonauer, A., Horrevoets, A. J., Didier, N., Girmatsion, Z., Biliczki, P., Ehrlich, J. R., Katus, H. A., Müller, O. J., Potente, M., Zeiher, A. M., Hermeking, H. & Dimmeler, S. (2013). MicroRNA-34a regulates cardiac ageing and function. *Nature*, *495*(7439), 107-110.

Castelo-Branco, G., Wagner, J., Rodriguez, F. J., Kele, J., Sousa, K., Rawal, N., Pasolli, H. A., Fuchs, E., Kitajewski, J., & Arenas, E. (2003). Differential regulation of midbrain dopaminergic neuron development by Wnt-1, Wnt-3a, and Wnt-5a. *Proceedings of the National Academy of Sciences*, *100*(22), 12747-12752.

Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., & Noller, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science*, *285*(5436), 2095-2104.

Chen, Z., & Zhang, Y. (2005). Dimethyl sulfoxide targets phage RNA polymerases to promote transcription. *Biochemical and biophysical research communications*, *333*(3), 664-670.

Cheong, C., Varani, G., & Tinoco, I. (1990). Solution structure of an unusually stable RNA hairpin, 5GGAC (UUCG) GUCC. *Nature*, *346*(6285), 680-682.

Chheda, N., & Gupta, M. K. (2014). *RNA as a Permutation*. Cornell University library.

Christiansen, M. E., & Znosko, B. M. (2009). Thermodynamic characterization of tandem mismatches found in naturally occurring RNA. *Nucleic acids research*, *37*(14), 4696-4706.

Christiansen, M. E., & Znosko, B. M. (2008). Thermodynamic characterization of the complete set of sequence symmetric tandem mismatches in RNA and an improved model for predicting the free energy contribution of sequence asymmetric tandem mismatches. *Biochemistry*, *47*(14), 4329-4336.

Corney, D. C., Hwang, C. I., Matoso, A., Vogt, M., Flesken-Nikitin, A., Godwin, A. K., Kamat, A. A., Sood, A. K., Ellenson, L. H., Hermeking, H., Nikitin, A. Y. (2010). Frequent downregulation of miR-34 family in human ovarian cancers. *Clinical Cancer Research*, *16*(4), 1119-1128.

Crichton, R. R., & Louro, R. O. (2012). *Practical approaches to biological inorganic chemistry*, Newnes.

Crowther, C. V., Jones, L. E., Morelli, J. N., Mastrogiacomo, E. M., Porterfield, C., Kent, J. L., & Serra, M. J. (2017). Influence of two bulge loops on the stability of RNA duplexes. *RNA*, *23*(2), 217-228.

Davis, A. R., & Znosko, B. M. (2008). Thermodynamic characterization of naturally occurring RNA single mismatches with GU nearest neighbors. *Biochemistry*, *47*(38), 10178-10187.

De Leon, G., Sherry, T. C., & Krucher, N. A. (2008). Reduced expression of PNUTS leads to activation of Rb-phosphatase and caspase-mediated apoptosis. *Cancer biology & therapy*, *7*(6), 833-841.

Draper, D. E. (2004). A guide to ions and RNA structure. *RNA*, *10*(3), 335-343.

Dumas, P. (2014). *Handbook of RNA Biochemistry*, 2nd edition, Wiley-VCH, pp. 445-446.

Faehnle, C. R., & Joshua-Tor, L. (2010). Argonaute MID domain takes centre stage. *EMBO reports*, *11*(8), 564-565.

Fürtig, B., Richter, C., Wöhnert, J., & Schwalbe, H. (2003). NMR spectroscopy of RNA. *Chembiochem*, *4*(10), 936-962.

Gaffarogullari, E. C., Krause, A., Balbo, J., Herten, D. P., & Jäschke, A. (2013). Microscale thermophoresis provides insights into mechanism and thermodynamics of ribozyme catalysis. *RNA biology*, *10*(12), 1815-1821.

Ghosh, A., & Bansal, M. (2003). A glossary of DNA structures from A to Z. *Acta Crystallographica Section D: Biological Crystallography*, *59*(4), 620-626.

Gjerde, D. T., Hoang, L., & Hornby, D. (2009). *RNA purification and analysis: sample preparation, extraction, chromatography*. John Wiley & Sons.

Grimson, A., Farh, K. K. H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, *27*(1), 91-105.

Ha, M., & Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nature reviews Molecular cell biology*, *15*(8), 509-524.

Hashimi, S. T., Fulcher, J. A., Chang, M. H., Gov, L., Wang, S., & Lee, B. (2009). MicroRNA profiling identifies miR-34a and miR-21 and their target genes JAG1 and WNT1 in the coordinate regulation of dendritic cell differentiation. *Blood*, *114*(2), 404-414.

Hermann, T., & Patel, D. J. (2000). RNA bulges as architectural and recognition motifs. *Structure*, *8*(3), R47-R54.

Hermeking, H. (2007). p53 enters the microRNA world. *Cancer cell*, *12*(5), 414-418.

Isaksson, J., & Chattopadhyaya, J. (2005). A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochemistry*, *44*(14), 5390-5401.

Iwasaki, A., & Medzhitov, R. (2010). Regulation of adaptive immunity by the innate immune system. *Science*, *327*(5963), 291-295.

Iwata, M., Izawa, M., Sasaki, N., Nagumo, Y., Sasabe, H., & Hayashizaki, Y. (2000). T7 RNA polymerase activation and improvement of the transcriptional sequencing by polyamines. *Bioorganic & medicinal chemistry*, *8*(8), 2185-2194.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., & Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, *1077*, 101-113.

Kao, C., Zheng, M., & Rüdisser, S. (1999). A simple and efficient method to reduce nontemplated nucleotide addition at the 3′ terminus of RNAs transcribed by T7 RNA polymerase. *RNA*, *5*(9), 1268-1272.

Katoh, M., & Katoh, M. (2003). CLDN23 gene, frequently down-regulated in intestinal-type gastric cancer, is a novel member of CLAUDIN gene family. *International journal of molecular medicine*, *11*(6), 683-689.

Keupp, K., Beleggia, F., Kayserili, H., Barnes, A. M., Steiner, M., Semler, O., Fischer, B., Yigit, G., Janda, C. Y., Becker, J. & Breer, S. (2013). Mutations in WNT1 cause different forms of bone fragility. *The American Journal of Human Genetics*, *92*(4), 565-574.

Khvorova, A., Reynolds, A., & Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell, 115(2), 209-216.

Kierzek, R., Burkard, M. E., & Turner, D. H. (1999). Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, *38*(43), 14214-14223.

Kim, H., Lee, O. H., Xin, H., Chen, L. Y., Qin, J., Chae, H. K., Lin, S., Safari, A., Liu, D. & Songyang, Z. (2009). TRF2 functions as a protein hub and regulates telomere maintenance by recognizing specific peptide motifs. *Nature structural & molecular biology*, *16*(4), 372-379.

Klostermeier, D., & Hammann, C. (2013). *RNA Structure and Folding: Biophysical Techniques and Prediction Methods*, 22th edition, DE GRUYTER.

Koculi, E., Hyeon, C., Thirumalai, D., & Woodson, S. A. (2007). Charge density of divalent metal cations determines RNA stability. *Journal of the American Chemical Society*, *129*(9), 2676.

Komiya, Y., & Habas, R. (2008). Wnt signal transduction pathways. *Organogenesis*, *4*(2), 68-75.

Kuzmine, I., Gottlieb, P. A., & Martin, C. T. (2003). Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *Journal of Biological Chemistry*, *278*(5), 2819-2823.

Landsverk, H. B., Mora-Bermúdez, F., Landsverk, O. J., Hasvold, G., Naderi, S., Bakke, O., Ellenberg, J., Collas, P., Syljuåsen, R. G., & Küntziger, T. (2010). The protein phosphatase 1 regulator PNUTS is a new component of the DNA damage response. *EMBO reports*, *11*(11), 868-875.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., & Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, *23*(20), 4051-4060.

Liu, C., Kelnar, K., Liu, B., Chen, X., Calhoun-Davis, T., Li, H., ... & Wiggins, J. F. (2011). The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nature medicine*, *17*(2), 211-215.

Makarova, K. S., Wolf, Y. I., Van der Oost, J., & Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology direct*, *4*(1), 29.

Maroof, H., Salajegheh, A., Anthony Smith, R., & King-Yin Lam, A. (2014). MicroRNA-34 family, mechanisms of action in cancer: a review. *Current cancer drug targets*, *14*(8), 737-751.

Martin, C. T., & Coleman, J. E. (1989). T7 RNA polymerase does not interact with the 5'-phosphate of the initiating nucleotide. *Biochemistry*, *28*(7), 2760-2762.

Masquida, B., & Westhof, E. (2000). On the wobble GoU and related pairs. *RNA*, *6*(1), 9-15.

McMahon, A. P., & Bradley, A. (1990). The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell*, *62*(6), 1073-1085.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell*, *15*(2), 185-197.

Miller, S., Jones, L. E., Giovannitti, K., Piper, D., & Serra, M. J. (2008). Thermodynamic analysis of 5′ and 3′ single-and 3′ double-nucleotide overhangs neighboring wobble terminal base pairs. *Nucleic acids research*, *36*(17), 5652-5659.

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D. L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K. & Saito, T. (2011). Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell stem cell*, *8*(6), 633-638.

Murray, J. B., Collier, A. K., & Arnold, J. R. (1994). A general purification procedure for chemically synthesized oligoribonucleotides. *Analytical biochemistry*, *218*(1), 177-184.

Mustoe, A. M., Brooks, C. L., & Al-Hashimi, H. M. (2014). Hierarchy of RNA functional dynamics. *Annual review of biochemistry*, *83*, 441-466.

Nakanishi, K. (2016). Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins?. *Wiley Interdisciplinary Reviews: RNA*, *7*(5), 637-660.

Nelson, D. L., Lehninger, A. L., & Cox, M. M. (2008). *Lehninger principles of biochemistry*, 5th edition, Macmillan.

Nusse, R., Theunissen, H., Wagenaar, E., Rijsewijk, F., Gennissen, A., Otte, A., Schuuring, E & Van Ooyen, A. (1990). The Wnt-1 (int-1) oncogene promoter and its mechanism of activation by insertion of proviral DNA of the mouse mammary tumor virus. *Molecular and Cellular Biology*, *10*(8), 4170-4179.

Nykänen, A., Haley, B., & Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, *107*(3), 309-321.

Oliver R. W. A. (1997). HPLC of macromolecules

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K., & Aravin, A. A. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. *Molecular cell*, *51*(5), 594-605.

Pan, B., & Sundaralingam, M. (1999). Mismatched base pairing in RNA crystal structures. *International journal of quantum chemistry*, *75*(3), 275-287.

Parisien, M., & Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, *452*(7183), 51-55.

Peters, L., & Meister, G. (2007). Argonaute Proteins: Mediators of RNA Silencing. Molecular Cell, 26(5), 611–623.

Peyret, N., Seneviratne, P. A., Allawi, H. T., & SantaLucia, J. (1999). Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A.A, C.C, G.G, and T.T Mismatches. *Biochemistry*, *38*(12), 3468-3477.

Roy, A., Panigrahi, S., Bhattacharyya, M., & Bhattacharyya, D. (2008). Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies. *The Journal of Physical Chemistry B*, *112*(12), 3786-3796.

SantaLucia Jr, J., Kierzek, R., & Turner, D. H. (1991). Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U. cntdot. U and C. cntdot. C+ pairs. *Biochemistry*, *30*(33), 8242-8251.

Sontheimer, E. J. (2005). Assembly and function of RNA silencing complexes. *Nature Reviews Molecular Cell Biology*, *6*(2), 127-138.

Šponer, J., Šponer, J. E., Mládek, A., Jurečka, P., Banáš, P., & Otyepka, M. (2013). Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*, *99*(12), 978-988.

Spring, F. A., Dalchau, R., Daniels, G. L., Mallinson, G., Judson, P. A., Parsons, S. F., Fabre, J. W. & Anstee, D. J. (1988). The Ina and Inb blood group antigens are located on a glycoprotein of 80,000 MW (the CDw44 glycoprotein) whose expression is influenced by the In (Lu) gene. *Immunology*, *64*(1), 37.

Stryer, L., Berg, J. M., & Tymoczko, J. L. (2000). *Biochemistry*. 4th edition, W.H. Freeman & Co Ltd.

Varani, G., & McClain, W. H. (2000). The G· U wobble base pair. *EMBO reports*, *1*(1), 18-23.

Vikesaa, J., Hansen, T. V., Jønson, L., Borup, R., Wewer, U. M., Christiansen, J., & Nielsen, F. C. (2006). RNA-binding IMPs promote cell adhesion and invadopodia formation. *The EMBO journal*, *25*(7), 1456-1468.

Wang, X. P., Wang, Q. X., Lin, H. P., Xu, B., Zhao, Q., & Chen, K. (2016). Recombinant heat shock protein 70 functional peptide and alpha-fetoprotein epitope peptide vaccine elicits specific anti-tumor immunity. *Oncotarget*, *7*(44), 71274-71284.

Westhof, E., & Auffinger, P. (2000). RNA tertiary structure. *Encyclopedia of analytical chemistry*.

Westhof, E., & Beveridge, D. L. (1990). Hydration of nucleic acids. *Water Science Reviews*, *5*, 24-123.

Wijmenga, S. S., & van Buuren, B. N. (1998). The use of NMR methods for conformational studies of nucleic acids. *Progress in nuclear magnetic resonance spectroscopy*, *32*(4), 287-387.

Carl E. Longfellow, Ryszard Kierzek,l & Douglas H. Turner (1990). Thermodynamic and Spectroscopic Study of Bulge Loops in Oligoribonucleotides Biochemistry, 29, 278-285.

Winter, J., Jung, S., Keller, S., Gregory, R. I., & Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nature Cell Biology, 11(3), 228–234.

Yi, R., Qin, Y., Macara, I. G., & Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development*, *17*(24), 3011-3016.

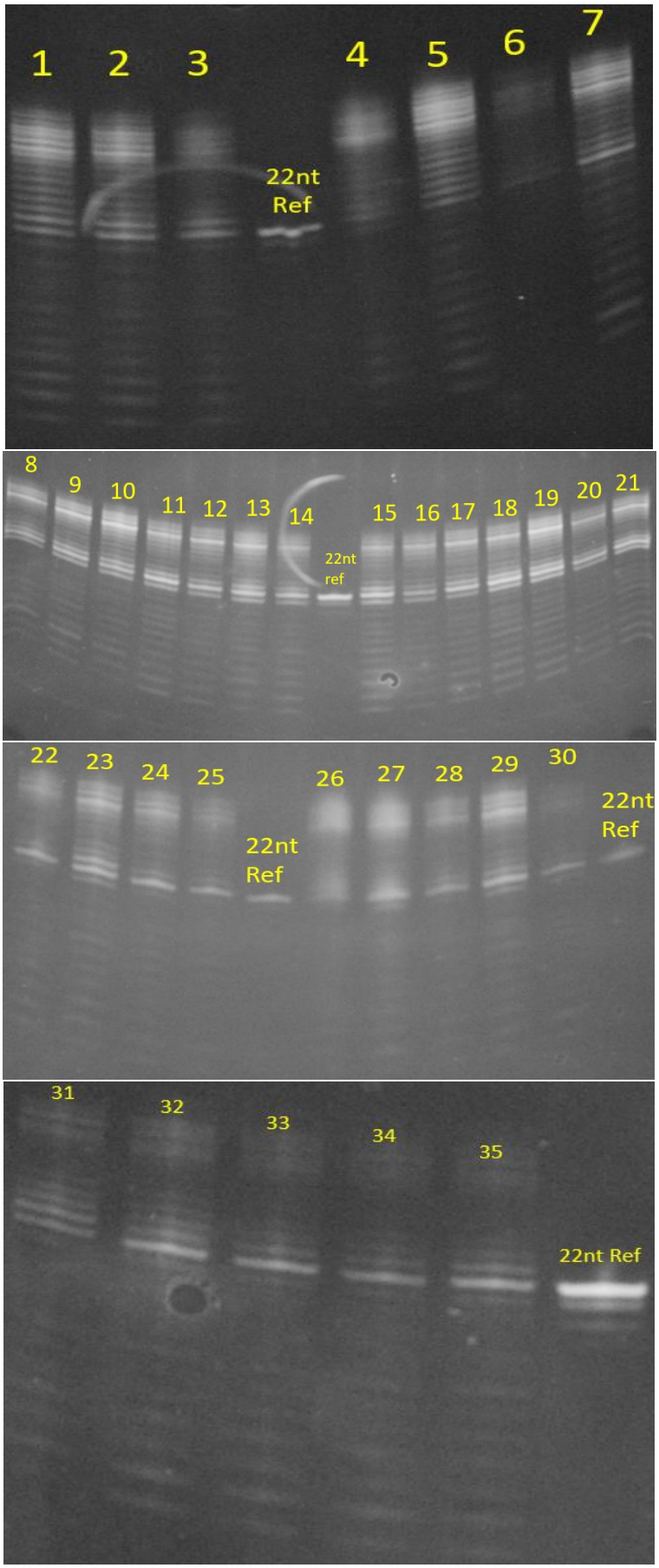
Yin, Y., & Carter Jr, C. W. (1996). Incomplete factorial and response surface methods in experimental design: Yield optimization of tRNATrp from in vitro T7 RNA polymerase transcription. *Nucleic acids research*, *24*(7), 1279-1286.

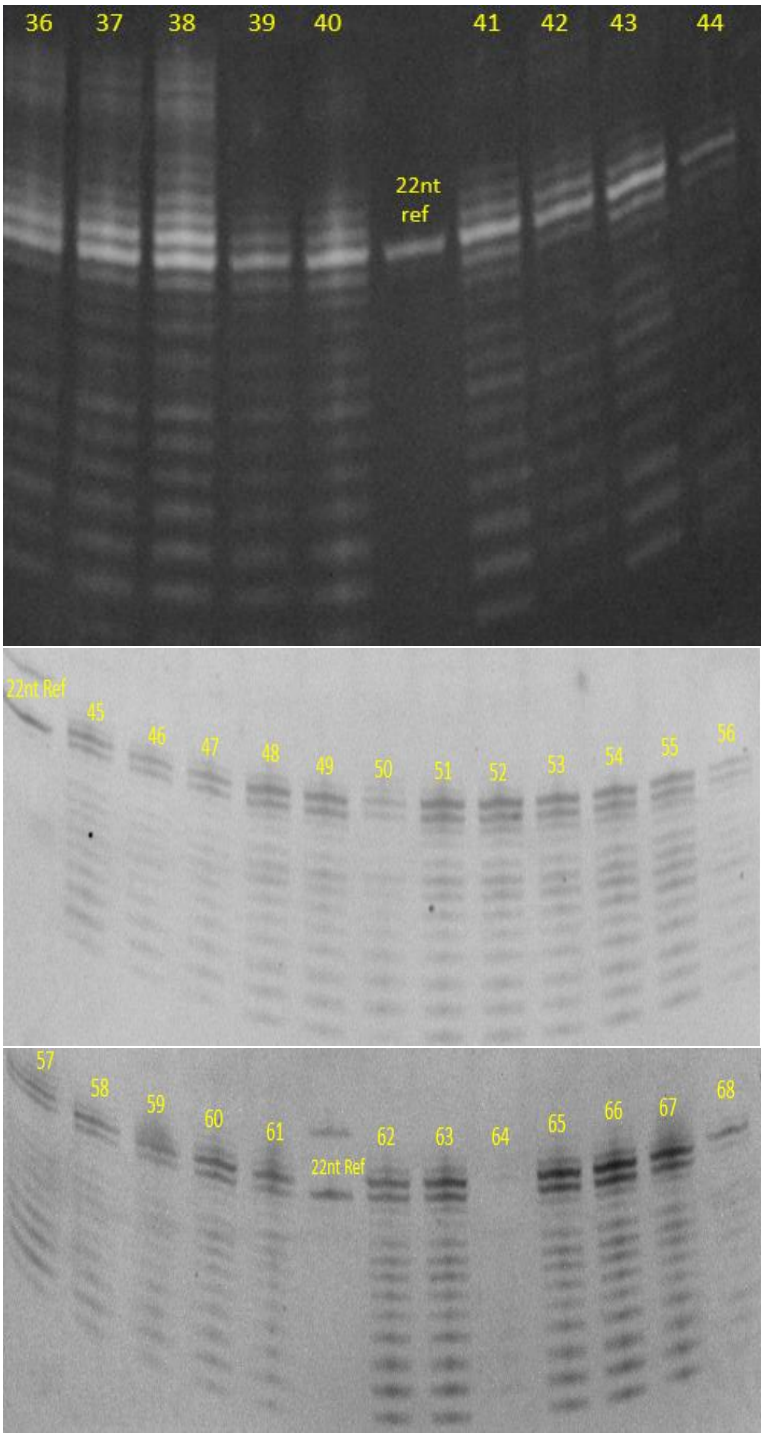
Zhan, Y., & Wu, L. (2012). Functional regulation of monocyte-derived dendritic cells by microRNAs. *Protein & cell*, *3*(7), 497-507.

Znosko, B. M., Kennedy, S. D., Wille, P. C., Krugh, T. R., & Turner, D. H. (2004). Structural features and thermodynamics of the J4/5 loop from the Candida albicans and Candida dubliniensis group I introns. *Biochemistry*, *43*(50), 15822-15837.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*, *31*(13), 3406-3415.

6) Supplementary material:




36
37
38
39
40
41
42
43
44
22nt
ref
22nt Ref
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
22nt Ref
62
63
64
65
66
67
68

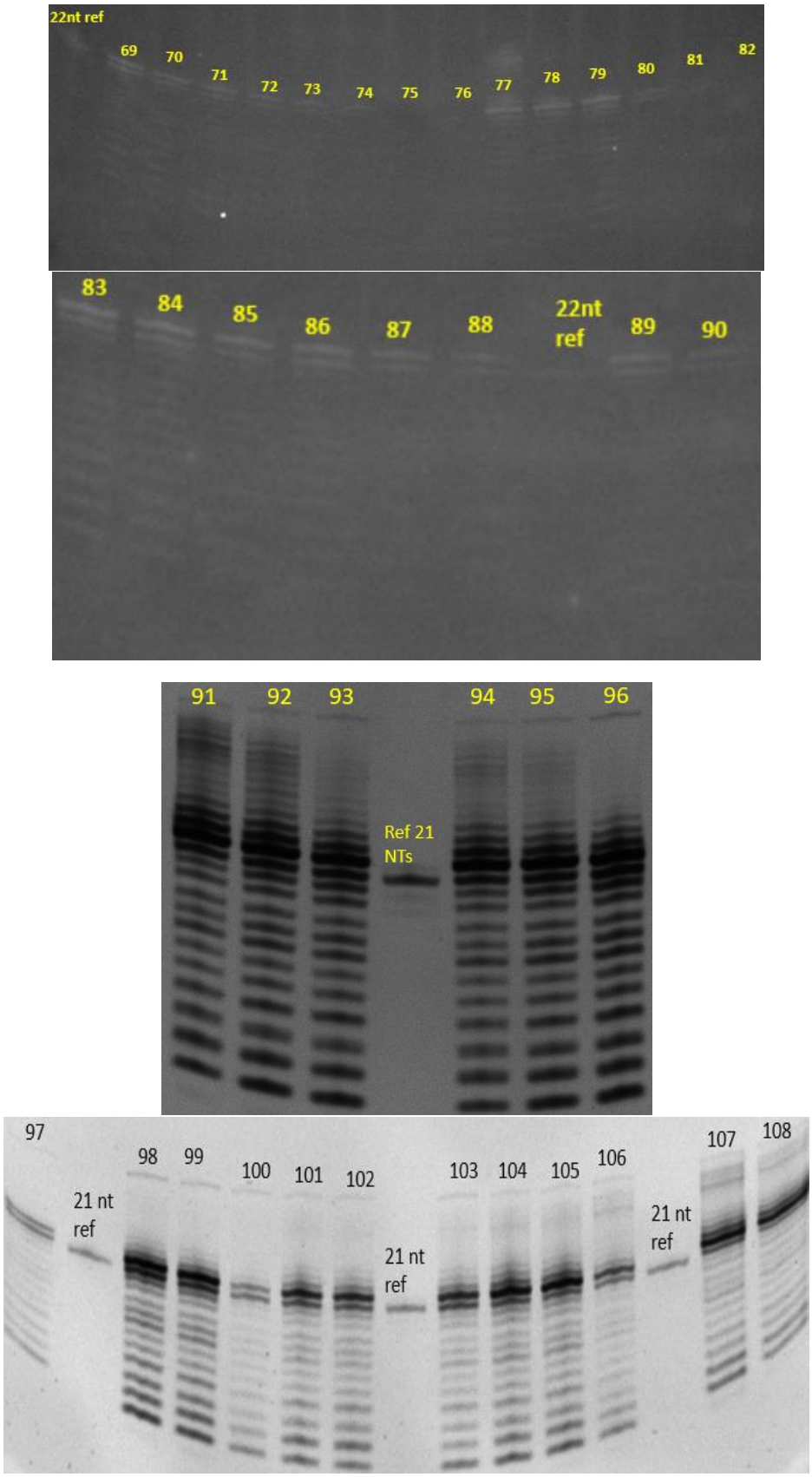
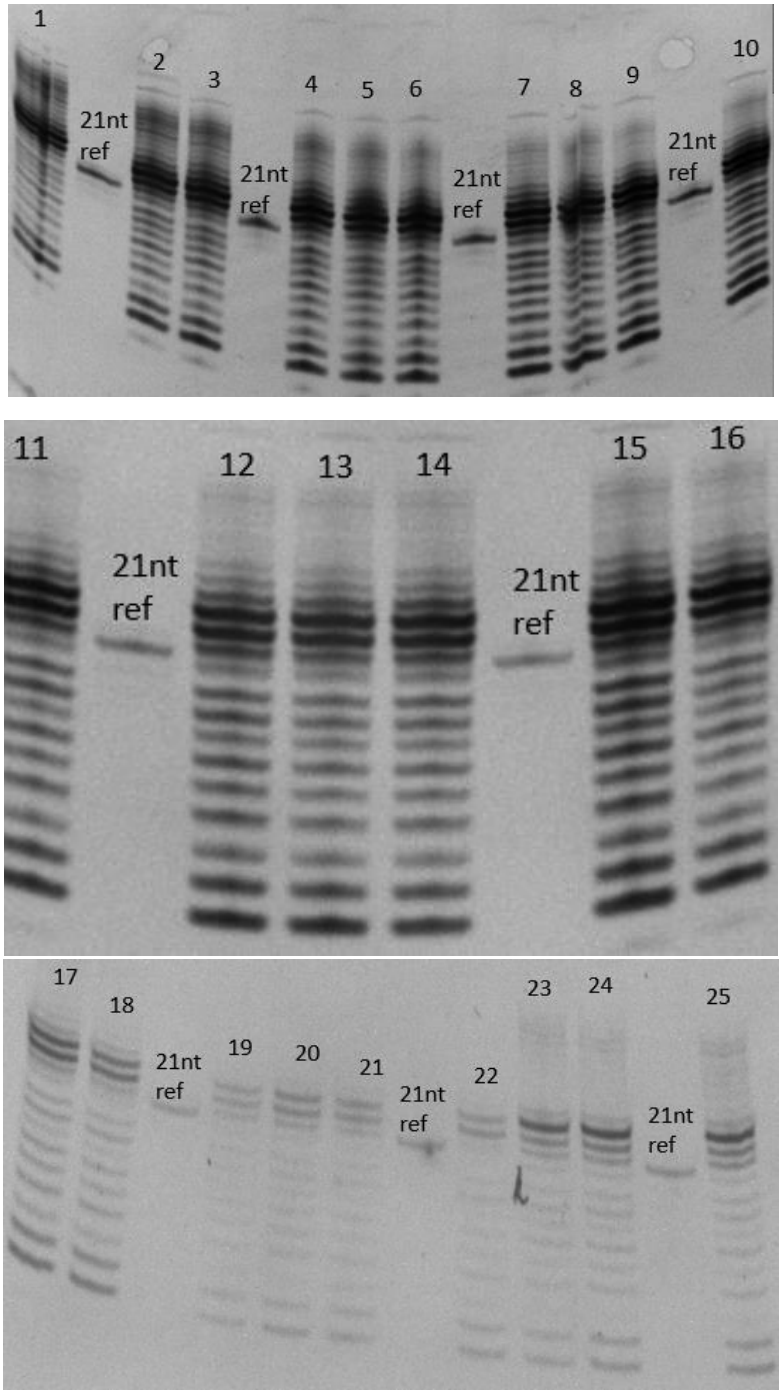
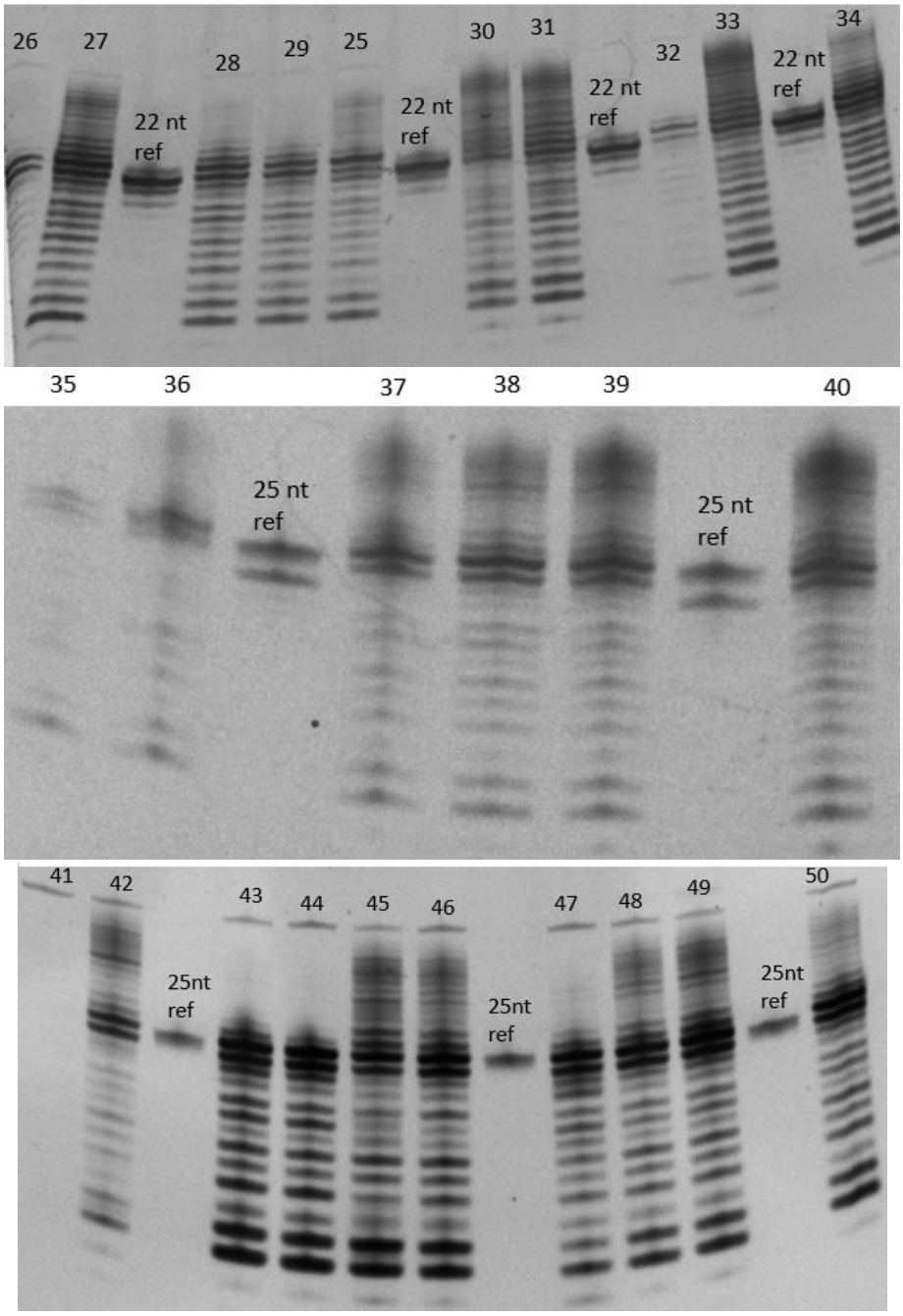


figure 18. All 20% polyacrylamide denaturing gels for transcription reaction optimization for CD44. Each well is labelled according to the reaction depicted in Table 2.


1
2
3
4
5
6
7
8
9
10
21nt ref
21nt ref
21nt ref
21nt ref
11
12
13
14
15
16
21nt ref
21nt ref
17
18
19
20
21
22
23
24
25
21nt ref
21nt ref
21nt ref


26
27
22 nt ref
28
29
25
22 nt ref
30
31
22 nt ref
32
33
22 nt ref
34
35
36
25 nt ref
37
38
39
25 nt ref
40
41
42
25nt ref
43
44
45
46
25nt ref
47
48
49
25nt ref
50

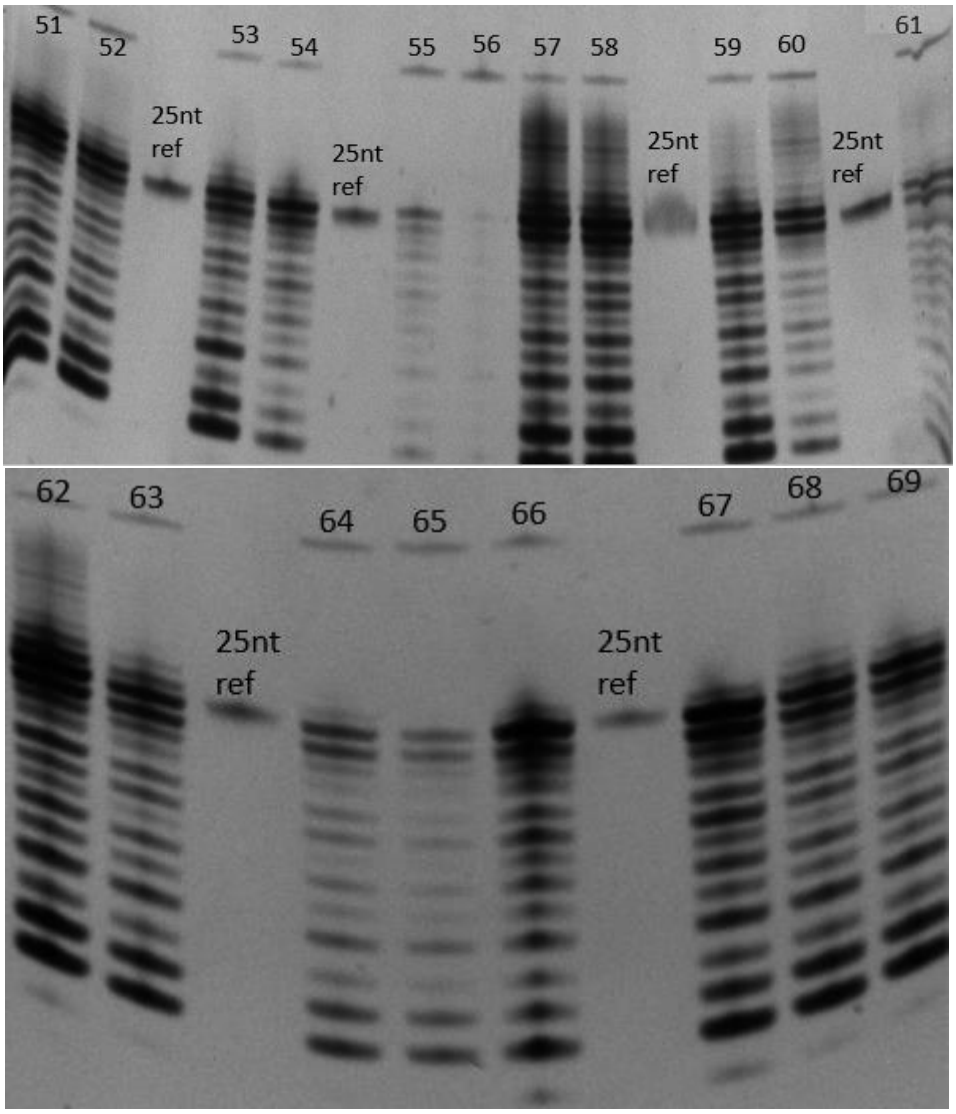
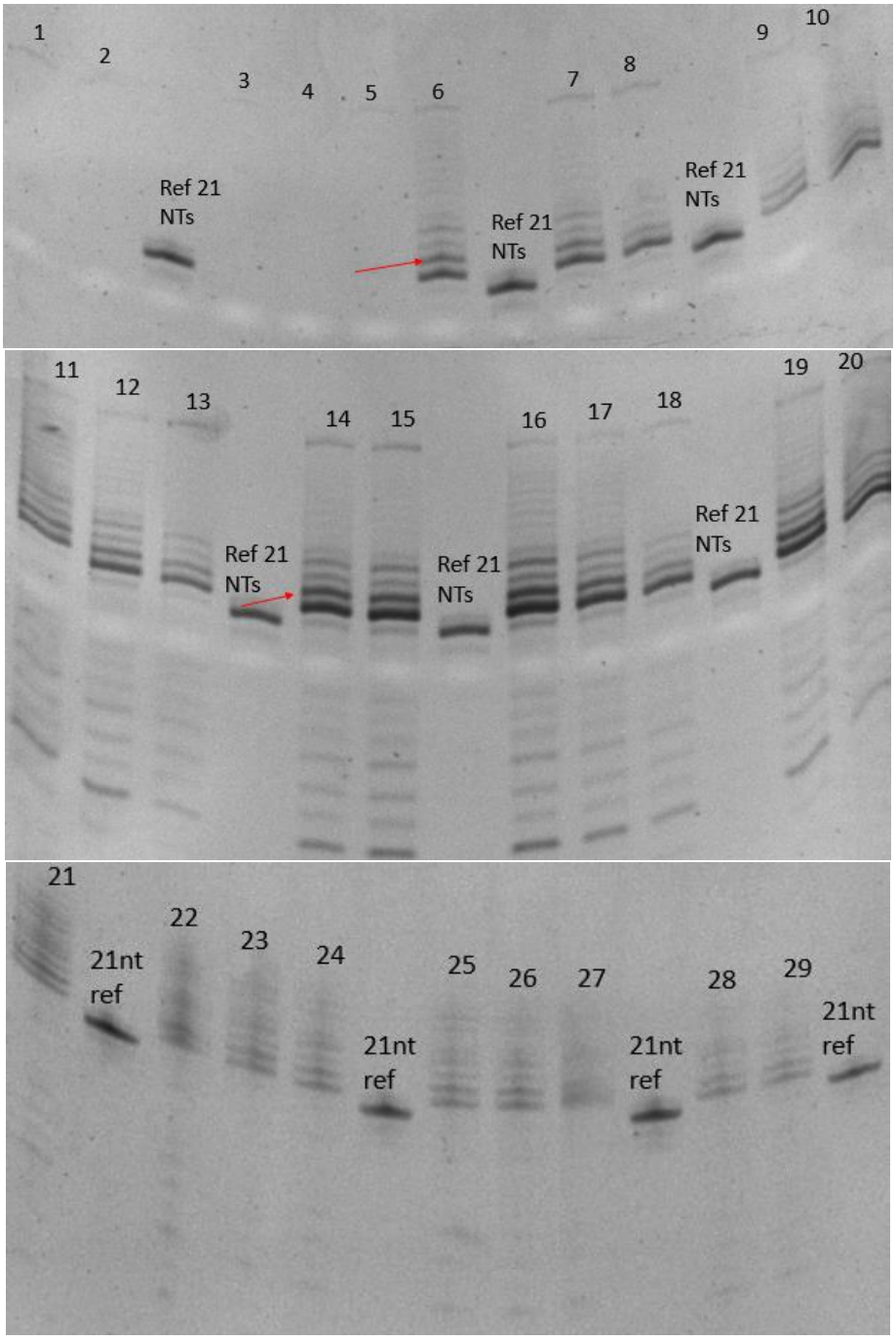
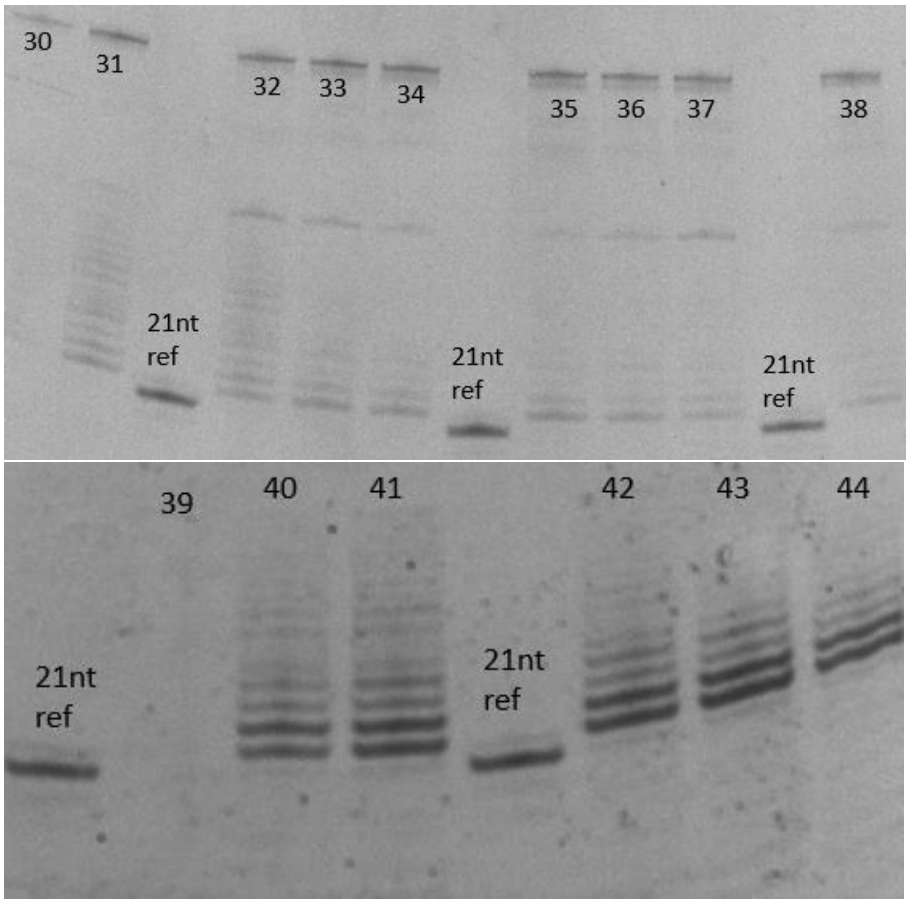


figure 19. All 20% polyacrylamide denaturing gels for transcription reaction optimization for gCD44. Each well is labelled according to the reaction depicted in Table 2.


1
2
3
4
5
6
7
8
9
10
Ref 21
NTs
Ref 21
NTs
Ref 21
NTs
11
12
13
14
15
16
17
18
19
20
Ref 21
NTs
Ref 21
NTs
Ref 21
NTs
21
22
23
24
25
26
27
28
29
21nt
ref
21nt
ref
21nt
ref
21nt
ref


30
31
32
33
34
35
36
37
38
21nt ref
21nt ref
21nt ref
39
40
41
42
43
44
21nt ref
21nt ref

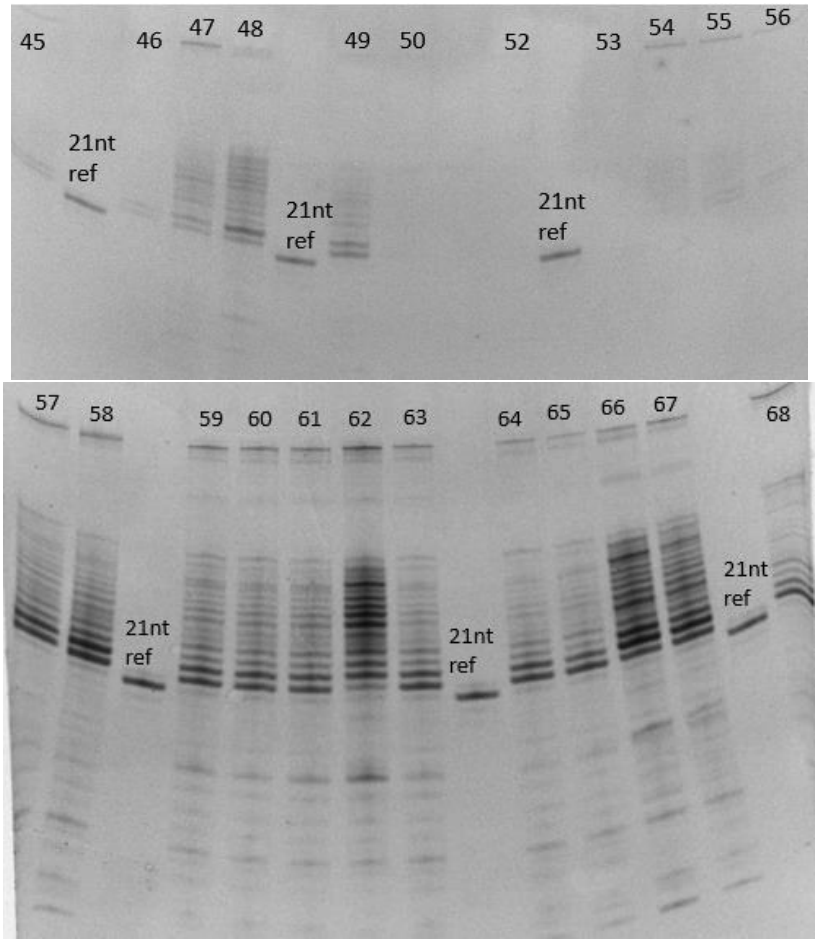


figure 20. All 20% polyacrylamide denaturing gels for transcription reaction optimization for WNT1. Each well is labelled according to the reaction depicted in Table 2.

mRNA	Number	Tris (M)	MgCl2 (M)	DTT (M)	SP (M)	DMSO (v/v)%	GMP (mM)	ATP (mM)	GTP (mM)	CTP (mM)	UTP (mM)	DNA (uM)
CD44	1	0,1	0,03	0,03	0,002	10	6	3	3	3	3	2
	2	0,1	0,03	0,03	0,002	10	12	3	3	3	3	2
	3	0,1	0,03	0,03	0,002	10	18	3	3	3	3	2
	4	0,1	0,03	0,03	0,002	10	3	1	3	3	3	2
	5	0,1	0,03	0,03	0,002	10	3	3	1	3	3	2
	6	0,1	0,03	0,03	0,002	10	3	3	3	1	3	2
	7	0,1	0,03	0,03	0,002	10	3	3	3	3	3	1
	8	0,1	0,03	0,03	0,002	3	12	3	3	3	3	1
	9	0,1	0,03	0,03	0,002	6	12	3	3	3	3	1
	10	0,1	0,03	0,03	0,002	10	12	3	3	3	3	1
	11	0,1	0,03	0,03	0,002	14	12	3	3	3	3	1
	12	0,05	0,03	0,03	0,002	10	12	3	3	3	3	1
	13	0,1	0,03	0,03	0,002	10	12	3	3	3	3	1
	14	0,15	0,03	0,03	0,002	10	12	3	3	3	3	1
	15	0,2	0,03	0,03	0,002	10	12	3	3	3	3	1
	16	0,1	0,02	0,03	0,002	10	12	3	3	3	3	1
	17	0,1	0,04	0,03	0,002	10	12	3	3	3	3	1
	18	0,1	0,06	0,03	0,002	10	12	3	3	3	3	1
	19	0,1	0,03	0,03	0,002	10	14	3	3	3	3	1
	20	0,1	0,03	0,03	0,002	10	16	3	3	3	3	1
	21	0,1	0,03	0,03	0,002	10	24	3	3	3	3	1
	22	0,1	0,06	0,03	0,002	14	14	3	3	3	3	1
	23	0,1	0,06	0,03	0,002	16	14	3	3	3	3	1
	24	0,1	0,06	0,03	0,002	18	14	3	3	3	3	1
	25	0,1	0,06	0,03	0,002	20	14	3	3	3	3	1
	26	0,1	0,05	0,03	0,002	10	14	3	3	3	3	1
	27	0,1	0,06	0,03	0,002	10	14	3	3	3	3	1
	28	0,1	0,07	0,03	0,002	10	14	3	3	3	3	1
	29	0,1	0,08	0,03	0,002	10	14	3	3	3	3	1
	30	0,1	0,09	0,03	0,002	10	14	3	3	3	3	1
	31	0,1	0,06	0,03	0,002	16	14	3	3	3	3	1
	32	0,1	0,06	0,03	0,002	16,6	14	3	3	3	3	1
	33	0,1	0,06	0,03	0,002	17,2	14	3	3	3	3	1
	34	0,1	0,06	0,03	0,002	17,6	14	3	3	3	3	1
	35	0,1	0,06	0,03	0,002	18	14	3	3	3	3	1
	36	0,15	0,06	0,03	0,002	16,6	14	3	3	3	3	1
	37	0,17	0,06	0,03	0,002	16,6	14	3	3	3	3	1

38	0,2	0,06	0,03	0,002	16,6	14	3	3	3	3	1
39	0,22	0,06	0,03	0,002	16,6	14	3	3	3	3	1
40	0,23	0,06	0,03	0,002	16,6	14	3	3	3	3	1
41	0,25	0,06	0,03	0,002	16,6	14	3	3	3	3	1
42	0,1	0,07	0,03	0,002	16,6	14	3	3	3	3	1
43	0,1	0,08	0,03	0,002	16,6	14	3	3	3	3	1
44	0,1	0,09	0,03	0,002	16,6	14	3	3	3	3	1
45	0,2	0,08	0,03	0,002	10	14	3	3	3	3	1
46	0,2	0,08	0,03	0,002	12	14	3	3	3	3	1
47	0,2	0,08	0,03	0,002	14	14	3	3	3	3	1
48	0,2	0,08	0,03	0,002	16	14	3	3	3	3	1
49	0,2	0,08	0,03	0,002	18	14	3	3	3	3	1
50	0,2	0,08	0,03	0,002	18	0	3	3	3	3	1
51	0,2	0,08	0,03	0,002	20	14	3	3	3	3	1
52	0,2	0,08	0,03	0,002	22	14	3	3	3	3	1
53	0,2	0,08	0,03	0,002	24	14	3	3	3	3	1
54	0,2	0,08	0,03	0,002	26	14	3	3	3	3	1
55	0,2	0,08	0,03	0,002	28	14	3	3	3	3	1
56	0,2	0,08	0,03	0,002	30	14	3	3	3	3	1
57	0,2	0,08	0,03	0,002	16,6	0	3	3	3	3	1
58	0,2	0,08	0,03	0,002	16,6	2	3	3	3	3	1
59	0,2	0,08	0,03	0,002	16,6	4	3	3	3	3	1
60	0,2	0,08	0,03	0,002	16,6	6	3	3	3	3	1
61	0,2	0,08	0,03	0,002	16,6	8	3	3	3	3	1
62	0,2	0,08	0,03	0,002	16,6	10	3	3	3	3	1
63	0,2	0,08	0,03	0,002	16,6	12	3	3	3	3	1
64	0,2	0,08	0,03	0,002	0	12	3	3	3	3	1
65	0,2	0,08	0,03	0,002	16,6	14	3	3	3	3	1
66	0,2	0,08	0,03	0,002	16,6	16	3	3	3	3	1
67	0,2	0,08	0,03	0,002	16,6	18	3	3	3	3	1
68	0,2	0,08	0,03	0,002	16,6	20	3	3	3	3	1
69	0,2	0,08	0,03	0,002	20	16	3	3	3	3	1
70	0,2	0,08	0,03	0,002	20	18	3	3	3	3	1
71	0,2	0,08	0,03	0,002	20	20	3	3	3	3	1
72	0,2	0,08	0,03	0,002	20	22	3	3	3	3	1
73	0,2	0,08	0,03	0,002	20	24	3	3	3	3	1
74	0,2	0,08	0,03	0,002	20	26	3	3	3	3	1
75	0,2	0,08	0,03	0,002	20	28	3	3	3	3	1
76	0,2	0,08	0,03	0,002	20	30	3	3	3	3	1
77	0,2	0,02	0,03	0,002	20	16	3	3	3	3	1

	78	0,2	0,04	0,03	0,002	20	16	3	3	3	3	1
	79	0,2	0,06	0,03	0,002	20	16	3	3	3	3	1
	80	0,2	0,1	0,03	0,002	20	16	3	3	3	3	1
	81	0,2	0,12	0,03	0,002	20	16	3	3	3	3	1
	82	0,2	0,14	0,03	0,002	20	16	3	3	3	3	1
	83	0,14	0,08	0,03	0,002	20	16	3	3	3	3	1
	84	0,16	0,08	0,03	0,002	20	16	3	3	3	3	1
	85	0,18	0,08	0,03	0,002	20	16	3	3	3	3	1
	86	0,22	0,08	0,03	0,002	20	16	3	3	3	3	1
	87	0,24	0,08	0,03	0,002	20	16	3	3	3	3	1
	88	0,26	0,08	0,03	0,002	20	16	3	3	3	3	1
	89	0,2	0,08	0,03	0,002	20	16	3	3	3	3	1
	90	0,2	0,08	0,03	0,002	20	16	3	3	3	3	1
	91	0,2	0,02	0,03	0,002	16	16	3	3	3	3	1
	92	0,2	0,02	0,03	0,002	24	16	3	3	3	3	1
	93	0,2	0,02	0,03	0,002	28	16	3	3	3	3	1
	94	0,2	0,04	0,03	0,002	16	16	3	3	3	3	1
	95	0,2	0,04	0,03	0,002	24	16	3	3	3	3	1
	96	0,2	0,04	0,03	0,002	28	16	3	3	3	3	1
	97	0,04	0,08	0,03	0,002	20	16	3	3	3	3	1
	98	0,08	0,08	0,03	0,002	20	16	3	3	3	3	1
	99	0,12	0,08	0,03	0,002	20	16	3	3	3	3	1
	100	0,04	0,08	0,03	0,002	12	16	3	3	3	3	1
	101	0,08	0,08	0,03	0,002	12	16	3	3	3	3	1
	102	0,12	0,08	0,03	0,002	12	16	3	3	3	3	1
	103	0,04	0,04	0,03	0,002	20	16	3	3	3	3	1
	104	0,08	0,04	0,03	0,002	20	16	3	3	3	3	1
	105	0,12	0,04	0,03	0,002	20	16	3	3	3	3	1
	106	0,04	0,04	0,03	0,002	12	16	3	3	3	3	1
	107	0,08	0,04	0,03	0,002	12	16	3	3	3	3	1
	108	0,12	0,04	0,03	0,002	12	16	3	3	3	3	1
gCD44	1	0,1	0,02	0,03	0,002	16	16	3	3	3	3	2
	2	0,14	0,02	0,03	0,002	16	16	3	3	3	3	2
	3	0,1	0,04	0,03	0,002	16	16	3	3	3	3	2
	4	0,14	0,04	0,03	0,002	16	16	3	3	3	3	2
	5	0,1	0,052	0,03	0,002	16	16	3	3	3	3	2
	6	0,14	0,052	0,03	0,002	16	16	3	3	3	3	2
	7	0,18	0,052	0,03	0,002	16	16	3	3	3	3	2
	8	0,1	0,06	0,03	0,002	16	16	3	3	3	3	2

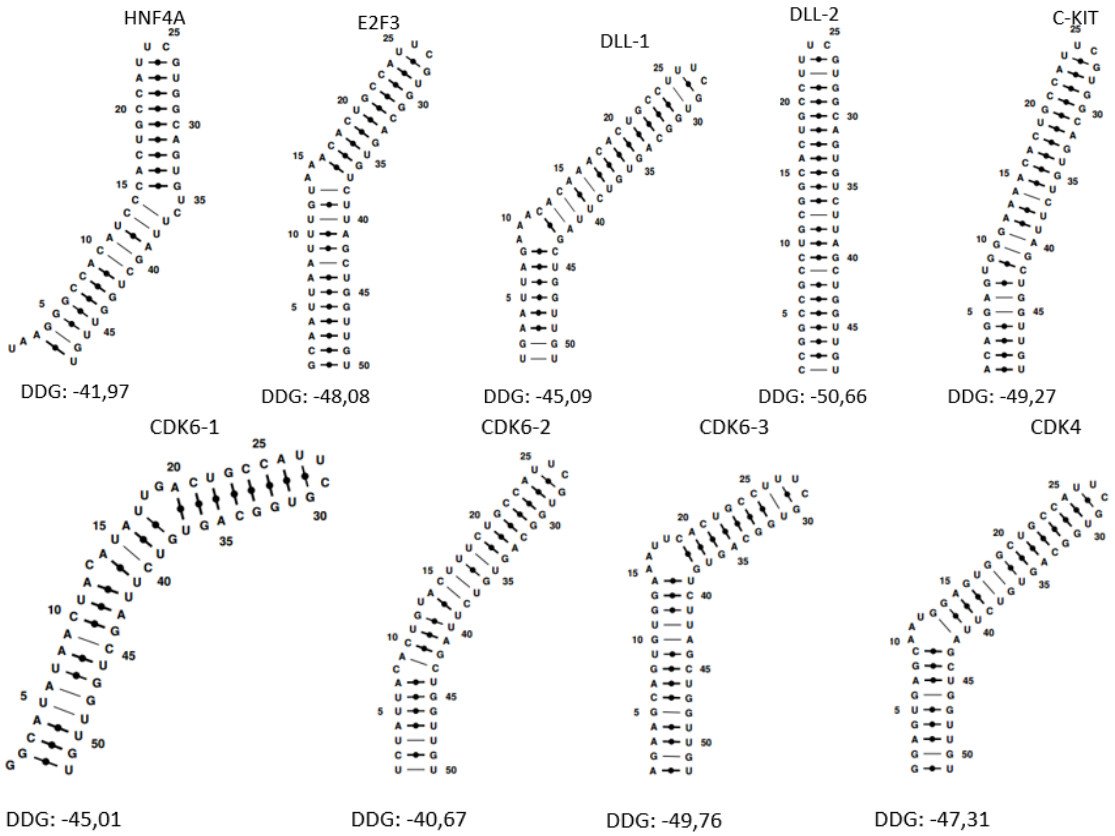
	9	0,14	0,06	0,03	0,002	16	16	3	3	3	3	2
	10	0,18	0,06	0,03	0,002	16	16	3	3	3	3	2
	11	0,1	0,072	0,03	0,002	16	16	3	3	3	3	2
	12	0,14	0,072	0,03	0,002	16	16	3	3	3	3	2
	13	0,18	0,072	0,03	0,002	16	16	3	3	3	3	2
	14	0,1	0,08	0,03	0,002	16	16	3	3	3	3	2
	15	0,14	0,08	0,03	0,002	16	16	3	3	3	3	2
	16	0,18	0,08	0,03	0,002	16	16	3	3	3	3	2
	17	0,1	0,092	0,03	0,002	16	16	3	3	3	3	2
	18	0,14	0,092	0,03	0,002	16	16	3	3	3	3	2
	19	0,18	0,092	0,03	0,002	16	16	3	3	3	3	2
	20	0,1	0,1	0,03	0,002	16	16	3	3	3	3	2
	21	0,14	0,1	0,03	0,002	16	16	3	3	3	3	2
	22	0,18	0,1	0,03	0,002	16	16	3	3	3	3	2
	23	0,1	0,01	0,03	0,002	16	16	3	3	3	3	2
	24	0,14	0,01	0,03	0,002	16	16	3	3	3	3	2
	25	0,18	0,01	0,03	0,002	16	16	3	3	3	3	2
	26	0,02	0,06	0,03	0,002	16	16	3	3	3	3	2
	27	0,06	0,06	0,03	0,002	16	16	3	3	3	3	2
	28	0,26	0,06	0,03	0,002	16	16	3	3	3	3	2
	29	0,3	0,06	0,03	0,002	16	16	3	3	3	3	2
	30	0,1	0,06	0,03	0,002	0	16	3	3	3	3	2
	31	0,1	0,06	0,03	0,002	8	16	3	3	3	3	2
	32	0,1	0,06	0,03	0,002	30	16	3	3	3	3	2
	33	0,1	0,06	0,03	0,002	16	0	3	3	3	3	2
	34	0,1	0,06	0,03	0,002	16	20	3	3	3	3	2
	35	0,1	0,006	0,03	0,002	16	16	3	3	3	3	2
	36	0,1	0,008	0,03	0,002	16	16	3	3	3	3	2
	37	0,1	0,01	0,03	0,002	16	16	3	3	3	3	2
	38	0,1	0,012	0,03	0,002	16	16	3	3	3	3	2
	39	0,1	0,016	0,03	0,002	16	16	3	3	3	3	2
	40	0,1	0,02	0,03	0,002	16	16	3	3	3	3	2
	41	0	0,012	0,03	0,002	16	16	3	3	3	3	2
	42	0,02	0,012	0,03	0,002	16	16	3	3	3	3	2
	43	0,26	0,012	0,03	0,002	16	16	3	3	3	3	2
	44	0,32	0,012	0,03	0,002	16	16	3	3	3	3	2
	45	0,1	0,012	0,03	0,002	0	16	3	3	3	3	2
	46	0,1	0,012	0,03	0,002	8	16	3	3	3	3	2
	47	0,1	0,012	0,03	0,002	30	16	3	3	3	3	2
	48	0,1	0,012	0,03	0,002	16	0	3	3	3	3	2
	49	0,1	0,012	0,03	0,002	16	8	3	3	3	3	2

	50	0,1	0,012	0,03	0,002	16	20	3	3	3	3	2
	51	0,38	0,012	0,03	0,002	16	16	3	3	3	3	2
	52	0,44	0,012	0,03	0,002	16	16	3	3	3	3	2
	53	0,52	0,012	0,03	0,002	16	16	3	3	3	3	2
	54	0,32	0,012	0,03	0,002	30	16	3	3	3	3	2
	55	0,32	0,012	0,03	0,002	36	16	3	3	3	3	2
	56	0,32	0,012	0,03	0,002	42	16	3	3	3	3	2
	57	0,32	0,012	0,03	0,002	16	0	3	3	3	3	2
	58	0,32	0,012	0,03	0,002	16	8	3	3	3	3	2
	59	0,32	0,012	0,03	0,002	16	30	3	3	3	3	2
	60	0,02	0,012	0,03	0,002	16	32	3	3	3	3	2
	61	0,02	0,012	0,03	0,002	30	16	3	3	3	3	2
	62	0,32	0,016	0,03	0,002	16	16	3	3	3	3	2
	63	0,52	0,016	0,03	0,002	16	16	3	3	3	3	2
	64	0,68	0,016	0,03	0,002	16	16	3	3	3	3	2
	65	0,84	0,016	0,03	0,002	16	16	3	3	3	3	2
	66	0,52	0,016	0,03	0,002	30	16	3	3	3	3	2
	67	0,52	0,016	0,03	0,002	36	16	3	3	3	3	2
	68	0,52	0,016	0,03	0,002	16	30	3	3	3	3	2
	69	0,52	0,016	0,03	0,002	16	36	3	3	3	3	2
WNT1	1	0,02	0,03	0,03	0,002	16	6	3	3	3	3	2
	2	0,02	0,03	0,03	0,002	20	6	3	3	3	3	2
	3	0,02	0,03	0,03	0,002	24	6	3	3	3	3	2
	4	0,02	0,03	0,03	0,002	20	3	3	3	3	3	2
	5	0,02	0,03	0,03	0,002	20	12	3	3	3	3	2
	6	0,04	0,03	0,03	0,002	16	6	3	3	3	3	2
	7	0,04	0,03	0,03	0,002	20	6	3	3	3	3	2
	8	0,04	0,03	0,03	0,002	24	6	3	3	3	3	2
	9	0,04	0,03	0,03	0,002	20	3	3	3	3	3	2
	10	0,04	0,03	0,03	0,002	20	12	3	3	3	3	2
	11	0,06	0,03	0,03	0,002	16	6	3	3	3	3	2
	12	0,06	0,03	0,03	0,002	20	6	3	3	3	3	2
	13	0,06	0,03	0,03	0,002	24	6	3	3	3	3	2
	14	0,06	0,03	0,03	0,002	20	3	3	3	3	3	2
	15	0,06	0,03	0,03	0,002	20	12	3	3	3	3	2
	16	0,08	0,03	0,03	0,002	16	6	3	3	3	3	2
	17	0,08	0,03	0,03	0,002	20	6	3	3	3	3	2
	18	0,08	0,03	0,03	0,002	24	6	3	3	3	3	2
	19	0,08	0,03	0,03	0,002	20	3	3	3	3	3	2
	20	0,08	0,03	0,03	0,002	20	12	3	3	3	3	2
	21	0,1	0,04	0,03	0,002	16	8	3	3	3	3	2

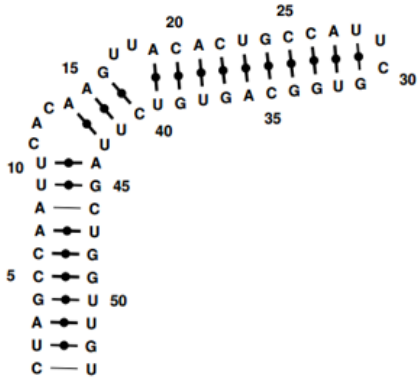
	22	0,14	0,04	0,03	0,002	16	8	3	3	3	3	2
	23	0,18	0,04	0,03	0,002	16	8	3	3	3	3	2
	24	0,1	0,052	0,03	0,002	16	8	3	3	3	3	2
	25	0,14	0,052	0,03	0,002	16	8	3	3	3	3	2
	26	0,18	0,052	0,03	0,002	16	8	3	3	3	3	2
	27	0,1	0,06	0,03	0,002	16	8	3	3	3	3	2
	28	0,14	0,06	0,03	0,002	16	8	3	3	3	3	2
	29	0,18	0,06	0,03	0,002	16	8	3	3	3	3	2
	30	0,04	0,04	0,03	0,002	16	8	3	3	3	3	2
	31	0,06	0,04	0,03	0,002	16	8	3	3	3	3	2
	32	0,08	0,04	0,03	0,002	16	8	3	3	3	3	2
	33	0,1	0,072	0,03	0,002	16	8	3	3	3	3	2
	34	0,14	0,072	0,03	0,002	16	8	3	3	3	3	2
	35	0,18	0,072	0,03	0,002	16	8	3	3	3	3	2
	36	0,1	0,08	0,03	0,002	16	8	3	3	3	3	2
	37	0,14	0,08	0,03	0,002	16	8	3	3	3	3	2
	38	0,18	0,08	0,03	0,002	16	8	3	3	3	3	2
	39	0,1	0,04	0,03	0,002	20	8	3	3	3	3	2
	40	0,1	0,04	0,03	0,002	24	8	3	3	3	3	2
	41	0,1	0,04	0,03	0,002	28	8	3	3	3	3	2
	42	0,1	0,04	0,03	0,002	16	0	3	3	3	3	2
	43	0,1	0,04	0,03	0,002	16	4	3	3	3	3	2
	44	0,1	0,04	0,03	0,002	16	8	3	3	3	3	2
	45	0,1	0,04	0,03	0,002	16	12	3	3	3	3	2
	46	0,1	0,04	0,03	0,002	16	16	3	3	3	3	2
	47	0,1	0,04	0,03	0,002	16	20	3	3	3	3	2
	48	0,14	0,018	0,03	0,002	16	8	3	3	3	3	2
	49	0,24	0,018	0,03	0,002	16	8	3	3	3	3	2
	50	0,02	0,024	0,03	0,002	0	8	3	3	3	3	2
	51	0,02	0,024	0,03	0,002	8	8	3	3	3	3	2
	52	0,02	0,024	0,03	0,002	24	8	3	3	3	3	2
	53	0,02	0,024	0,03	0,002	16	0	3	3	3	3	2
	54	0,02	0,024	0,03	0,002	16	16	3	3	3	3	2
	55	0,02	0,024	0,03	0,002	16	24	3	3	3	3	2
	56	0,02	0,024	0,03	0,002	16	32	3	3	3	3	2
	57	0,14	0,018	0,03	0,002	16	0	3	3	3	3	2
	58	0,14	0,018	0,03	0,002	16	16	3	3	3	3	2
	59	0,14	0,018	0,03	0,002	16	24	3	3	3	3	2
	60	0,14	0,018	0,03	0,002	16	32	3	3	3	3	2
	61	0,14	0,018	0,03	0,002	0	8	3	3	3	3	2
	62	0,14	0,018	0,03	0,002	8	8	3	3	3	3	2

	63	0,14	0,018	0,03	0,002	24	8	3	3	3	3	2
	64	0,06	0,018	0,03	0,002	16	8	3	3	3	3	2
	65	0,1	0,018	0,03	0,002	16	8	3	3	3	3	2
	66	0,18	0,018	0,03	0,002	16	8	3	3	3	3	2
	67	0,22	0,018	0,03	0,002	16	8	3	3	3	3	2
	68	0,14	0,012	0,03	0,002	16	8	3	3	3	3	2
	69	0,14	0,008	0,03	0,002	16	8	3	3	3	3	2

Table 2. All transcription reactions numbered accordingly for each mRNA: CD44, gCD44 and WNT1.

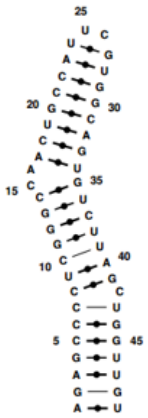


CCNE2



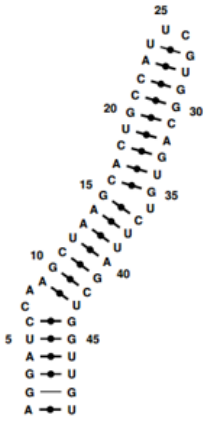
DDG: -48,2

BIRC5



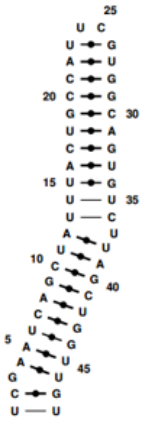
DDG: -47,85

AXL



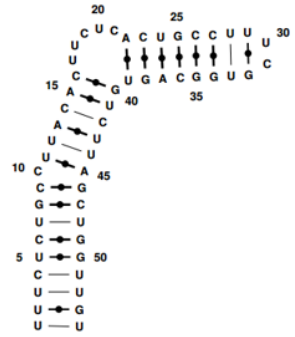
DDG: -52,71

BCL2



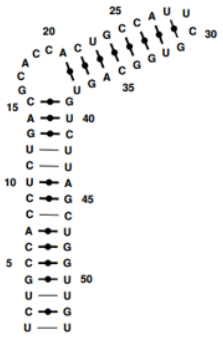
DDG: -47,23

ARHGAP1



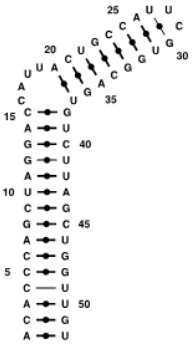
DDG: -42,09

LDHA



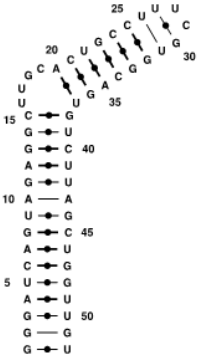
DDG: -51,04

SIRT1



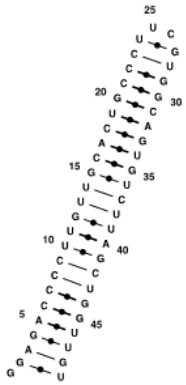
DDG: -41,5

SIRT6



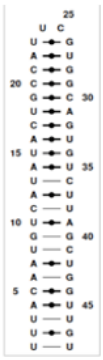
DDG: -47,02

WNT1



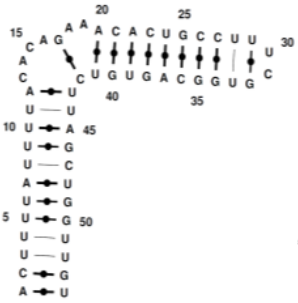
DDG: -44,84

CCND1



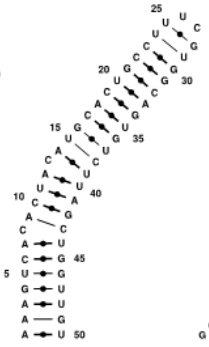
DDG: -44,18

NOTCH-1



DDG: -46,08

PNUTS



DDG: -45,94

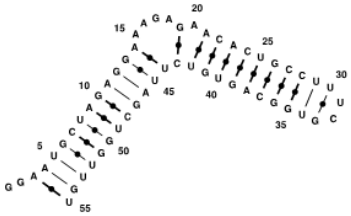
CD44

DDG: -48,26

NOTCH2

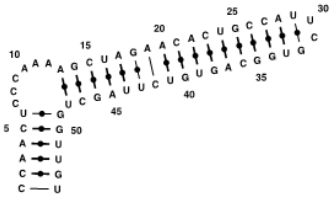
DDG: -42,31

ACSL1



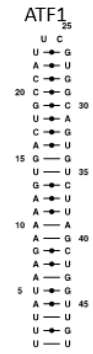
DDG: -48,37

ACSL4



DDG: -54,62

ATF1



DDG: -47,52

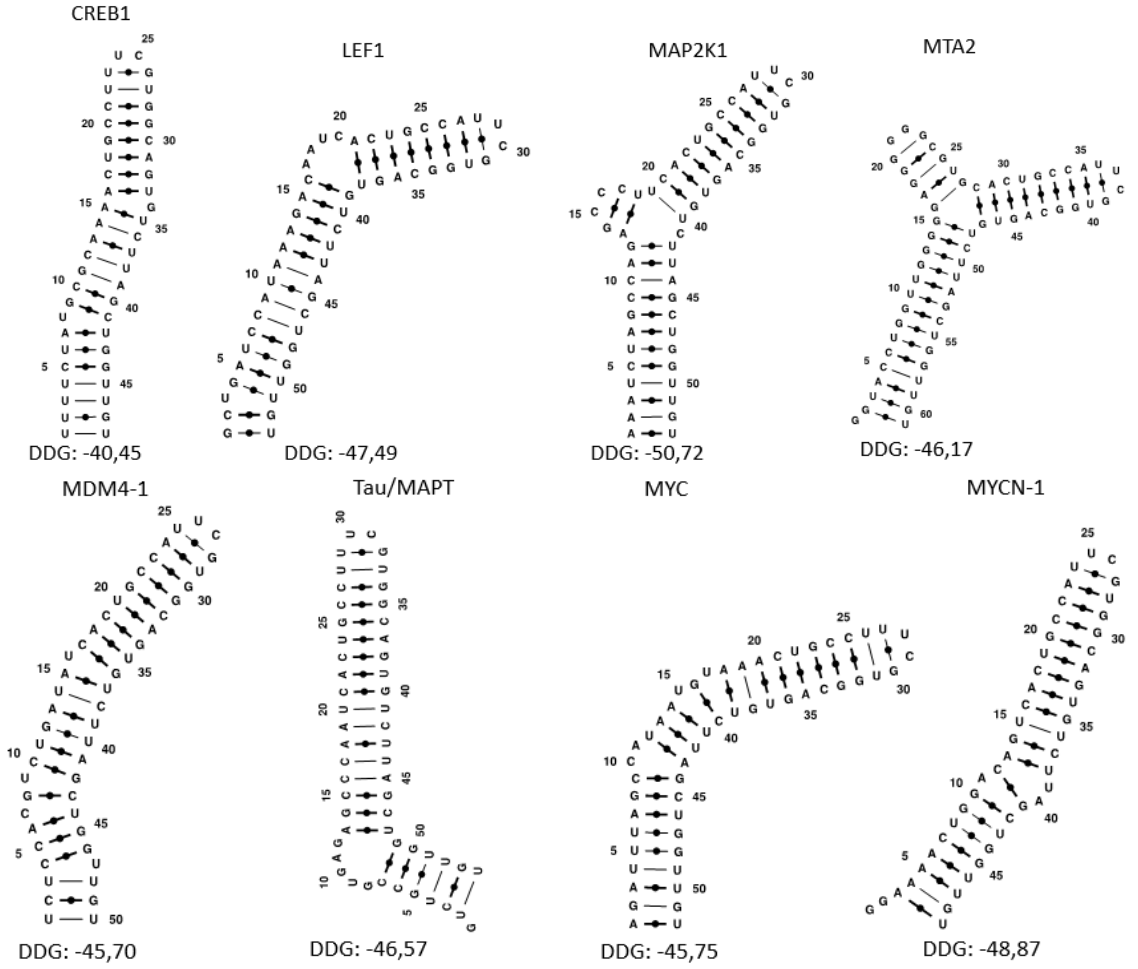


figure 21. All 34 simulated structures with MCfold (Parisien, M., & Major, F., 2008). Strand on the left corresponds to mRNA and right to miRNA. A trend is observed in general, where a kink on the miRNA side is present.