Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 2508

Insights into Musashi-1 interactions with RNA

From in vitro kinetics to cell biology

GUILLERMO E. PÉREZ ROPERO

ACTA UNIVERSITATIS
UPSALIENSIS
2025

ISSN 1651-6214
ISBN 978-91-513-2408-1
urn:nbn:se:uu:diva-552069

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in A1:107, BMC, Husargatan 3, Uppsala, Friday, 25 April 2025 at 09:00 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Ruth Brenk (University of Bergen).

Abstract

Pérez Ropero, G. E. 2025. Insights into Musashi-1 interactions with RNA. From *in vitro* kinetics to cell biology. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2508. 78 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-2408-1.

The interaction between proteins and RNA has gained increased attention as a key regulatory mechanism of RNA translation beyond conventional postulated mechanisms. It is also involved in various pathologies, including cancer and neurodegenerative disorders. In this thesis, the interactions between the RNA-binding protein Musashi-1 (MSI1) and both linear and hairpin-like RNA strands have been investigated using time-resolved and structural technologies, such as SPR biosensors, LigandTracer, and NMR.

Initially, the interactions of the two RNA recognition motifs (RRM1 and RRM2) in MSI1 were characterized separately. Both motifs exhibited high affinity for the UAG motif, with no significant difference between linear and hairpin-like RNA. These interactions followed a 1:1 binding mechanism, with relatively fast, but well-defined association and dissociation rate constants. Next, the interaction of the N-terminal region of MSI1, containing both RRMs, was analyzed, revealing a bivalent binding mechanism. Commercial software was not suitable for distinguishing the fast monovalent and slow bivalent interactions. Therefore, a novel method was developed to quantify kinetic rate constants and binding affinities.

To explore the determinants for the specificity of the interaction, RNA mutants and protein variants were designed to shift RRM2 recognition from the UAG to the CAG motifs. Comprehensive characterization confirmed the impact of these mutations and RNA structural modifications on the interaction.

Additionally, to examine MSI1 regulatory role in synthetic biology, a post-transcriptionally controlled circuit was engineered in E. coli using MSI1 binding to RNA strands. This system, developed on solid media, utilized sfGFP fluorescence as a reporter. Data analysis using the Gompertz growth model correlated well with previously published data on bacteria in suspension. The regulatory effect of oleic acid, a potential MSI1 allosteric inhibitor, was confirmed using this assay. Similarly, MSI1 inhibition effects were assessed in HCT-116 colorectal cancer cells, revealing that both oleic acid and luteolin impacted cell proliferation and CD44v6 receptor expression.

In conclusion, this work has resulted in novel methods and insights for studying and understanding the kinetics and mechanisms of protein-RNA interactions.

Keywords: Interaction kinetics, biochemistry, RNA-protein interactions, RNA binding proteins, synthetic biology

Guillermo E. Pérez Ropero, Department of Chemistry - BMC, Biochemistry, Box 576, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-2408-1
URN urn:nbn:se:uu:diva-552069 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-552069)

A mis padres Ángela y Emilio
y a mi hermana Laura

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. **Pérez-Ropero, G**., Pérez-Ràfols, A., Martelli, T., Danielson, U. H., Buijs, J. (2024) Unraveling the Bivalent and Rapid Interactions Between a Multivalent RNA Recognition Motif and RNA: A Kinetic Approach. *Biochemistry*. 63, 21, 2816–2829

II. Pérez-Ràfols, A., **Pérez-Ropero, G**., Cerofolini, L., Sperotto, L., Roca-Martínez, J., Anahí Higuera-Rodríguez, R., Russomanno, P., Kaiser, W., Vranken, W., Danielson, U. H., Provenzani, A., Martelli, T., Sattler, M., Buijs, J., Fragai, M. (2024) Deciphering the RNA recognition by Musashi-1 to design protein and RNA mutants for in vitro and in vivo applications. *bioRxiv. Submitted, Nucleic Acid Research.*

III. **Pérez-Ropero, G**., Dolcemascolo, R., Pérez-Ràfols, A., Andersson, K., Danielson, U. H., Rodrigo, G., Buijs, J. (2025) Regulatory Effects of RNA–Protein Interactions Revealed by Reporter Assays of Bacteria Grown on Solid Media*. Biosensors.* 15(3), 175

IV. **Pérez-Ropero, G.**, Ortstad, N., Nestor, M., Danielson, U. H., Buijs, J. Exploring real-time interaction analysis to monitor CD44 expression modulation in cancer cells by inhibiting Musashi-1 interactions with RNA. *Manuscript*

Reprints were made with permission from the respective publishers.

Papers not included in this Thesis

Devignes, M-D., Smaïl-Tabbone, M., Dhondge, H., Dolcemascolo, R., Gavaldá-García, J., Anahí Higuera-Rodriguez, R., Kravchenko, A., Roca Martínez, J., Messini, N., Pérez-Ràfols, A., **Pérez Ropero, G**., Sperotto, L., Chauvot de Beauchêne, I., Vranken, W (2023) Experiences with a training DSW knowledge model for early-stage researchers. *Open Research Europe*.

Dolcemascolo, R., Heras-Hernández, M., Goiriz, L., Montagud-Martínez, R., Requena-Menéndez, A., Ruiz, R., Pérez-Ràfols, A., Anahí Higuera-Rodríguez, R., **Pérez-Ropero, G**., Vranken, W., Martelli, T., Kaiser, W., Buijs, J., Rodrigo, G. (2024) Repurposing the mammalian RNA-binding protein Musashi-1 as an allosteric translation repressor in bacteria. *eLife.*

Agosta, L., Fiore, L., Colozza, N., **Pérez-Ropero, G**., Lyubartsev, A., Arduini, F., Hermansson, K. (2024) Adsorption of Glycine on TiO2 in Water from On-the-fly Free-Energy Calculations and In Situ Electrochemical Impedance Spectroscopy. *Langmuir.*

Contribution report

Paper I	Conceptualized the study. Designed, performed and analysed SPR-biosensor experiments. Wrote the manuscript with input from all the co-authors. Corresponding author.
Paper II	Designed, performed and analysed SPR-biosensor experiments. Co-wrote the manuscript with input from the other authors.
Paper III	Conceptualized the study. Designed, performed and analysed RT-PEA experiments. Wrote the manuscript with input from all the co-authors. Corresponding author.
Paper IV	Conceptualized the study. Designed and supervised RT-CBA experiments performance and analysis. Designed, performed and analysed SPR-biosensor experiments. Wrote the manuscript with input from all the co-authors.

Contents

This thesis was written wholly by the author, Guillermo E. Pérez Ropero, alone without any content generated by artificial intelligence (AI) tools.

Abbreviations

COX	Cyclooxygenase
CSD	Cold-shock domain
DKK3	Dickkopf-related protein 3
DNA	Deoxyribonucleic acid
dsRNA	Double strand RNA
EC_{50}	Half maximal effective concentration
ePABP	Embryonic polyadenylate-binding protein 2
IC_{50}	Half maximal inhibitory concentration
IDR	Intrinsically disordered region
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
k_a	Association rate constant
k_d	Dissociation rate constant
K_A	Equilibrium association constant
K_D	Equilibrium dissociation constant
KH	K-homology
LT	LigandTracer
lncRNA	Long non-coding RNA
mRNA	Messenger RNA
miRNA	MicroRNA
MSI1	Musashi-1
NMR	Nuclear magnetic resonance
RBP	RNA binding protein
RBD	RNA binding domains
RNA	Ribonucleic acid
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RT-CBA	Real-time cell-based assay
RT-PEA	Real-time protein Expression assay
SCD	Stearoyl-CoA desaturase
siRNA	Small interfering RNA
sfGFP	Superfolder green fluorescent protein
SPR	Surface plasmon resonance
ssRNA	Single strand RNA
tRNA	Transfer RNA
xrRNA2	Exoribonuclease-resistant RNA2

Introduction

When we look at a photograph, we open the door to a condensation of time that contains some information. It can be used to illustrate some news, to preserve a good memory or to teach others something new. Although it can be a true representation of what we want to transmit, preserve, or research, it can also be misleading. One way of grasping or obtaining more knowledge would be to learn how that moment evolves over time. The introduction of the variable of time allows us to grasp subtle or major changes that are not obvious in a static context. In this case, it would correspond to watching a video of the whole sequence, instead of just a single frame.

Although this reflection can be obvious to anybody, we face the same challenge when doing research. This thesis is about time, and mainly about how different biomolecules, in this case proteins and ribonucleic acid (RNA), interact over time. Instead of extracting information based on a single or a few time points, here the analysis is focused on how these molecules get together, how fast they form complexes and how long it takes for them to fade away. This way of interacting can resemble human interactions, with some being quick in recognition and dissociation, while others tend to form more stable complexes, that last longer over time. The study of kinetics, which is the study of movement over time, has a deep implication on how these molecules can exert their biological functions, and how the alteration of interaction patterns can lead to pathological states. Moreover, the study of the interaction kinetics is of capital importance for any molecule that aspires to become a drug. Let us now see how the interactions between a human RNA binding protein, Musashi-1 (MSI1), and various RNA strands evolve over time, and their implications for biological systems.

Molecular interactions

Molecular interactions comprise the combined effects of all attractive and repulsive forces that are established between different molecules. These forces depend on the chemical nature of the molecules that interact. They may form reversible or irreversible interactions. An example of an irreversible reaction would be the formation of covalent bonds, as happens when acetylsalicylic

acid interacts with its target, resulting in acetylation of a serine residue of cyclooxygenase (COX) enzymes, thereby permanently inhibiting the enzyme[1].

In this thesis the focus is on the study of reversible interactions between RNA and proteins. The forces involved in establishing reversible intermolecular interactions are of different nature, being hydrogen bonds, van der Waal forces, hydrophobic effects and π-stacking as the main forces of interest for this thesis.

Hydrogen bonds can form between hydrogen and an electronegative atom, usually fluorine, oxygen or nitrogen, with a strength of 0.5 to 8 kcal/mol[2]. Both nitrogen and oxygen are abundant in biomolecules, making hydrogen bonds one of the most common interaction forces. Van der Waal forces can be divided into several groups depending on its exact nature such as Debye forces, London dispersion forces and a series of dipole-dipole interactions[3]. These forces are also of electrostatic nature but are weaker than hydrogen bonds, with energy values ranging from 0.5 to 1 kcal/mol and can occur at longer distances (more than 3 Å)[4]. Hydrophobic effects arise between chemical structures that are non-polar and therefore have an unfavourable interaction with polar water molecules in solution. The interaction between two hydrophobic structures, rather than water, results in a driving force that promotes the binding between organic molecules and the formation of stable structures in proteins[5]. Finally, π-stacking refers to the interaction between electron dense aromatic groups, these are abundant in RNA and proteins[6].

RNA-protein interactions

The nature of the interactions between RNA and proteins is a consequence of the characteristics of both biomolecules.

RNA structure and types

RNA is a polymer of nucleotides connected through phosphodiester bonds. These nucleotides are composed by ribose, a nitrogenous base (adenine, cytosine, guanine or uracil) and a phosphate group. Along with deoxyribonucleic acid (DNA), RNA is one of the nucleic acids, primarily responsible for storing and translating genetic information. However, RNA is both structurally and functionally different when compared to DNA. Structurally, RNA is composed of a sugar of different nature than the one present in DNA, *i.e.*, ribose instead of deoxyribose. Ribose presents a hydroxyl group in the 2' position that is indispensable for interactions with other molecules, including proteins. Another difference resides on the secondary structure adopted by the biomolecules; while two DNA strands form a double-helix, RNA is mainly present as a single strand, with a wide range of possible conformations. This allows RNA to display several functions beyond its basic role as mediator of protein synthesis. These roles may be structural (tRNA, rRNA), catalytic (ribozymes)

and regulatory[7]. In this last group we find most non-coding RNAs, which are the most abundant type of RNA, including long non-coding RNA (lncRNA), microRNA (miRNA) and small interfering RNA (siRNA)[8].

The formation of RNA secondary structures is mainly driven by intramolecular interactions within the same RNA strand. Conventionally, Watson-Crick base interactions have been considered as the main contributors to RNA structure, while the focus has shifted nowadays into π-stacking and hydrophobic effects as the main contributors[9]. The variety of conformations that RNA adopts[10] (Figure 1) are intimately related to their role as regulators[11]. In the case of mRNA, the formation of different structures influences its processing[12], translation[13], decay[14], localization[15] and the formation of cellular condensates such as stress granules[16].

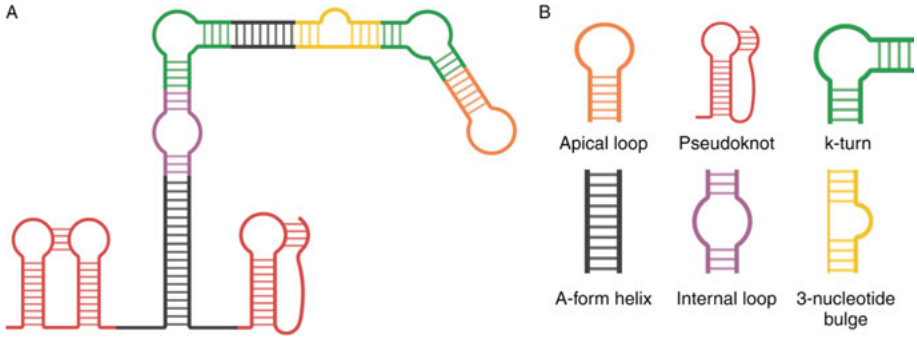


Figure 1. A) Structural landscape of a folded RNA strand, with different secondary structures highlighted in different colours. B) Common RNA secondary structures. Created in BioRender. Perez, G. (2025) https://BioRender.com/x07r548

RNA binding proteins

Proteins are composed of amino acid chains linked through peptide bonds. If the number of amino acids is between 2 and 50, they are considered peptides[17], while longer chains of amino acids proteins are referred to as proteins. Proteins present different structural levels that influence in how they interact with other biomolecules. Secondary structures arise locally through the establishment of hydrogen bonds between carbonyl and amino groups present in the amino acid residues, resulting in α-helixes, β-sheets or β-turns. When the structure does not have a recognizable secondary structure, it is considered to be an intrinsically disordered structure. Tertiary structure refers to the establishment of recognizable subunits or domains based on the interactions between secondary structures, while the quaternary structure applies to multidomain proteins and how these domains assemble.

RNA binding proteins (RBPs) are proteins that interact with both coding and non-coding single-strand RNA (ssRNA) or double-strand RNA (dsRNA)[18]. Some of them are also able to bind ssDNA as K-homology (KH), Zinc fingers or cold-shock domains (CSD) domains[4]. Most RBPs present RNA binding domains (RBDs) that are relatively short, typically less than 100 amino acids in length. These domains have only few residues involved in RNA binding and usually recognize short RNA motifs just a few nucleotides long[19]. In this context, specificity is often not driven by the RNA-binding motif alone, as many motifs bind to different proteins. Instead, it is driven by the presence of multiple RBDs within each protein[20].

A variety of RBDs in different combinations occur in RBPs, which makes it sometimes difficult to characterize and classify them. In addition, they often have a high prevalence of intrinsically disordered regions (IDRs) that also interact with RNA[21]. IDRs also play roles in the formation of cellular condensates which have significant regulatory consequences[22,23].

RBPs are present in all kingdoms of life, with more than 3 million proteins identified across different species[24] . More than 3000 RBPs have been identified in humans[24]. Of all RBPs families, the most abundant is the RNA Recognition Motif (RRM) family which encompasses proteins harbouring RRM domains.

RNA Recognition Motif (RRM)

Proteins in the RRM family are characterized by containing RRM domains. RRM domains are composed of around 90 amino acids that presents a well-conserved topology defined by four antiparallel β-sheets between two α-helixes (βαββαβ)[25] (Figure 2). This domain structure has been evolutionary preserved in different organisms. Although the overall structure has been well conserved, the amino acid sequences of the RRM domains vary between different proteins[26], influencing their recognition of different RNA motifs. The most conserved residues in the RRM domains are aromatic amino acids in the β-sheets that establish π-stacking interactions with the RNA nitrogenous bases[25,27].

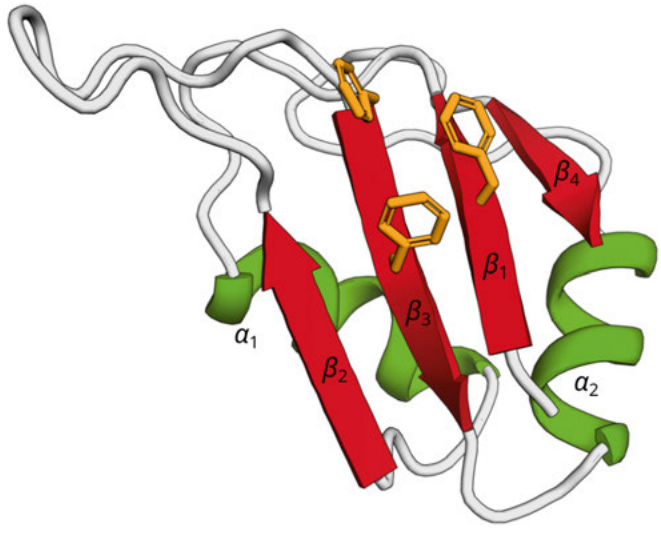


Figure 2. Structure of an RRM domain with secondary structures and conserved residues highlighted. From AlphaFold (AF-O43347) and modified using Pymol.

RRM domains are usually part of multidomain proteins. This is common for RNA binding proteins as they interact with short RNA motifs that are abundant across different RNA strands. The lack of specificity in the binding motif is compensated by the simultaneous binding of different domains[28]. The IDRs and other regions involved in protein-protein interactions also contribute to the specificity and strength of the binding.

Biochemistry of the interactions

As mentioned above, RBP-RNA interactions are characterized by relatively weak interactions between the RNA strand, the amino acid residues and the protein backbone[4]. These interactions are mostly mediated by hydrogen bonds and van der Waals forces, which arise from the polar groups in the nitrogenous bases, the 2'OH in the ribose, and the phosphate groups. Polar amino acids, such as serine and asparagine, as well as positively charged amino acids (lysine and arginine) can establish hydrogen bonds from the protein sidechain[29].

Hydrophobic effects arise as a result of non-polar nitrogenous bases in RNA and hydrophobic residues of the protein. Π-stacking interactions are relatively strong interactions compared with the previously mentioned interactions and occur between the aromatic rings of the nitrogenous bases of the RNA and the aromatic amino acids (tryptophan, histidine, tyrosine and phenylalanine)[30]

Musashi-1

The Musashi family of proteins encompasses Musashi-1 and Musashi-2, that have a 69% of sequence identity[31]. They are both RRM proteins that bind to single stranded mRNA and regulate protein expression. In this thesis the focus is on human Musashi-1.

Human Musashi-1 structure and interaction with RNA

Musashi-1 (MSI1) is 362 amino acids long, 39 kDa multidomain protein with two RRM domains, RRM1 and RRM2 (Figure 3). The two domains are united by a flexible linker and a long IDR tail (residues 194 to 362) is positioned in the C-terminal region.

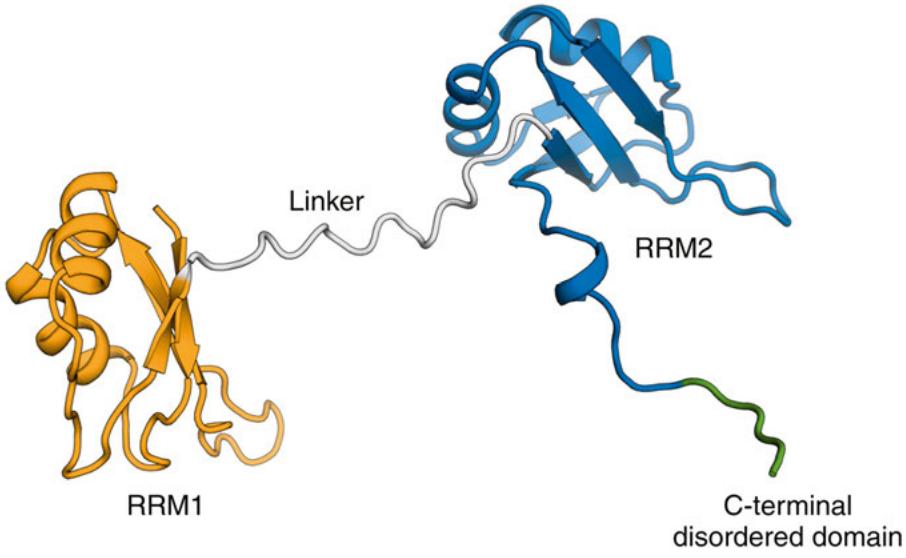


Figure 3. AlphaFold-predicted structure of the truncated MSI1 (AF-O43347) used in this thesis (residues 1-200). The different regions of the protein are; in yellow, RRM1; in grey, the linker; in blue, RRM2 and in green, a portion of the disordered tail. Structure rendered using Pymol.

Both RRMs recognize a three-nucleotides core motif UAG[32]. An extended recognition motif ((G/A)$U_{1\text{-}3}$AGU)) has been proposed for RRM1[33]. The interactions of both RRM1[31] and RRM2[34] with RNA have been demonstrated experimentally, although there has been no information on the tandem domain or full length-protein interacting with RNA prior to the currently presented studies.

The interaction between RRM1 and $G^1U^2A^3G^4U^5$ is mainly mediated by π-stacking and hydrogen bonds[31]. G^1 interacts with W29 through π-stacking and

establishes a hydrogen bond with K88. U^2 fits in a pocket formed by a rim with D91, K93 and F23 and G25 and G26 at the bottom. A^3 establishes π-stacking interactions with F23 and F96, with its aromatic ring between both phenylalanine, and a hydrogen bond is established between the carbonyl group of V94 and the nitrogenous base. G^4 establishes a π-stacking interaction with F65 and a hydrogen bond with K21. In the N-terminus, an electrostatic interaction is established between R61 and the phosphate backbone of the RNA molecule. The interaction of RRM2 with the same sequence revealed a similar interaction pattern[34], although with no involvement of G^1. U^2 establishes two hydrogen bonds with K182, and E180, while at the same time it fits in a pocket where K182, E180 and F112 form a rim and G114 and G115 the bottom. A^3 is situated between two F112 and M190, establishing π-stacking interactions, while hydrogen bonds are formed with Q185 and K193. G^4 establishes a π-stacking interaction with F154, hydrogen bonds with Q185 and K110 and a salt bridge between its phosphate group and K187.

Other than the RRM domains, disordered regions present in the protein also play important roles. In MSI1, IDRs are present in the linker between the RRM domains and in the C-terminal tail. The linker is a flexible 14 amino acid long region that, without steric hindrance, allows binding to RNAs with UAG motifs spaced between 1 and 50 nucleotides apart[34]. The MSI1 C-terminal tail is around 200 amino acids long. It can interact with both mRNA[35,36] and poly(A) binding proteins such as the embryonic polyadenylate-binding protein 2 (ePABP)[37].

Both multivalency[38] and the presence of IDRs have been implicated in the formation of cellular condensates, as they thermodynamically favour the formation of non-soluble polymeric structures[39]. These condensates exert a wide range of biological functions, from transcriptional regulation to molecular transport. They are composed of a network of different molecules, including cellular condensates formed by mRNA and proteins, such as Cajal bodies, P condensates, and stress granules[16,40]. These structures are highly dynamic and can quickly be formed in response to stimuli, establishing high local concentration environments that accelerate reaction rates[41] and influence molecular interactions, for example, by favouring molecule rebinding. MSI1 is both a multivalent and partially disordered protein, and it has been related with the formation of stress granules, especially in pathological conditions[42].

Apart from RNA, MSI1 interacts also with other endogenous ligands such as monounsaturated fatty acids, with oleic acid being worth highlighting[43]. Its effects are explained in detail the next section of this thesis. Other non-endogenous ligands of MSI1 are luteolin[44] and gossypol[45] that inhibit the protein binding to RNA in a competitive manner.

Musashi-1 biology

Musashi-1 was first identified as a neural RNA binding protein in *Drosophila*[46] and has since been found in many other eukaryotic organisms, including humans. It has mainly been related to differentiation and proliferation of neural cells in early developmental stages[47,48], primarily by regulating *Notch*[49–52] and *Wnt*[51,52] pathways. MSI1 exerts its function by binding to mRNAs, such as Numb mRNA, and inhibiting its translation[33]. Since Numb is a Notch pathway inhibitor[53], its suppression increases Notch signalling, which is mainly involved in cell proliferation and differentiation at early development stages but is expressed at low levels in adult cells[54]. Another important target of MSI1 is $p21^{WAF\text{-}1/CIP\text{-}1}$, a cyclin-dependent kinase inhibitor that regulates cell cycle progression[48,55]. By binding to p21 mRNA and inhibiting its translation, MSI1 promotes cell cycle progression and increases proliferin-1 levels that, in turn, contributes to the inhibition of Dickkopf-related protein 3 (DKK3)[56], thereby upregulating the *Wnt* pathway. Through these mechanisms, MSI1 enhances cell growth by stimulating *Notch* and *Wnt* pathways.

The role of MSI1 in early developmental processes, as embryogenesis, has been shown in *Xenopus*[57] and mice[35], but not in humans.

Allosteric regulation of MSI1

The allosteric regulation of Musashi-1 binding to RNA by oleic acid and other 18–22 carbon ω-9 monounsaturated fatty acids has been extensively described by Clingman et al[43]. In their study it was proposed that oleic acid binds to a hydrophobic cavity that is present on the RRM1 α-helix with micromolar affinity. The specificity of the binding was driven by the presence of two arginine residues in the vicinity of the pocket, R53 and R61 that interacted with the carboxy terminal of oleic acid. The binding of oleic acid disrupts the interaction between G^1 and W29, thereby decreasing affinity of MSI1 for RNA.

Biologically, it has been postulated that MSI1 acts as a regulator of the synthesis of long monounsaturated fatty acids by inhibiting translation of stearoyl-CoA desaturase (SCD), therefore reducing the production of oleic acid[43]. Fatty acids are especially relevant in glial cells, as astrocytes and oligodendrocytes[58], where MSI1 is expressed in early developmental stages[47]. In cell culture, they successfully showed a reduction of cell proliferation in an oligodendrocyte immortalized cell line[43].

Musashi-1 in disease

In view of MSI1 biology, and its role as a protein that regulates stem cell differentiation and growth, it is not surprising to state its role as an oncoprotein[59,60]. MSI1 overexpression has been detected in several cancers, where it

is usually associated with undifferentiation, stem cell-like properties and a worse prognosis[56,61,62]. Some of the cancers where MSI1 has been described to play a role are glioblastoma[63–65], breast cancer[60,66–68], colorectal cancer[50,62], urothelial carcinoma[42] and gastric cancer[69,70].

Musashi-1 and CD44 regulation

The development of metastasis is associated with advanced cancer stages and a worse prognosis. CD44 is a transmembrane cell surface receptor that interacts with extracellular matrix components, most importantly hyaluronic acid[71]. It comprises a series of ten isoforms created through alternative splicing, whose biological functions are related to cell migration and growth[72]. CD44 overexpression in cancer cells has been related to the appearance of metastasis and angiogenesis[73–75], making it an attractive therapeutic target[76,77]. Although the exact mechanism of action is unclear, MSI1 expression has been correlated with a higher expression of CD44, while the expression of CD44 has been found to be lower when MSI1 was knocked-down.[68,78]

Musashi-1 in Alzheimer's disease and Zika-virus infection

Beyond its role in cancer, MSI1 has been related to Alzheimer's disease[79–81]. In contrast to the previously mentioned mechanism of pathway deregulation, the role of MSI1 in the initiation of Alzheimer's pathology[79] is related to the establishment of insoluble protein aggregates. As mentioned before, MSI1 has the structural characteristics that are needed for the establishment of liquid-liquid phase separation and therefore cellular condensates[82]. The formation of such structures, where the protein concentration is high, has been associated with the potential appearance of fibrils and other protein aggregates[39,83].

Zika-virus is a *Flavivirus* endemic to certain tropical areas of South America, the Caribbean, East Africa and Southeast Asia. Infections can result in severe consequences, especially for pregnant women, causing miscarriages or neonatal microcephaly[84]. Musashi-1 has been identified as a binder of Zika virus mRNA, promoting viral replication and affecting neurodevelopment. Recently it has been proposed that MSI1 binding to exoribonuclease-resistant RNA (xrRNA2) in Zika virus is mediated by a non-canonical tetraloop with an AGAA motif[85]. This highlights the importance of RNA secondary structures in enhancing protein binding.

Interaction kinetics and affinity

All reversible chemical reactions are characterized by the formation of a new species after the interaction between molecules. This process has a characteristic change in the Gibbs free energy (ΔG) (Equation 1) (Figure 4), with two components, an enthalpy change (ΔH) and an entropy change (ΔS).

$$\Delta G = \Delta H - T\Delta S \qquad \text{(Eq. 1)}$$

ΔG can also be expressed in relation to the equilibrium constant, which in case of describing the formation of a complex would be K_A (equilibrium association constant) or K_D (equilibrium dissociation constant) in case of representing the complex break up.

$$\Delta G = -RTlnK_D \qquad \text{(Eq. 2)}$$

In both equations T is the temperature in *K* and *R* is the gas constant in $J \cdot K^{-1} \cdot mol^{-1}$.

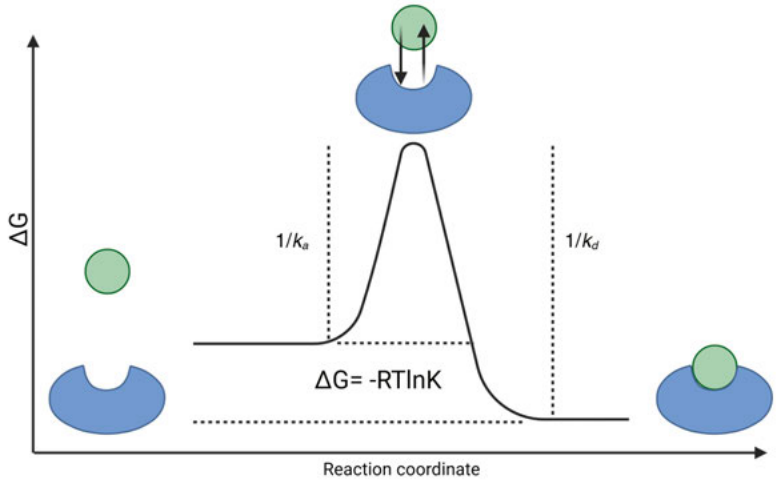


Figure 4. Free energy diagram depicting a single-step binding process. It highlights the relationship between free energy (ΔG), the equilibrium constant (K_D) and the rate constants (k_a and k_d). Created in BioRender. Perez, G. (2025) https://BioRender.com/r40q516

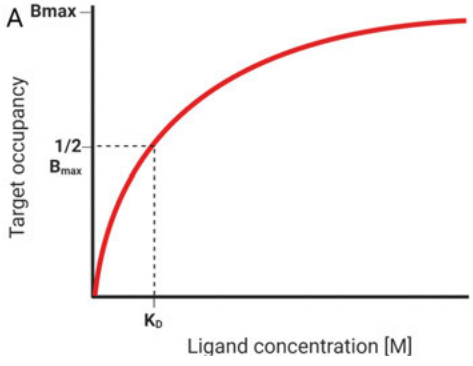
Affinity concept in biological research

The concept of affinity is related to the ratio of the concentrations of the free ligand and target and the formed complex at equilibrium for the interaction process given in Scheme 1. The affinity is usually described as the equilibrium dissociation constant (K_D), with units of concentration (M) as given in Eq. 3. In a situation where the concentration of ligands is much higher than that of targets, the ligand concentration is hardly affected by complex formation. In this context, K_D represents the ligand concentration when half of the targets are bound by a ligand and is, therefore, an intuitive measure that relates the dose of a ligand to the level of target binding at equilibrium. The association

rate constant at equilibrium K_A it is not often used because of it less intuitive M^{-1} unit.

$$[Ligand] + [Target] \rightleftarrows [Ligand{:}Target] \qquad \text{(Scheme 1)}$$

K_D values are thus used as an indicator of the strength of an interaction, where strong interactions with high affinity are correlated with low K_D-values, and *vice versa*. Variations in ligand concentration around the K_D value have the largest impact on target occupancy as shown in Figure 5.



$$K_D = \frac{[L][T]}{[LT]} = \frac{k_d}{k_a} \qquad \text{(Eq. 3)}$$

Figure 5. Relationship between the signal and the ligand concentration, with the K_D being marked as the concentration where half of the targets are occupied. Equation 3, relationship between equilibrium dissociation constant, concentration of interacting molecules at equilibrium and kinetic rate constants. Created in BioRender. Perez, G. (2025) https://BioRender.com/n58i628

Experimentally, K_D-values can be quantified using a wide range of techniques, that can be classified into steady-state or real-time binding approaches. In a classic steady state experiment, the ligand and the target are incubated in a concentration series, and the complex formation is measured after a certain time resulting in a target occupancy vs. ligand concentration graph as shown in Figure 5. For this data to be useful for quantification of K_D values, a certain curvature must be acquired and, ideally, a saturation reached[86].

The relationship between enthalpy, entropy and affinity allows the contribution of both state functions to be determined for each specific interaction. Enthalpy is mainly driven by the formation of specific interactions that are often of electrostatic nature, such as hydrogen bonds or van der Waals forces that guide the formation of the complex, releasing energy. Entropy, however, drives the interaction when it increases the overall disorder of the system, as for example occurs when highly ordered water molecules are released from hydrophobic moieties of the interacting molecules. On other hand, interactions

can decrease the entropy of the system, as newly formed complexes probably have less freedom to move than both interacting molecules separately.

Interaction kinetics

Real-time assays quantify the K_D-value through the ratio between kinetic rate constants (Equation 3). This approach does not require the establishment of an equilibrium state or to reach saturation. However, it requires enough curvature in shape of the bound signal vs time curves that are being monitored, along with other requirements that are further discussed in the Methodology chapter of this thesis.

The term interaction kinetics is associated with the quantification of the association and dissociation rate constants that characterize reversible interactions between molecules. Differently to affinity, that is used as an indicator of the interaction strength, the rate constants add a new layer of information. The association rate constant (k_a) gives information on the recognition of the interaction partners[87], this means the probability of both molecules, once in proximity, to form a complex. The concentrations do not influence the recognition, but at higher concentrations, the likelihood of an encounter increases. The dissociation rate constant (k_d) gives information on the stability of the formed complex[88]. It is important to highlight that molecules with completely different rate constant relationship can have the same affinity value[89] (Figure 6A, 6B). This means that different biological effects, as far as they depend on the rate constants, can be achieved with molecules with a similar affinity[90]. This makes interaction kinetic characterization critical in drug discovery and lead optimization[91].

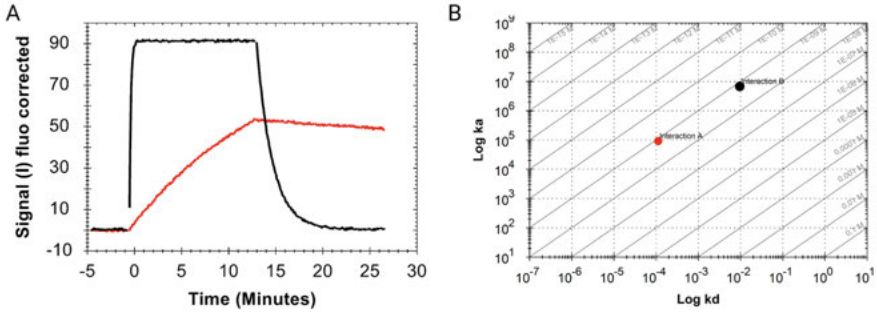


Figure 6. A) Kinetic traces of two interactions with same K_D value and different kinetic profile. B) k_{on}-k_{off} plot showing the interactions in part A, it can be observed that both have the same K_D but different rate constants. Created in BioRender. Perez, G. (2025) https://BioRender.com/d67z895

When optimising therapeutic ligands, it has been a standard procedure to focus on obtaining slow dissociation rate constants and thus long residence times, to enhance biological effects[92,93], while nowadays, there is a trend to also optimize association rate constants[94,95]. Despite efforts in understanding and optimizing kinetic parameters, their relation to pharmacokinetics presents discordances[95].

Implications for biological regulation and avidity

In a biological setting, K_D values are not just giving information about the strength of an interaction but are related to the biological effect that a specific interaction can trigger. In drug discovery, affinity quantification and optimization are typically performed prior to *in vivo* experiments. For that reason, it is key to find correlations between K_D, and EC_{50} or IC_{50} values, to understand at which ligand concentrations certain biological effects are obtained.

Molecules that have more than one binding domain and thus can establish several interactions simultaneously are referred as multivalent. In this case valency refers to the number of interactions, *i.e.,* MSI1 has two RRM domains and therefore can establish bivalent interactions with RNA. Each of these interactions is defined by a specific affinity, however, is the combined strength of all interactions what characterizes the overall interaction of the molecule with its target. This combined strength is what is defined as avidity, and it is often translated into interactions that present a slower dissociation rate, thus forming more stable complexes than the individual binding domains. For a bivalent interaction, the effect of avidity is explained by the increased local concentration of the second interaction pairs, and the high probability of rebinding rather than dissociation. This prevents the complex to dissociate as both domains would need to do it simultaneously, therefore rendering a more stable complex.

Multivalency is a common feature of RBPs, and it has been proposed to enhance the recognition of short and common RNA motifs[96,97]. Moreover, this multivalency has been described to play a key role in cellular condensates[98–100], helping to establish high local concentrations that could act as fast regulatory centres in situations such as cellular stress[101].

Aims of the Thesis

The work presented in this thesis was performed within the RNAct Innovative Training Network (ITN) a Marie Skłodowska-Curie Action (MSCA) and should be understood as a piece in a collective scientific effort to expand our understanding of RNA-protein interactions.

The aims of the thesis were to:

I. Characterize the kinetic parameters and affinities of Musashi-1 RRM1, RRM2 and the tandem protein interactions with different RNAs, with special emphasis on resolving the contributions of the RRMs to the kinetics contribution to the binding. **Papers I** and **II**.

II. Design new protein variants with specificities for a different RNA motif. **Paper II**.

III. Develop a new cell-based assay for evaluation of RNA binding proteins. **Paper III**.

IV. Determine the impact of allosteric modulation of Musashi-1 binding to RNA and its impact in a cancer cell model. **Papers III** and **IV**.

Methodology

In this thesis, several techniques have been used to characterize the interaction between MSI1 and RNA, as well as to investigate the potential regulatory effects of MSI1. The techniques range from biophysical *in vitro* techniques, using purified molecules, to methods using living cells, including both bacteria and mammalian cell lines.

When describing the use of different techniques, such as real time interaction analysis methods, there are differences in nomenclature worth to mentioning. For Surface Plasmon Resonance (SPR)-based biosensors, the interaction partners are denoted ligand (the immobilized molecule) and analyte (the molecule in solution). This is because SPR technology was established in a chromatographic setting, and it was logical to use the same nomenclature as for affinity chromatography. However, when performing binding assays using membranes or living cells, the ligand is the molecule in solution, and the molecules present on the surface are the receptors, often called targets, more in consonance with the biological nomenclature used in drug discovery. When using techniques were both molecules are in solution, as NMR spectroscopy, the ligand is referred as the smaller molecule or the molecule whose concentration varies, while the target is usually a larger molecule which is present in a constant concentration.

Surface Plasmon Resonance (SPR) biosensors

Surface plasmon resonance (SPR) is the physical principle behind a class of optical-based biosensor devices that are widely used to characterize interactions[102], with special focus on interaction kinetics. Surface plasmon resonance can occur on certain metal surfaces (such as gold) when hit by a beam of light that induces the formation of electron (plasmon) waves that resonate on the surface[103]. These resonating plasmons absorb the incident light at a certain angle[103] (Figure 7A). When the metal layer is between two compartments with different refractive indexes, such as glass and a liquid, the change on the refracted light angle (θ_r) that is absorbed by the plasmons reflects the fluctuations of refractive index and thus the concentration of molecules in the liquid in contact with the metal layer[104]. This principle is used to characterize binding

events with high sensitivity. Molecules denoted as ligands are immobilized on the surface and different concentrations of the analyte are injected over it in a continuous flow of buffer. Changes in the angle where the incident light has a minimum due to the surface plasmon phenomena as a function of time (Figure 7B) are translated into sensorgrams, that are graphic representations of the interaction binding events (Figure 7C). The signal is expressed as RU (resonance unit) and 1 RU corresponds to roughly 1 pg/mm^2 of organic molecules bound to the sensor chip[105].

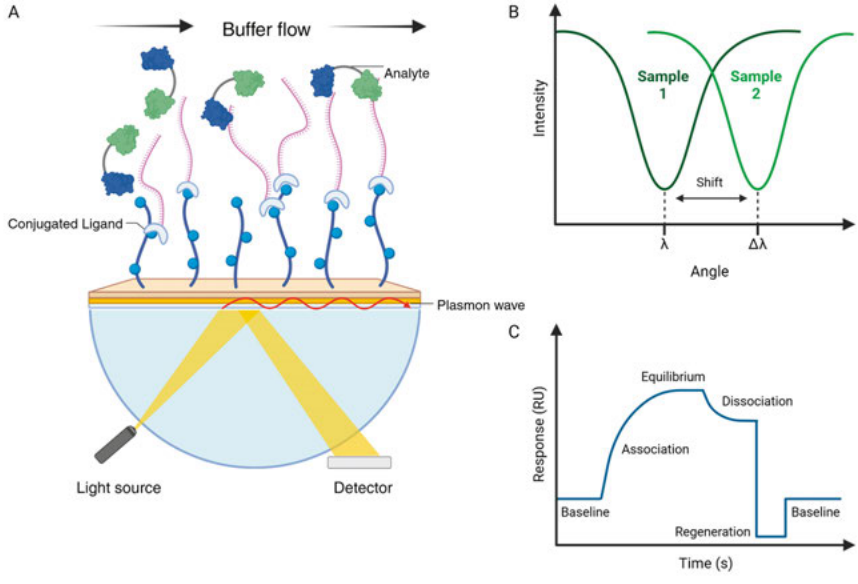


Figure 7. A) Schematic representation of a prism-based SPR sensor in Kretschmann configuration with immobilized target molecule. B) Graphic representing the change in the angle of the incident light absorbed by the plasmons upon concentration change C) Different phases of a sensorgram, including a regeneration step to force ligand-target dissociation. Created in BioRender. Perez, G. (2025) https://BioRender.com/e43r342

SPR biosensors have been widely used to characterize RNA-protein interactions, usually having the RNA immobilized on the sensor surface (as a ligand) and the protein as an analyte in the buffer[106–108].

There are several manufacturers of SPR based biosensors. In this thesis we used Biacore™ 3000 and Biacore™ T200 devices to characterize the interactions. The sensor chips were made of gold and coated with a dextran layer where streptavidin was immobilized first, therefore allowing the immobilization of biotinylated RNA.

Nuclear Magnetic Resonance (NMR) spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy can be used to study biophysical interactions between molecules in solution at a residue level detail and without the need of additional modification of the studied species (*e.g* as immobilization or tagging). The technique takes advantage of the spin of a nucleus. This spin, arising from the intrinsic angular momentum of protons and neutrons, can adopt either values equal to 0 (when $N_{protons}=N_{neutrons}$) or half-integer/integer numbers, creating a magnetic moment. The spins of these atoms align when exposed to a strong external magnetic field in different states depending on their magnetic quantum number. For spins equal to ½ (as ^{1}H, ^{13}C or ^{15}N), the particles can align with or against the magnetic field creating a net magnetization which are perturbated with a radiofrequency (RF) magnetic pulse, changing their orientation out of the equilibrium. After the RF pulse the spins relax to their equilibrium state, creating a signal that can be detected and depends on the nucleus type and its chemical environment. The changes in the chemical environment of the molecules (*e.g* due to different conditions or interactions) will cause shifts in the detected signals that can be tracked, allowing for monitoring of binding events. The shifts are detected by comparing the spectra of the free species with the spectra of the complex[109]

For this study, different NMR experiments (from 1D to 3D) were used depending on the information that was required in combination with properly labelled proteins (^{15}N and/or ^{13}C). Mono-dimensional ^{1}H-NMR experiments were used to get fast and easy overall information on the state of the protein, particularly the folding state. Bi-dimensional (2D) ^{15}N-^{1}H HSQC, were used to monitor the amide groups of the protein (one for every amino acid in the protein excepting proline, plus additional amide group in the side chains), while 3D NMR were employed to gain supplementary information on the protein taking advantage of the carbon atoms that are part of the peptide bonds. Together, the information from these experiments allow to gain information structural and dynamical features of the complex.

Real Time Cell Binding Assays (RT-CBA)

Detection and characterization of biological interactions in early research stages is mainly done by using *in vitro* techniques, but more complete assays are needed to address the complexity of biological settings[110]. When characterizing interactions with isolated proteins, there are several limitations in using *in vitro* approaches, including protein production and purification issues[111], the absence of post-translational modifications[112], protein immobilization[113], and stability, as well as the absence of other interaction partners or regulation process that could influence the interaction in an *in vivo* context.

In contrast to *in vitro* assays, cell assays with living cells allow characterization of interaction with an additional layer of information that is closer to the reality of *in vivo* interactions. This is especially important when characterizing potential therapeutic molecules that are meant to exert their effects in living organisms.

In this thesis, a LigandTracer (LT) device has been used to monitor real-time protein-protein interactions in living cells[114], specifically anti-CD44 and anti-CD44v6 antibodies binding to their respective targets on the cell surface. For a standard Real Time-Cell Binding Assay (RT-CBA), cells are seeded in a plastic Petri dish that can be treated for cell culture or have a specific coating to improve cell adherence (as poly-lysine, poly-dopamine or collagen). Once cells are attached to the so-called detection area, which is the spot were signal will be recorded by a fluorescence detector, the dish is place on the LT rotative support with a certain volume of media (Figure 8). A baseline signal is recorded for both a spot with cells and a control spot, which can be empty or occupied by a negative cell line. After the baseline, the first concentration of the fluorescently labelled ligand is added, and the signal increase is recorded until an acceptable amount of curvature is obtained. A second concentration is needed for an accurate kinetic parameter determination[115], although three concentrations experiments are preferred due to the increased amount of information that is obtained, especially when working with molecules that do not interact with a 1:1 stoichiometry. After the concentration series, the medium with the ligand is removed and fresh medium is added, favouring the complex dissociation. The decrease in signal is then recorded (Figure 8).

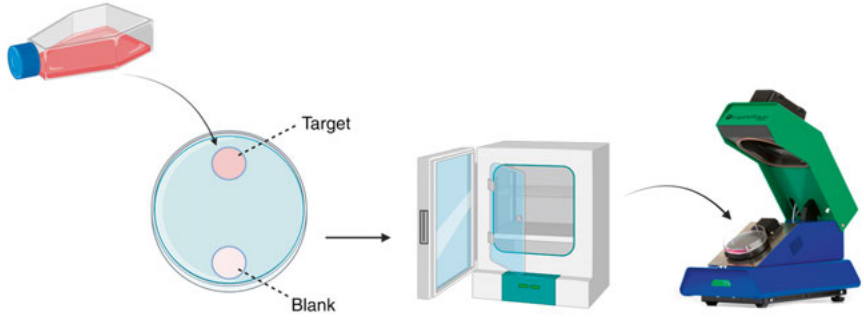


Figure 8. Schematic representation of a standard RT-CBA experimental setup. Created in BioRender. Perez, G. (2025) https://BioRender.com/*w80w510*

The obtained sensorgram-like curves (that resemble a traditional single-cycle kinetic experiment) (Figure 9) are then analysed with a suitable interaction model to be able to quantify the kinetic rate constants and K_D values.

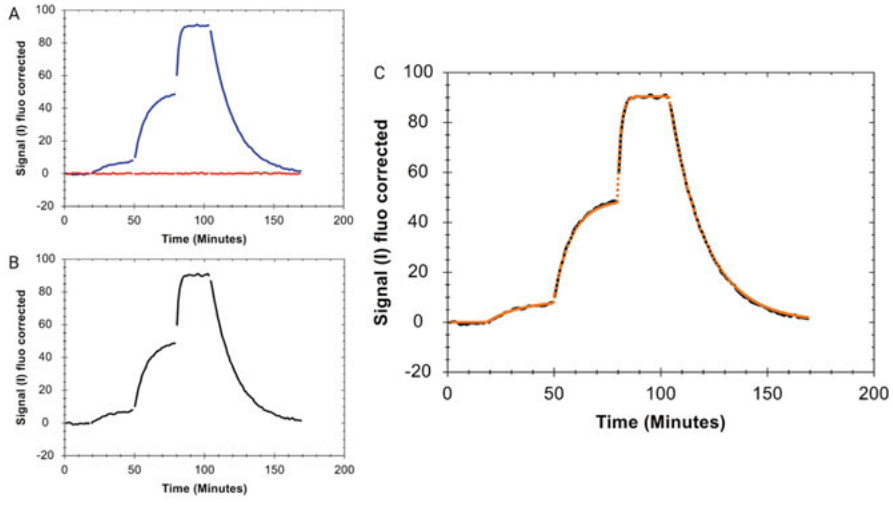


Figure 9. A) Kinetic traces corresponding to the target (blue) and blank (red) B) Blank-subtracted kinetic trace C) Kinetic trace (black) fitted with a 1:1 model (orange dash line). Created in BioRender. Perez, G. (2025) https://BioRender.com/r39j852

Real Time Protein Expression Assays (RT-PEA)

A Real Time Protein Expression Assay (RT-PEA) was developed to monitor real time synthesis of a fluorescent protein in bacteria. For this purpose, a LT protocol was developed to work with bacteria. Although some bacteria-ligand interactions were previously monitored with LT[116], the setting that was used did not allow for the physiological conditions that are needed for an optimal growth and protein synthesis. In this thesis, we developed a new approach using agar coated MultiDishes where we could monitor four independent experimental conditions. Bacteria were seeded in the different compartments that contained agar with different chemical compositions. This allowed us to control protein synthesis, as well as to add compounds that could modulate it. The output of the experiment was sigmoidal-shaped curves that had three different stages, firstly a lag phase until protein synthesis started, secondly a linear production phase and lately a plateau phase, where protein production and degradation reached an equilibrium (Figure 10). A fourth phase, cell death, characterized as an abrupt signal decrease, was not present in the experiments included in this thesis.

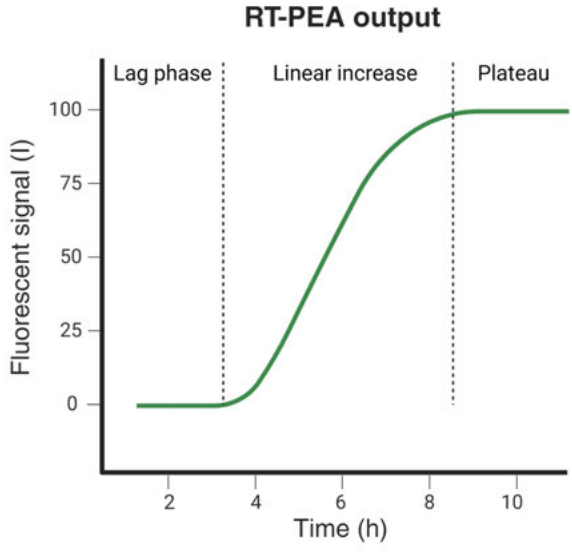


Figure 10. RT-PEA output, indicating the extension of the lag phase, the linear increase and the plateau. Created in BioRender. Perez, G. (2025) https://BioRender.com/u34n909

Interaction kinetic models

To be able to quantify the kinetic parameters and calculate the affinities that characterize an interaction, a model that reflects the interaction mechanism is chosen to analyse the data through non-linear regression analysis[117]. The simplest of all interaction mechanisms, that is also the base for most of the models, is the so-called 1:1 model.

1:1 model and mass transport correction

The 1:1 model assumes that each ligand binds one target molecule reversibly in a single step as illustrated in Scheme 2 and is derived from the Langmuir model describing the adsorption of an ideal gas to a surface[118].

$$[Ligand] + [Target] \underset{k_d}{\overset{k_a}{\rightleftharpoons}} [Ligand{:}Target] \qquad \text{(Scheme 2)}$$

The interaction is described by a single set of kinetic rate constant; *i.e.*, the association rate constant (k_a) and the dissociation rate constant (k_d), from which a dissociation constant at equilibrium (K_D) is calculated. The classical output of a 1:1 interaction is reflected in Figure 11. From the different part of

the curves, different information about the kinetics of the interactions is obtained. The initial part of the association phase, which is almost linear, reflects the association of the first ligands and results in a complex from which dissociation is insignificant. The signal in this phase is mainly influenced by the association rate constant and the concentration of ligand. After this initial segment, curvature starts to appear, with this part of the curve containing information on both rate constants. In some curves, a steady-state phase can be reached where the formation and dissociation of complexes happens at the same speed, therefore not resulting in any change in the signal. Upon dissociation, a signal reduction occurs as the formed complex start to fall apart, and this segment of the curve contains information about the dissociation rate constant only. As dissociation is proportional to the number of bound ligands, the dissociation phase has the shape of a single exponential decay and is independent of the concentration used during the association.

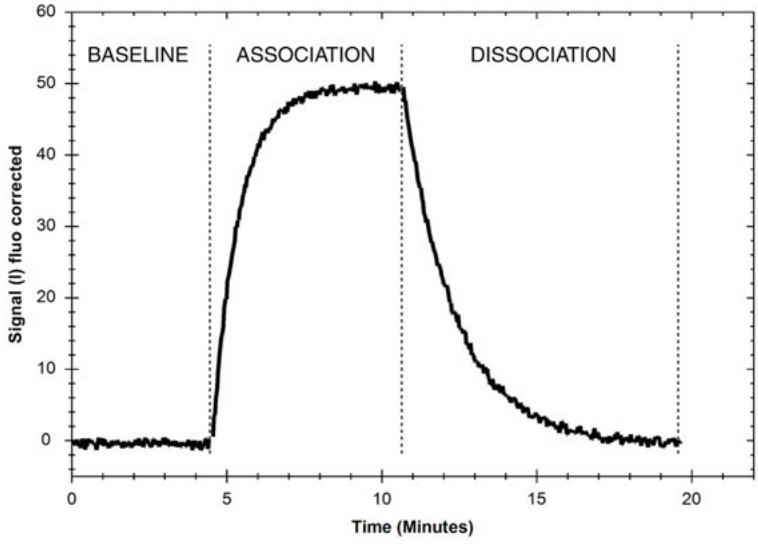


Figure 11. Kinetic trace of a 1:1 interaction. Baseline corresponds to the signal when no complexes are formed. Association starts with a linear phase were first complex are formed, and it is followed by curvature formation when complexes are formed and dissociate simultaneously. The last part of the association is characterized by an equilibrium where the rate of complex formation is equal to the dissociation rate. Finally, in the dissociation phase there is no free ligand and complex dissociate over time. Created in BioRender. Perez, G. (2025) https://BioRender.com/y88x274

The 1:1 mass transport corrected model differs from the basic model on addition of an initial step that accounts for mass transport of ligands to the target. This phenomenon occurs when, in a flow system (as a SPR-based biosensor), the ligand-target association occurs so rapidly that it creates a concentration

gradient between the bulk ligand concentration and the surface where the target is immobilized[119,120]. This artifact can be spotted as a biphasic association in theoretically 1:1 interactions together with a non-stable plateau, that presents a steady increase in signal[121]. To quantify the effect of mass transport, the model includes a mass transfer coefficient[122] (Scheme 3). If the value for the mass transport coefficient is closer to the association rate constant value, the data may be affected by mass transport. Some experimental strategies can be applied to avoid or minimize this effect, for example increasing the flow rate or reducing the concentration of target immobilized on the surface.

$$[Ligand_{bulk}] \underset{k_m}{\overset{k_m}{\rightleftarrows}} [Target_{surface}] + [Ligand] \underset{k_d}{\overset{k_a}{\rightleftarrows}} [Ligand{:}Target] \quad \text{(Scheme 3)}$$

1:1 ligand depletion model

The 1:1 ligand depletion model, as recently described in Weber et al.[123] accounts for a depletion phenomenon than can occur when an assay is performed under conditions where the number of ligands is close to the total number of targets in the system. Such an assay can, in some cases, be designed for closed (non-flow) systems such as RT-CBA. In such an assay design, a significant amount of ligand gets bound to the targets, leading to a reduction in the ligand concentration during an incubation. This reduction in the ligand concentration will lead to a slowing down of the binding rate and thus more curvature during an incubation, as shown in Figure 12. When such a condition is met, the number of targets in the assay can be quantified, as well as the number of targets per cell, in case a cell counting can be performed after the experiment.

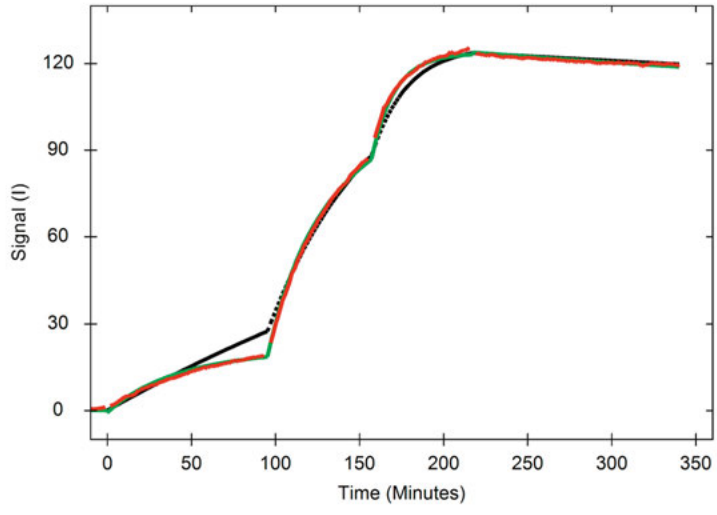


Figure 12. Experimental trace of ligand interacting with a cell receptor under ligand depletion conditions (red), with a 1:1 fit (black) and a 1:1 depletion fit (green). In the first incubation, with a low ligand concentration, the experimental signal start to deviate from the 1:1 model after one hour due to depletion of trastuzumab from solution. This deviation can be used to quantify the total number of receptors in the assay.

Standard RT-CBA assays are designed to not be affected by the depletion, but depletion can be induced by maximizing the number of target (seeding cells on a relatively large area of the Petri dish) and incubating with low ligand concentrations, typically below 1 nM. For an interaction to occur at low ligand concentrations, the affinity should be high, with sub-nanomolar K_D-values.

Ligand induced conformational changes (1:1 two states)

The induced fit model, or 1:1 two states model assumes a sequential binding process where there is a single ligand interacting with a target, after which the formed complex undergoes a change (with rate constant k_{a2}) that leads to a stable state from which it cannot dissociate unless it changes back to its less stable state (with rate constant k_{d2})[124,125]. This model is often employed to characterize interactions that come along with a conformational change, as described in Scheme 4.

(Scheme 4)

$$[Ligand] + [Target] \underset{k_{d1}}{\overset{k_{a1}}{\rightleftarrows}} [Ligand - Target] \underset{k_{d2}}{\overset{k_{a2}}{\rightleftarrows}} [Ligand - Target']$$

The 1:1 two states model results in two sets of kinetic rate constant that can be used to calculate the affinity of the initial interaction (k_{d1}/k_{a1}) and the affinity of the stabilization process (k_{d2}/k_{a2}) that represents the ratio between the

concentration of the initial state of the complex and the stabilized state. Often, the product of both affinity values is presented as the overall affinity. To study an induced fit mechanism, experiments with variation in the duration of association phase are typically employed, as they result in differences in the dissociation rate that do not occur for 1:1 interactions.

Heterogeneous ligand model (1:2)

The heterogeneous ligand model, or 1:2 model, assumes the existence of two parallel 1:1 interactions, where one ligand can bind to two different targets (Scheme 5). Therefore, this model will result in two sets of kinetic rate constants and two different B_{max} values. Two interactions can be quantified if at least one of their rate constants differs sufficiently. In this way, it is possible to decompose the obtained signals into two interactions *i.e.*, one slow and one fast dissociation interaction (Figure 13). Additionally, by calculating the different B_{max} values, the contribution of each interaction to the overall binding process can be determined. As a rule, an interaction should contribute to more than 10% of the overall interaction pattern if meaningful and reliable information on the interaction is to be extracted.

(Scheme 5)

$$[Ligand] + [Target\ A] \underset{k_{d1}}{\overset{k_{a1}}{\rightleftarrows}} [Ligand{:}Target\ A]$$

$$[Ligand] + [Target\ B] \underset{k_{d2}}{\overset{k_{a2}}{\rightleftarrows}} [Ligand{:}Target\ B]$$

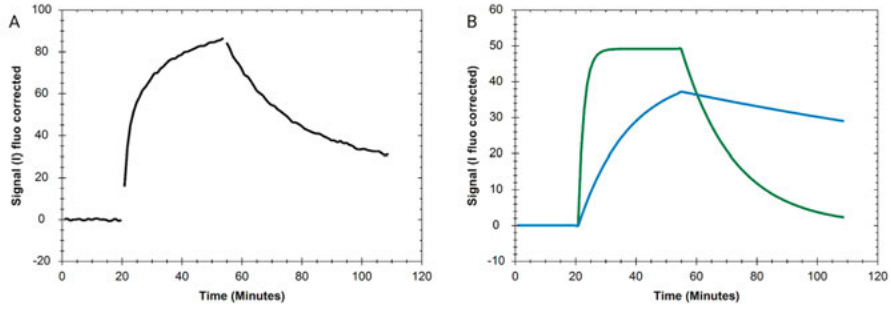


Figure 13. A) Kinetic trace for a heterogeneous ligand interaction. B) Deconvoluted signals of the interaction where the fast-on/fast-off interaction (green) and the slow-on/slow-off interaction (blue) are separated. Created in BioRender. Perez, G. (2025) https://BioRender.com/w49t654

In contrast to 1:1 interactions, heterogeneous interactions present a biphasic behaviour, that can occur during the association and/or dissociation phase. As

the 1:1 interaction has a monophasic exponential decay in the dissociation phase, the appearance of biphasic dissociation is easily recognized. 1:2 interactions are typically studied in more detail by examining interactions at different ligand concentrations, as strong interactions are more apparent at low concentrations and *vice versa*.

Bivalent model (2:1)

The bivalent or 2:1 model assumes that a ligand has two binding sites to a target. It describes a sequential binding process that is composed of two different steps. First, one of the binding sites interacts with the first target, and afterwards, once in proximity, the second binding site binds a second target. This dependence does not mean that it is a cooperative process, *i.e.*, a process in which the first binding event leads to a conformational change that increases the affinity for the second binding step. The model (Scheme 6) quantifies two sets of kinetic constants and two affinity values, one for the monovalent process (the first binding event) and a second one for the bivalent process (the second binding event). However, the second set of rate constants and affinity values is difficult to interpret, as the concentration of the monovalently bound ligand is only known in relative quantities (signal) and not in concentration (molar). Only when the signal can be translated into concentrations, meaningful values for the association rate and affinity of the second, bivalent interaction, can be obtained.

(Scheme 6)

$$[Ligand\ +\ Target] \underset{k_{d1}}{\overset{k_{a1}}{\rightleftharpoons}} [Ligand^{Target}] + [Target'] \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} [Ligand^{Target}_{Target'}]$$

The second affinity values refer exclusively to the second, bivalent, binding event, and does not represent the stabilization due to avidity, or the overall affinity of the process. So far there is not a consensual way of extracting the affinity corrected avidity for the whole process[126].

When using the bivalent model, it is important to realize that most of the signal change comes from the first interaction event, especially in bivalent interactions with a fast recognition, while the second binding event only affects how the overall net binding evolves over time. This implies that the actual information that the curves contain on the second interaction can be minimal and therefore make the fitted parameter unreliable[127]. This interaction is also dependent on the target availability. At lower ligand concentrations, the formation of bivalent complexes is favoured, while at higher ligand concentrations, a competition for the available targets can be observed between ligands, favouring an initial formation of monovalent bound complexes. In the specific case of slow binders at high concentrations, the bivalent mode of binding can

be spotted as a decrease of the signal once the interaction gets closer to equilibrium. This signal change (Figure 14) can be explained as monovalent bound ligands being expelled by other ligands that are binding bivalently.

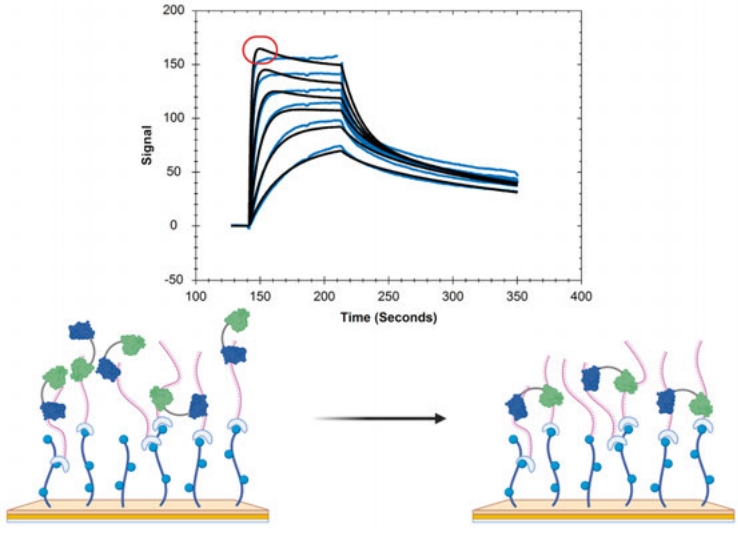


Figure 14. Kinetic traces (blue lines) fitted with a bivalent model (black lines). Highlighted with a red circle appears a reduction in the fit signal that is explained by the scheme below, where monovalent complexes dissociate allowing to the remaining one to form bivalent interactions. Created in BioRender. Perez, G. (2025) https://BioRender.com/u25h512

In general, despite the vast number of molecules with a theoretical bivalent binding mode, this model is underused due to the complexity of interpreting its results as well as the variability from the fitted parameters, that is higher than of other heterogeneous models, such as the 1:2 model[127].

New approach for bivalent affinity estimation

In this thesis a new method to quantify the K_D value of the second bivalent interaction was developed. The approach was designed to account for the characteristics of the interaction pattern displayed by the tandem MSI1 interaction with RNA, with a fast association and a biphasic dissociation, that contained a fast-dissociating fraction followed by a slow dissociating one (this is described in detail under the Results section of this thesis). An example of such an interaction process is shown in Figure 15. Due to the intrinsic difficulties (mentioned in the previous section) to extract meaningful information on the bivalent interaction by using the bivalent model, a new approach was developed.

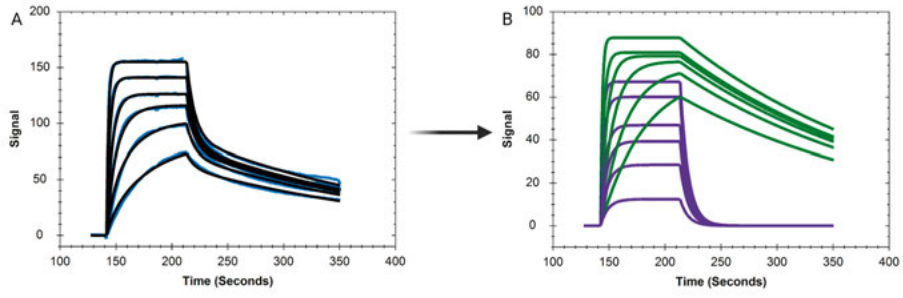


Figure 15. A) Sensorgram (blue lines) fitted with a heterogeneous ligand model (black lines). B) Deconvolution of both interactions, with a fast-dissociating fraction (purple) and a slow dissociating fraction (green). This experiment corresponding to this figure is further described in page 47. Created in BioRender. Perez, G. (2025) https://BioRender.com/f24w028

In this new method, we firstly fitted the data with a 1:2 model. Due to the clearly biphasic dissociation, the 1:2 model resolves both dissociation rate constants accurately, while it was assumed that both association rate constants were similar with no meaningful differences between RRMs, based on the kinetic information that was obtained from experiments with single RRM1 and RRM2 domains. Due to the fast association of the RRM domains, bivalent complexes are readily formed and only at higher concentrations, monovalently bound MSI1 appears due to the lack of free RNA motifs on the sensor chip, and this monovalently bound fraction dissociates rapidly (within a few seconds) as given by Eq. 4. Once enough free RNA becomes available, MSI1 primarily binds in a bivalent manner in a pseudo-equilibrium state where the slow dissociation observed represents the small fraction of monovalently bound MSI1 that dissociates with the same rate as monovalently bound MSI1 rather than rebinding into a bivalent state (Eq.6). The ratio between slow and fast dissociation rate constants thus describes the fraction of monovalent complexes at a certain RNA concentration on the sensor chip. The affinity of the bivalent interaction is defined as the ratio between the concentrations of the interacting molecules and the complex (Eq.7). The ratio between monovalent and bivalent concentrations can be obtained by rewriting Eq. 6, and if you multiply this ratio with the free RNA concentration, the affinity of the bivalent interaction can be calculated.

$$\frac{\delta B}{\delta t} \approx [MSI_{RRM2}^{RRM1:RNA}] * k_{d1} \qquad \text{(Eq. 4)}$$

$$B \approx \left[MSI_{RRM2}^{RRM1:RNA}\right] + \left[MSI_{RRM2:RNA}^{RRM1:RNA}\right] \qquad \text{(Eq. 5)}$$

$$k_{obs} = \frac{\delta B/B}{\delta t} = \frac{\left[MSI_{RRM2}^{RRM1:RNA}\right]}{\left[MSI_{RRM2}^{RRM1:RNA}\right]+\left[MSI_{RRM2:RNA}^{RRM1:RNA}\right]} * k_{d1} \quad \text{(Eq. 6)}$$

$$K_{D2} = \frac{[MSI_{RRM2}^{RRM1:RNA}][RNA]}{[MSI_{RRM2:RNA}^{RRM1:RNA}]} = \left(\frac{k_{d1}}{k_{d2}} - 1\right)^{-1} * [RNA] \quad \text{(Eq. 7)}$$

The RNA concentration at a certain time point during the slow dissociation phase can be estimated from the signal by converting signals into concentrations using the knowledge that 1 RU corresponds to 1 pg/mm^2 [105] and that the thickness of the hydrophilic layer to which RNA is immobilized is 200 nm. With a molecular weight of 39 kDa for MSI1, it means that one signal unit corresponded to 0.22 µM. The signals corresponding to binding of all RNA for the mono (B_{m1}) and bivalent (B_{m2}) fractions are also obtained by fitting the 1:2 model to the curves where both mono- and bivalent binding is clearly visible. The total RNA concentration can thus be calculated by adding the maximum signal for monovalent MSI1 and, as bivalent MSI1 is bound to two RNA strands, two times the maximum signal of the bivalent interaction and multiplying the sum of these signals by the conversion factor (Eq. 8).

$$[RNA]_{total} = 0.22\ \mu M * (B_{m1} + 2B_{m2}) \quad \text{(Eq.8)}$$

Upon dissociation, the free RNA concentration increases over time. The average signal from 100s to 130s after the start of dissociation was used to estimate the bound RNA concentration, as this time frame aligns with the slow dissociation interaction from the 1:2 model. The total signal was divided into mono- and bivalent fractions using the ratio in Eq. 6 and subtracted from the max signals for both fractions as presented in Eq. 8, resulting in Eq. 9.

$$[RNA] = 0.22\mu M * ((B_{m1} - B_{obs}(MR)) + 2(B_{m2} - B_{obs}(MRR) \quad \text{(Eq.9)}$$

In this way, the affinity of the second bivalent process can be estimated in a more robust way that employing the traditional bivalent model.

Interaction Map

Interaction Map (IM) is an analytical tool based on the fitting of many parallel 1:1 interactions to experimental data, quantifying their different rate constants and contribution to the overall binding process. It is displayed as a k_a-k_d plot where each interaction is plotted as a heat map that indicates the interaction contribution (Figure 16)[128].

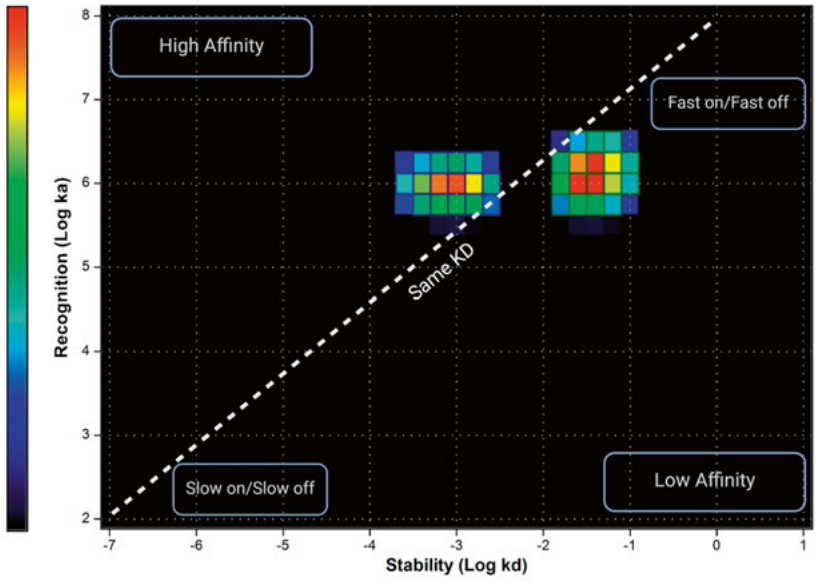


Figure 16. InteractionMap representation indicating the spatial localization of different affinities and kinetic profiles. The colour scale represents the contribution of each interaction.

Growth models

Gompertz model

Growth models are used to characterize biological processes that produce sigmoidal shaped curves. The Gompertz model was developed by Benjamin Gompertz in 1825 as a model that described the increase of mortality as a function of age[129]. In the early 20th century, it began to be used as a model to describe biological growth processes[130]. Since then, the Gompertz has been used for a wide range of applications, from tumour growth[131], bacteria growth[132] or metabolic rates[133] in between others.

There are variations to the Gompertz model, where different parameters are added to fine-tune the model for a certain purpose. In this thesis a traditional three-parameter type I Gompertz model (Eq. 10)[134] was used to characterize the real-time protein synthesis data from a reporter system engineered to detect the influence of MSI1 binding to different RNA strands.

$$Signal\,(t) = A\,e^{-e^{-k(t-T)}} \qquad \text{(Eq.10)}$$

Of the three-parameters, two parameters affect the shape of the curve while the third parameter only influences the location of the curve with respect to the time axis, reflecting the delay in growth. The curve shape affecting parameters are the upper asymptote (A) which describes the maximum signal at the plateau of the curve, and the growth-rate coefficient (k) that describes the maximum growth rate relative to the signal at the plateau. The third parameter is the time of inflection (T) and is situated in around 37% (1/e) of the signal height related to the upper asymptote value.

Interaction kinetic data evaluation

The quantification of the kinetic parameters and equilibrium dissociation constants (affinities) is performed by applying mathematical models, corresponding to interaction kinetic models, that theoretically represent the underlying biochemical interaction mechanisms. However, it can be tempting to choose a model based on the best fit instead of putting the focus on the model that best describes the interaction mechanism[127]. A detailed knowledge of the interaction mechanism is therefore desirable before choosing a model, however this is not always possible, as interaction traces are used to explore and understand the interaction itself.

A first step in data evaluation is visual inspection of the curves and, identification of deviations from a 1:1 model, that for example indicate a heterogeneous interaction, a more complex interaction mode, or to the presence of experimental artifacts[135]. It is key to assure that baseline signals are stable without a significant amount of drift[136], and that there is a logical concentration dependency on the curve signals. Once this is checked, it is critical to assure that the curves included in the evaluation have enough curvature and enough signal decrease upon dissociation[136,137]. If all these parameters are fulfilled, it is possible to perform a reliable model fitting.

At this stage, it is important to remember that the more complex an interaction model is, the more flexibility it has to adjust the fitting due to the number of parameters. For example, a 1:1 interaction can be fitted with a 1:2 model and obtain an excellent fit. However, this does not contribute to understanding the biology behind the data and can even be misleading. The more complex the models get, as frequently observed when using the bivalent and 1:1 two states models, the more diverse the fitting landscape will be[127]. This results in the presence of multiple local minima that can produce a reasonable fit but not necessarily the best fit. Additionally, the outcome may depend on the initial values assigned to various parameters during the fitting process[127]. For these situations, it is important to perform a so-called sensitivity test and assess the presence of local minima and/or flat fitting landscape by systematically

varying the starting values for each parameter and observe the variation in resulting values for the extracted parameters.

Most of the commercial software offer two quality control values to assess the quality of the fitting. The first one is the χ^2 value, which is a numerical value that points out the difference between the experimental trace and the fitted line. A higher value indicates a higher deviation of the fitted line; however, this parameter is signal dependent, which hampers its use to compare different experiments without being corrected by the signals beforehand. The second parameter is the U-value, which indicates the variation that a fitted parameter can suffer without affecting the fitted trace. This gives an understanding of the uniqueness of the data, with high values indicating that a parameter or parameter pair can change a lot without affecting the fitted line, therefore not being unique and trustworthy.

Results and Discussion

Biophysical characterization of MSI1 interactions with RNA

The interactions between MSI1 and RNA were first discovered in *Drosophila*, where it is involved in sensory organ development[46]. Later studies investigated the interactions of the isolated RRM domains using NMR[31] and fluorescence polarization[34]. However, until now, no study has focused on the interaction kinetics of both the isolated RRM domains and the tandem MSI1 protein against the same array of RNA strands. For the results presented in this thesis, a truncated form of MSI1, containing both RRMs in tandem has been used. It lacks the disordered tail, that has been described to also have an influence the interaction process but makes biophysical studies more challenging[35].

Different names for the same RNA sequences are used in **Papers I** and **II**. To simplify comparisons of experiments in the papers and to better compare their results, the name equivalents are summarized in the table below.

Table 1. RNA nomenclature in the different manuscripts and sequences.

RNA name			Sequence
Herein	**Manuscript 1**	**Manuscript 2**	
RNA-L1	RNA-L1	Oligo-L2	5′-UU**GUUAGU**UACCCCUU-3′
RNA-L1a	RNA-L1a	Oligo-L2.1	5′-UU**GUCAGU**UACCCCUU -3′
RNA-L2	-	Oligo-L3	5′-UU**GUUAGU**UAU**UAG**UU-3’
RNA-L2a	-	Oligo-L3.2	5′-UU**GAUAGU**UAG**CAG**GU-3’
RNA-HP2a	RNA-HP2a	Oligo-HP4	5′-AAGC**GUUAGU**UAUU**UAG**UCGCUU-3′
RNA-HP2b	RNA-HP2b	Oligo-HP5	5′-CACUCU**GUAGU**AUG**UAG**GGUUUAUUU-3′

RRM1 and RRM2 interactions with RNA

RRM1 and RRM2 are structurally very similar, RRM1 has a tryptophan residue that is not present in RRM2 that allows the establishment of additional hydrophobicity in the interaction with a guanine or adenine. This residue was key to justify the hypothesis that RRM1 and RRM2 recognizes different RNA motifs[34]. Based on this, the RRM1 domain was described to

bind a longer RNA motif, G/AU$_{1-3}$AGU, compared to the triplet UAG that is recognized by RRM2[34].

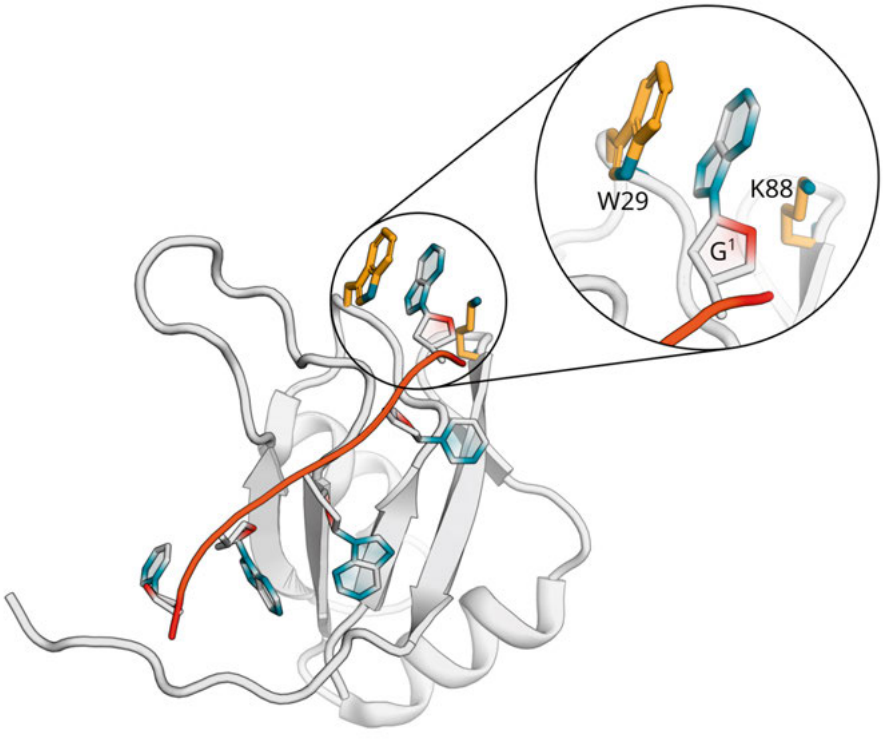


Figure 17. RRM1 residues involved in the interaction with the first guanine of the GAUAGU RNA sequence. Rendered using Pymol from PDB: 2RS2.

RRM1 and RRM2 interactions with linear RNA

In **Paper I**, a comprehensive set of experiments was performed to generate interaction kinetic information on the MSI1-RNA binding. A series of experiments with the single RRM1 and RRM2 domains against linear (RNA-L1 and RNA-L2) and hairpin-like RNA (RNA-HP2a and RNA-HP2b) was performed. RNA-L1 presents a single GUUAGU motif, while RNA-L2 encompasses both a GUUAGU and a UAG motif. These experiments were performed using a SPR-based biosensor where the RNA was immobilized to a streptavidin-coated surface and the protein was the analyte.

Injection of either RRM1 or RRM2 over immobilized RNA-L1 resulted in similar interaction kinetic profiles, with fast association and dissociation rate constants (Table 2), that resemble a classic 1:1 interaction, and a 1:1 diffusion corrected model could be fitted to the data. The estimated affinities for both RRMs were in the double digit nanomolar range, with less than a 2-fold difference. This indicates that RRM1 does not has a specific higher affinity for the

GUUAGU motif than RRM2, and that both motifs are able to compete for the same RNA motif. This is somewhat contradictory with the current literature.

Table 2. Kinetic rate constants and KD values for RRM1 and RRM2 interacting with different RNA strands obtained by fitting a 1:1 mass transport corrected model to the data. Average values are shown, number of replicates and variation are given in Paper I and II.

	RRM1			RRM2		
RNA	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)
RNA-L1	$3.5 \cdot 10^6$	0.167	47	$4.6 \cdot 10^6$	0.112	25
RNA-L2	$4.5 \cdot 10^5$	0.069	153	$4.7 \cdot 10^5$	0.038	82
RNA-HP2a	$8.3 \cdot 10^6$	0.178	21	$1.1 \cdot 10^6$	0.038	37
RNA-HP2b	$8.0 \cdot 10^6$	0.268	33	$2.1 \cdot 10^6$	0.055	26

To further explore the role of the two domains in the interactions with UAG or GUUAGU motifs, the kinetic parameters of RRM1 and RRM2 binding to RNA-L2 (Table 2) were quantified. It was found that RRM1 interaction with RNA-L2 had an almost 8-fold lower association rate constant than its interaction with RNA-L1, while the dissociation rate constant was 2-fold lower. This led to an almost 3-fold weaker affinity. This profile was also observed for RRM2, that had a 10-fold lower association rate constant and almost 4-fold lower dissociation rate constant than the interaction with RNA-L1, leading to a 3-fold weaker affinity. In both cases the interaction traces resembled 1:1 interactions, with no signs of heterogeneity that would be expected if binding UAG or GUUAGU motifs occur with different association and dissociation rate constants. This suggests that both RRMs can recognize and bind to both motifs with similar kinetics and affinities. Therefore, it was concluded that UAG is the minimal recognition motif for both RRMs. The suitability of the 1:1 interaction mode was confirmed by NMR experiments for all RRMs interacting with linear oligoes. For RNA-L2, it was also confirmed that both RRMs can interact with both RNA motifs.

RRM1 and RRM2 interactions with hairpin-like RNA

It has been suggested that the presence of secondary structures, such as hairpins, can influence the interactions between proteins and RNA[138]. In **Paper I** the interactions of the isolated MSI1 RRM domains with two hairpin-like RNA oligonucleotides, RNA-HP2a and RNA-HP2b, were characterized.

In contrast to the RRM interactions with linear RNAs, the RRM1 and RRM2 showed a different kinetic profile when interacting with hairpin-like RNAs. Both RNA-HP2b and HP2a showed a 4- and 8-fold higher k_a when binding RRM1 compared to RRM2 (Table 2). In the case of RRM1, this could be an indication that the recognition is favoured by the presence of a secondary

structure. The dissociation rate constants of RRM1 are in the same range as for the linear RNA, and therefore not affected by the hairpin structure.

Surprisingly, the RRM2 had a slower k_d when interacting with hairpin-like RNAs, indicating that it forms a more stable complex despite not having an association rate constant as high as RRM1. Although the kinetic parameters differed, both RRMs had similar K_D values. For RRM1 the high affinity is primarily a result of a fast association rate while for RRM2 it is mostly due to a slower dissociation rate.

MSI1 interaction with linear RNA

We next assessed the interaction of the MSI1 tandem. The truncated MSI1 protein used in this study comprises both RRMs and a short part of the disordered tail. Firstly, the interaction between MSI1 and the two linear strands RNA-L1 and RNA-L2 was evaluated. RNA-L1 presents a single GUUAGU motif that can interact with both RRM1 and RRM2. Theoretically, and assuming that MSI1 binds only to one RNA strand, we expected to observe a 1:1-like interaction due to the similar kinetic profile of both RRMs when binding RNA-L1. However, the kinetic traces (Figure 15) showed a fast association followed by a clearly biphasic dissociation. This dissociation pattern is characterized by a fast-dissociating fraction followed by a stabilization with a much slower dissociating fraction. The slow dissociating fraction did not match any of the single RRM results; therefore, it must be due to the formation of a different complex. The experimental setup, with flexible RNA strands attached to the surface, together with the presence of a flexible interdomain linker in the protein, suggested that the formation of a bivalent complex with two RNA strands could explain the dissociation pattern. This hypothesis was confirmed by SPR-based biosensor studies using sensor chips with different RNA immobilization levels, as well as by NMR studies, where the formation of intermolecular bivalent complexes was observed. Also, InteractionMap analysis showed two well defined peaks with a similar association and different dissociation rate constants.

To analyse the data, the bivalent model available in the commercial evaluation software was tried first. Theoretically, this model accurately describes the underlying interaction and aligns well with the expected multivalent binding modes of RBPs[100,108]. However, it was challenging to obtain a good fit. The fitting process took longer than expected (>30 min) and it mostly rendered suboptimal fitting traces with high U-values. To improve the fitting process, different starting value combinations were explored. However, the obtained kinetic parameters were influenced by the starting values, which led to the conclusion that the bivalent model could not be used to accurately characterize the interaction.

Therefore, a new approach was used to quantify the bivalent affinity (as previously explained in the Methodology section). For that, a heterogeneous ligand model (1:2) was used to quantify the kinetic parameters of the monovalent interaction, and the affinity of the bivalent interaction. The heterogeneous ligand model rendered lower χ^2 and U-values compared to the bivalent model. The kinetic parameters and K_D values showed a fast-dissociating interaction, identified as the monovalent fraction, and a slow dissociating interaction identified as the bivalent fraction. This bivalent fraction is larger when intermolecular interactions are favoured, *e.g.*, by higher RNA immobilization levels or higher affinities between RRM domains and RNA strands. The new approach takes the variation in immobilization levels into account for the calculation of the bivalent affinity (K_{D2}).

The fast interaction showed similar association and dissociation rates as the ones obtained for RRM1 and RRM2 (Table 3). The monovalent affinity was 58 nM, aligning well with the values obtained for the isolated RRMs.

Table 3. Kinetic parameters and K_D values for the interaction between MSI1 and different RNA strands obtained by applying the new bivalent method for data analysis. Average values are shown, number of replicates and variation are given in Paper I and II.

	MS1			
RNA	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	K_{D2} (μM)
RNA-L1	$8.8 \cdot 10^6$	0.100	58	1.4
RNA-L2	$1.6 \cdot 10^5$	0.041	248	2.8
RNA-HP2a	$1.9 \cdot 10^6$	0.119	77	3.1
RNA-HP2b	$1.9 \cdot 10^6$	0.093	49	3.0

For the bivalent interaction, the K_{D2} of the MSI1 interaction with RNA-L1 was in the low micromolar range. It is important to remember that this second affinity value refers to the bivalent step (Scheme 6) of the interaction and not to the overall avidity-influenced affinity of the interaction. The relative weak affinity is probably due to the lower flexibility or restricted mobility that a tethered molecule has compared with one in solution. It is possible that some steric hindrance impacts the affinity due to the crowded experimental space.

To investigate the differences between intermolecular and intramolecular binding, RNA-L2 interactions with MSI1 were characterized. This RNA harbours two RNA motifs, allowing the formation of intramolecular bivalent complexes. Similar to the interaction with RNA-L1, MSI1 presented a fast association phase followed by a clearly biphasic dissociation. The monovalent interaction presented a 55-fold slower association than MSI1 binding RNA-L1. The dissociation was less than 3-fold slower rate than the interaction with RNA-L1 and was in the same range as the one obtained with the isolated

RRMs. The monovalent interaction had a 4-fold weaker affintiy than with RNA-L1, while the bivalent interaction had a 2-fold weaker affinity. InteractionMap analysis of this interaction showed broader peaks that in the interaction with RNA-L1, suggesting a heterogeneity in the recognition of the RNA motifs that could be explained by the formation of intramolecular complexes, where the binding of the first RRM domain slows down the binding of a second domain. Also, NMR spectroscopy suggested the presence of a heterogeneity of species indicative of the formation of both intra- and intermolecular complexes.

MSI1 interaction with hairpin-like RNA

MSI1 interactions were characterized against two hairpin-forming RNA strands, with two binding motifs each, theoretically allowing for an intramolecular binding mode (Figure 18). Monovalent binding of MSI1 to both hairpins was characterized by an association rate constant that was 10-fold higher than the interaction with RNA-L2 but lower than binding to RNA-L1 while the dissociation rate constant is very similar to that of RNA-L1 (Table 3). The bivalent interaction had an affinity in the same range as the linear RNA, with low micromolar values.

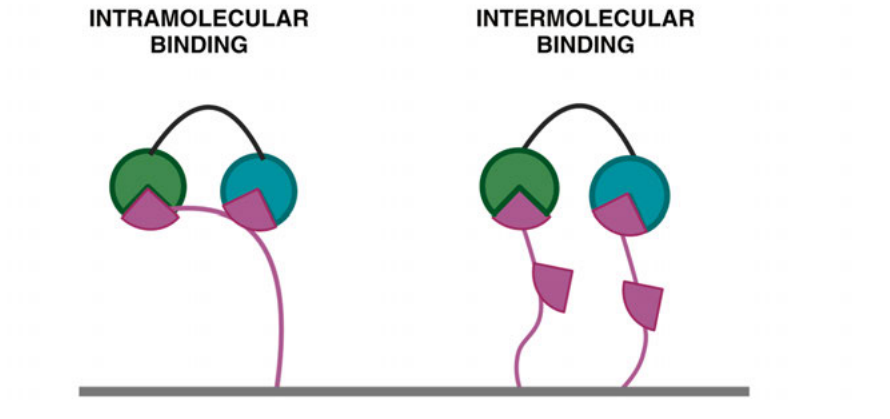


Figure 18. Schematic representation of intra- and intermolecular binding modes. Created in BioRender. Perez, G. (2025) https://BioRender.com/p78i422

Although intramolecular interactions may occur, no difference were observed in the traces, kinetic analysis or InteractionMap analysis that indicated a different binding mode. However, the NMR results showed the presence of different species in solution, leading to a heterogeneous spectrum that could be influenced by the opening of the hairpin structures.

Insights into MSI1 recognition of RNA motifs

The characterization of interactions between MSI1 and GUUAGU and UAG RNA motifs, along with the development of the new method to quantify affinities of bivalent interactions, enabled the design of a new set of experiments. In **Paper II**, the focus was on deciphering the possible competition between RRM1 and RRM2 as well as the design of protein mutants with different binding specificities. For this part of the thesis, a software developed in the RNAct ITN, RRMscorer, was used. This software predicts the likelihood of interactions of pentanucleotides and isolated RRM domains[25]. For the specific case of MSI1, it was first confirmed that both RRMs have a good score against the UAG motif, and secondly it was used to find an RNA sequence which had better affinity for RRM2 than RRM1, resulting in the selection of the CAG motif. Both linear RNAs were modified to produce RNA-L1a, where the motif was switched from GUUAGU to GUCAGU and RNA-L2a, where the second UAG motif was switched from UAG to GCAGG.

In addition to increasing selectivity for one of the RRMs based on the RNA sequences, a series of RRM2 and MSI1 mutants were designed to enhance the affinity of RRM2 for the CAG motif. Based on the previously published structural information on RRM2 binding to RNA, residues E180 and K182 were substituted with E180N and K812M to disrupt RNA recognition for the uracil and improve affinity for cytosine in the CAG motif.

General understanding of the interplay between RRMs and RNA motifs and the ability to switch protein recognition from one motif to another, such as UAG to CAG, were the main goals of this part of the thesis. These mechanisms play a crucial role in developing synthetic biology applications, including synthetic pathways and cell-based biosensors.

Characterization of MSI1 interactions with RNA variants

RRMs interaction with RNA-L1a and RNA-L2a

The interactions of the two new designed RNA variants were investigated to assess the impact of the RNA sequence variations on the interaction kinetics. The interactions of both RRMs with RNA-L1a were similar, with affinity values in the nanomolar range (Table 4). RRM2 presented a slightly higher affinity, mainly due to a slightly lower dissociation rate constant and thus a more stable complex. Compared to the parent strand, RNA-L1, RRM1 had a 3-fold weaker affinity while RRM2 presented an almost 5-fold weaker affinity, mainly due to lower association rate constants.

Table 4. Kinetic rate constants and K_D values for RRM1 and RRM2 binding to different RNA strands obtained after fitting the 1:1 mass transport corrected model to the data. Average values are shown, number of replicates and variation are given in Paper I and II.

	RRM1			RRM2		
RNA	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)
RNA-L1a	$2.6 \cdot 10^6$	0.370	151	$1.8 \cdot 10^6$	0.218	123
RNA-L2a	$1.4 \cdot 10^6$	0.120	86	$2.8 \cdot 10^6$	0.121	45

Compared to the parent strand, the interactions of both RRMs with RNA-L2a become faster both in terms of association and dissociation. RRM1 showed a 3-fold faster association and almost 2-fold faster dissociation rate, while RRM2 showed a 6-fold faster association and a 3-fold faster dissociation rate. This led to affinity values 2-fold stronger for both RRMs, compared with interactions with RNA-L2.

MSI1 interactions with RNA-L1a and RNA-L2a

When interacting with RNA-L1a, the tandem MSI1 showed a different interacting profile compared with the parent strand, RNA-L1. The association was fast, but the stabilized fraction that characterized the bivalent binding was strongly reduced (**Paper II**, Figure 7B). The limited contribution of the bivalent interaction led, in this case, to a characterization of the monovalent interaction only. Compared to RNA-L1, the monovalent interaction with RNA-L1a showed a slower association and faster dissociation, leading to an almost 3-fold weaker affinity (Table 5).

Table 5. Kinetic parameters and K_D values for the interaction between MSI1 and different RNA strands obtained applying the new bivalent method for data analysis. Average values are shown, number of replicates and variation are given in Paper I and II.

	MS1			
RNA	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	K_{D2} (μM)
RNA-L1a	$1.0 \cdot 10^6$	0.201	218	-
RNA-L2a	$3.2 \cdot 10^5$	0.045	144	1.2

The interaction of MSI1 with RNA-L2a still showed a biphasic dissociation phase, therefore both the monovalent and bivalent interactions could be quantified. The monovalent interaction showed a 2-fold faster association and a similar dissociation pattern, this produced a 2-fold stronger affinity. The bivalent affinity also became 2-fold stronger compared with the parent strand, RNA-L2.

The disruption of the bivalency indicates that the new CAG motif successfully influences MSI1 binding, which is backed by the higher dissociation rate constants observed for the isolated RRMs. The fact that the bivalency is disrupted when binding RNA-L1a but not when binding RNA-L2a suggests that intermolecular species are formed between UAG motifs in different strands, while that process is not possible with RNA-L1a strands.

Characterization of RRM2 and MSI1 mutant interactions with RNA

In total, three different protein mutants were chosen to be characterized. RRM2-M1, harbouring the E180N substitution, RRM2-M2, with the K182M, and MSI-1 DM, with RRM2 harbouring both E180N and K182M substitutions.

The interactions between RRM2 mutants and linear and hairpin-like RNA

RRM2-M2 and RRM2-M1 interactions with all linear and hairpin-like RNA were well described by a 1:1 interaction model and the results are summarized in Table 6.

Table 6. Kinetic rate constants and K_D values for RRM2-M1 and RRM2-M2 binding to different RNA strands obtained after fitting the data with a 1:1 mass transport corrected model.

	RRM2-M1			**RRM2-M2**		
RNA	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)
RNA-L1	$8.4 \cdot 10^6$	0.288	36	$9.6 \cdot 10^5$	0.100	108
RNA-L1a	$6.7 \cdot 10^6$	0.389	59	$3.5 \cdot 10^6$	0.227	63
RNA-L2	$7.0 \cdot 10^6$	0.507	75	$1.5 \cdot 10^6$	0.094	65
RNA-L2a	$4.2 \cdot 10^6$	0.187	50	$1.0 \cdot 10^6$	0.187	89
RNA-HP2a	$1.4 \cdot 10^7$	0.178	14	$1.1 \cdot 10^7$	0.177	39
RNA-HP2b	$2.6 \cdot 10^7$	0.375	15	$7.7 \cdot 10^6$	0.112	15

The more relevant interactions for this study were the ones comprising the non-mutated and mutated linear RNAs. RRM2-M1 interaction with RNA-L1a showed a similar binding profile than when interacting with RNA-L1, with an affinity value in the same range as RNA-L1. The interaction with RNA-L2a showed a lower dissociation rate constant, with an almost 3-fold change compared with the interaction with RNA-L2, that lead to an almost 2-fold stronger affinity. RRM2-M2 binding to RNA-L1a had an almost 4-fold faster association than its binding to RNA-L1, while the dissociation was 2-fold faster, leading to a slightly stronger affinity. RRM2-M2 mutation seemed to impact the recognition of the UAG motif, as its k_a was almost 5-fold slower than RRM2

when binding RNA-L1, with no effect on the k_d. Thus, it led to a 4-fold weaker interaction, confirming the effect of the mutations on the RNA recognition (**Paper II**, Figure 8). The isolated analysis of both mutations allows to point out the role of each of the mutations on the interaction processes. RRM2-M1 seems to have a stronger effect on the dissociation rate constant, especially when binding RNA-L2, although surprisingly it also seems to improve the recognition, with an almost 9-fold faster association rate constant. RRM2-M2 seems to improve CAG motif recognition while UAG recognition is impaired, while it does not seem to affect the dissociation rate constant of the complexes.

Compared to non-mutated RRM2, the mutated versions slightly improved recognition when interacting with the hairpins, leading to a stronger affinity of the interaction.

MSI1-DM interactions with linear and hairpin like RNA

The mutated version of MSI1 was characterized against binding of both linear RNAs and their variants, RNA-L1a and RNA-L2a. Interactions between MSI1-Double Mutated (MSI1-DM) and RNA-L1 showed the classical biphasic shape that characterizes the bivalent interaction between MSI1 and RNA, suggesting that bivalency is not affected by the mutation. However, both the monovalent and bivalent interactions resulted in higher K_D-values and thus weaker affinities. The monovalent interaction had an almost 9-fold slower association rate constant that led to a 2-fold weaker affinity, while the bivalent interaction showed an almost 10-fold weaker affinity (Table 7). This indicates a worse recognition, confirming the effect of the mutation, that also affects the bivalent part of the interaction. The interaction between RNA-L1a had the same shape as the interaction with MSI1, with a small, stabilized fraction that barely contributed to the overall interaction. Therefore, the mutation did not improve recognition of CAG enough to recover the bivalent binding. However, the monovalent interaction had a 3-fold stronger affinity due to a 2-fold slower dissociation, indicating an effect of the double mutation in the formation of a more stable monovalent complex.

Table 7. Kinetic parameters and K_D values for the interaction between MSI1-DM and different RNA strands obtained by applying the new bivalent method for data analysis.

	MSI1-DM			
RNA	k_{a1} ($M^{-1}s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	K_{D2} (μM)
RNA-L1	$1.0 \cdot 10^6$	0.100	104	10.6
RNA-L1a	$1.6 \cdot 10^6$	0.094	67	-
RNA-L2	$1.4 \cdot 10^6$	0.140	99	2.3
RNA-L2a	$5.2 \cdot 10^5$	0.101	99	1.9

After evaluation of MSI1-DM binding to linear RNAs containing a single RNA motif, the interaction with linear RNAs harbouring two motifs was studied. Surprisingly, and compared with the interaction of MSI1, the mutated protein showed better recognition (higher k_a) when binding RNA-L2, with an almost 10-fold higher association rate constant, and a 3-fold faster dissociation, that led to a more than 2-fold stronger affinity. The binding to the mutated RNA-L2a also rendered a slightly stronger affinity, although this time due to a slower dissociation rate constant, while the recognition was kept in similar range. These results were not completely surprising, as the RNA still has one UAG motif that can be recognized by the non-modified RRM1.

Characterization of MSI1 interactions with RNA in living cells

The characteristic properties attributed to interactions between different biomolecules can differ depending on the research method that is used, with cell-based techniques giving more biologically relevant information than biochemical or biophysical methods. After detailed *in vitro* characterization of the isolated RRM domains, MSI1 and RNA strands with various mutations, a new method was devised to study the interaction between MSI1 and RNA in living cells (**Paper III)**. This method relies on the indirect characterization of the interaction through a fluorescence reporter system in living bacteria. The engineered system consisted in two different plasmids, one inducible plasmid that coded for truncated MSI1, under the control of a pLac promoter, and a constitutive plasmid, that had an MSI1 RNA binding motif fused to the sequence coding for superfolder green fluorescent protein (sfGFP), that would act as reporter. Upon addition of the lactose analogue Isopropyl ß-D-1-thiogalactopyranoside (IPTG), the MSI1 gene would be expressed, and the subsequently produced protein could interact with a specific RNA motif near the motif regulating the translation of sfGFP. Binding of MSI1 to this motif was expected to impede the translation of the transcript and thereby reduce the fluorescence (Figure 19).

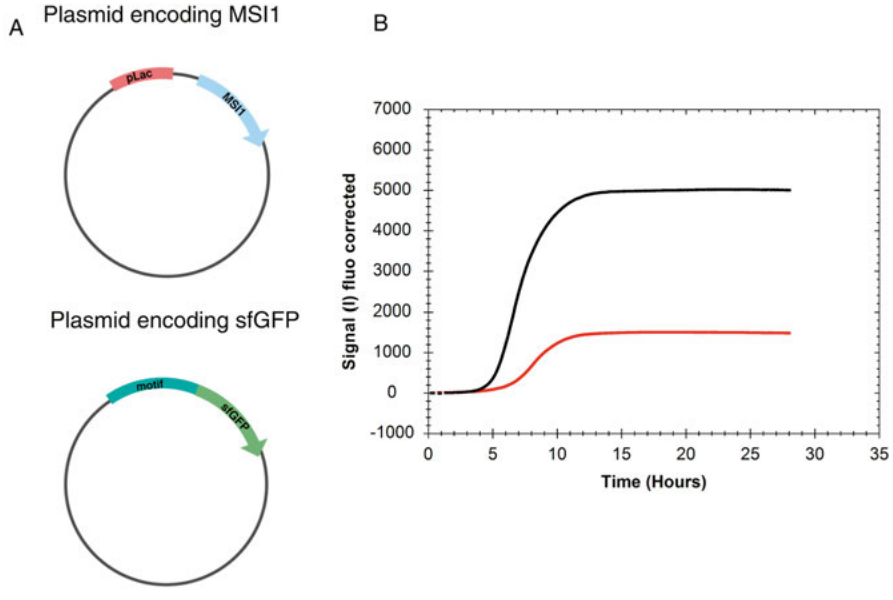


Figure 19. A) Schematic representation of the plasmids forming the post-transcriptional control circuit. B) Standard output of a RT-PEA experiment, IPTG-induced bacteria (red) and non-induced (red). Created in BioRender. Perez, G. (2025) https://BioRender.com/b96b361

MSI1 interaction with RNA in living bacteria

To monitor the production of a fluorescent reporter protein (sfGFP) in living bacteria over time under different conditions, we used a LigandTracer device, traditionally used for RT-CBA, as explained in the Methodology section.

A set of RNA strands, placed individually in a plasmid close to the sfGFP coding sequence, was used to test the binding to the protein. To optimize the method, a primary strand containing two UAG motifs was used. The interaction of MSI1 with this RNA strand was considered as the standard for the comparison of interactions with four different strands (Figure 20); two of these strands (M2 and M4) had substitutions in the UAG core. M2 had one motif where guanine was substituted by cytosine, rendering a UAC motif, while M5 had both UAG motifs turned into UAC motifs. M3 presented a U to C substitution near the first binding motif, and M4 presented a U to G substitution in the region between both UAG motifs, in both cases the binding motifs were unaffected.

Figure 20. Schematic representation of the different RNA motifs that were characterized by RT-PEA. The MSI1 binding motif is highlighted, and the mutated nucleotides are circled in red.

The results showed that both M2 and M5 resulted in a stronger reduction in fluorescence compared to M3 and M4. This suggests less effective inhibition of sfGFP expression, indicating a weaker binding of MSI1 to the RNA segment. M4 presented a minimum reduction in the fluorescence and M3 rendered an intermediate result between M2, M4 and M5, indicating that modifications near the binding core could affected the binding efficiency.

The results were comparable to those from a previous study[139] that employed the same system in suspension instead of solid media, obtaining a similar fluorescence fold-reduction for each interaction pair between MSI1 and RNA strands (**Paper III**, Figure 3). When comparing the correlation between *in vitro* affinity values and fluorescent reduction, the results obtained with the RT-PEA were in line with previously published[139] studies.

Oleic acid as an allosteric modulator of MSI1-RNA interactions

Different concentrations of oleic acid were added to the agar to test the effects on the MSI1-RNA interaction. Based on a series of concentrations, it was observed that a 60 µM OA concentration produced the lowest fluorescence decrease, with an average 1.14-fold change in the fluorescence signal (Figure 21) which implied an almost complete reversion of the binding in the presence of IPTG. This suggests that oleic acid can effectively interfere with the interaction between MSI1 and RNA in physiological conditions, which confirms previously published results describing the influence of oleic acid on MSI1[43,139].

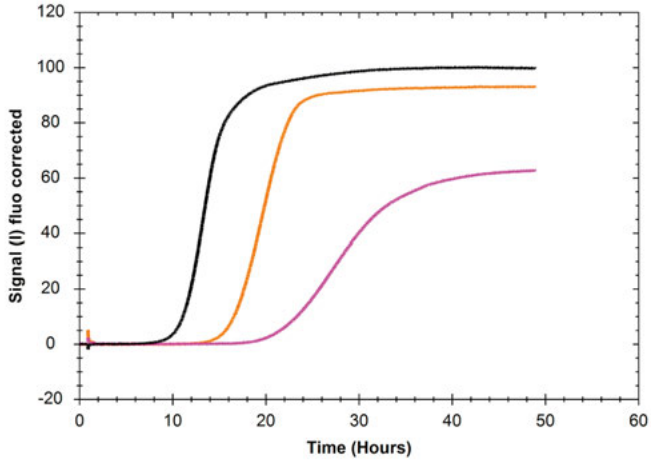


Figure 21. Effect of oleic acid and MSI1 on the synthesis of reporter protein sfGFP, which is regulated by MSI1-RNA interaction. Traces corresponding to the fluorescence signal of transformed bacteria with MSI1 and sfGFP reporter plasmids treated with IPTG (magenta), with IPTG and 60 μM oleic acid (orange) and non-treated (black). Figure from Paper III.

MSI1 regulation of CD44 and CD44v6 receptors

The inhibition of RBPs with RNA in cancer has been devised as an attractive therapeutic option. MSI1 is involved in regulation of various pathways and is expected to regulate expression of cancer marker CD44. In **Paper IV** we quantified the number of CD44 and CD44v6 receptors in a colorectal cancer cell line (HCT-116) that expresses MSI1. The experiments were performed under standard conditions and after treatment with two known MSI1 regulators, oleic acid and luteolin.

The receptor quantification was performed using the recently described method of ligand depletion (explained in detail in Methodology). This method allows receptor quantification and calculation of the kinetic parameters and affinities within the same experimental setting. The number of CD44 and CD44v6 receptors was successfully quantified in HCT-116 cells. Considering that the anti-CD44 antibody is non-isoform specific, while the anti-CD44v6 antibody is specific for the splice variant, we could estimate the proportion of CD44v6 antibodies in the total population to be 77%. As a control, interaction kinetic parameters for the interaction between the antibodies and their targets was performed under standard and depletion conditions, obtaining similar rate constants and affinity values (Table 8).

Table 8. Kinetic parameters, K_D values and number of CD44 and CD44v6receptors per cell quantified using RT-CBA in HCT-116 cells.

Parameters	CD44v6 RT-CBA (n=3)	CD44v6 Receptor # assay (n=7)	CD44 RT-CBA (n=2)	CD44 Receptor # assay (n=2)
k_a (10^5 $M^{-1}s^{-1}$)	3.67 ± 0.90	3.86 ± 2.05	0.76 ± 0.05	1.43 ± 0.26
k_d (10^{-5} s^{-1})	2.94 ± 1.37	2.70 ± 1.62	0.97 ± 0.03	1.36 ± 0.43
K_D (pM)	82 ± 38	116 ± 102	129 ± 9	95 ± 19
Receptors/cell (10^6)	NA	0.22 ± 0.21	NA	0.28 ± 0.08

Effect of oleic acid and luteolin on CD44v6 receptor number

Oleic acid has been shown to inhibit MSI1 binding to RNA and therefore has potential to affect CD44v6 receptor number on the cell surface. Cells were first seeded in a 6-well plate and treated for 48 or 24h with oleic acid or luteolin. Afterwards, cells were detached and seeded in a MultiDish and then a RT-CBA was performed. After a concentration series test, it was found that a concentration 300 μM oleic acid had the major effect on the number of receptors while avoiding the toxic effects observed at higher concentrations. Treatment of HCT-116 cells during 48 h showed the largest reduction in the number of receptors per cell, with a 35% reduction. Also, a reduction in the total number of cells was observed, indicating an effect on cell growth.

Treatment with 20 μM luteolin for 48 h seemed to render the largest reduction in the number of receptors per cell to 20%. Also, the effect on cell growth seemed to be higher upon luteolin treatment, with a reduction of 59% compared to the control groups. This could be due to the wider inhibition range of luteolin, that is known to affect several protein mediated processes.

RRM1 interaction with RNA in the presence of oleic acid

As a control to explore the effect of oleic acid in the RRM1 interaction kinetics with RNA, an SPR biosensor experiment with the RRM1 and two RNA strands was performed. Musashi-1 RRM1 domain was incubated with 0.1 mM oleic acid to explore the effect of the fatty acid in the interaction kinetics profile and affinity. The interaction between RRM1 and a linear RNA strand, RNA-L1, and a hairpin like RNA strand, RNA-HP2b, were evaluated. The results showed an effect on of the association rate constants for both strands, with a 2-fold slower association for RNA-L1 and a 3-fold slower association for RNA-HP2b. Subsequently, the affinities were 2-fold weaker for RNA-L1 and almost 3-fold weaker for RNA-HP2b.

Conclusion

This thesis explores various aspects of the interactions between MSI1 and RNA. The research started by using biophysical techniques, mainly SPR-based biosensors and NMR spectroscopy, to study the affinity and interaction kinetic profile of the isolated RRMs and MSI1 to different RNA strands. After this basic characterization, a new method to analyse the complex data was devised, RNA mutants and different protein versions were designed to modify RRM2 RNA recognition. This was followed by investigating the interaction between Musashi-1 and RNA in living bacteria by using a reporter system. The allosteric effect of oleic acid on the binding was also studied. Finally, the role of MSI1 in the regulation of a cancer surface marker, CD44v6, was analysed by quantifying cell proliferation and CD44v6 expression.

The main findings of these thesis can be summarized as follows:

Paper I and II

- Isolated RRM1 and RRM2 interact similarly with linear and hairpin-like RNA, and the results are well described by a 1:1 model. The interaction is fast, for both association and dissociation, with K_D values in the two-digits nanomolar range. There is no indication that the binding to hairpin-like RNA has a higher affinity, or a different kinetic profile compared to linear RNA.

- The minimum recognition sequence for both RRM1 and RRM2 is UAG. There is no indication that these RRMs have different roles in the recognition and stabilization of the interaction, as was earlier described[32,34].

- A new method developed to analyse MSI1 interactions was suitable for extraction of information from fast, bivalent interactions while considering fluctuations in the RNA concentration on the chip surface. It was also useful for quantifying the bivalent affinity, K_{D2}. The bivalent affinity was in all cases found to be weaker than the monovalent one. Overall, these results were well aligned with previously published affinities as well as with the expected role of multivalent RBPs, *i.e.* in the formation of cellular condensates**.**

- When MSI1 interacts with RNAs with a single motif, the formation of bivalent complexes reflected an intermolecular binding mode, with one protein interacting with two RNA strands. When binding to a linear RNA with two binding motifs, both NMR and SPR-based biosensor experiments suggested that intramolecular binding dominates.

- The analysis of the isolated RRM2 variants, each of them harbouring one substitution, allowed to identify E180N as the change that mostly improved recognition and affinity for the CAG motif.

- RRMScorer was validated as a useful tool for the design of new RRMs with different binding specificities, while protein variants of MSI1 where successfully modified to recognize a different RNA motif**.**

- The different effects that each of the RRM2 substitutions had in the interaction profile highlights the importance of performing time-resolved analysis as a complement of equilibrium experiments as NMR.

Paper III and IV

- A new method developed for the real-time characterization of RNA-protein interactions in bacteria on solid media, using a post-transcriptional controlled circuit, was shown to work as expected. The results were validated using different RNA sequences and were equivalent to data obtained from bacteria in suspension.

- Oleic acid, a known MSI1 allosteric modulator, showed to reduced binding of MSI1 to RNA in the engineered circuit. These results could not be obtained using suspension-based methods, due to the turbidity of the media, and highlights the contributions of this new method to study reporter assays in solid media.

- Oleic acid resulted in a time-dependent reduction in the number of CD44v6, while luteolin mainly affected cell proliferation. Although it is difficult to draw any conclusions on the effect directly on MSI1, due to the possible affection of other pathways, the results suggest a possible MSI1 influence on CD44v6 and opened the door for the development of therapies that could target RNA binding proteins.

The findings in this thesis, alongside with the joint effort from the RNAct ITN, represents a significant step forward in understanding RBP-RNA interactions. This opens the door to new drug discovery screening strategies for RNA-based therapeutics as well as an expanded toolkit for innovative synthetic biology applications.

Future Perspectives

As is common in research, testing our hypothesis led to new questions. Most new questions were related to the specifics of the interaction between MSI1 and RNA.

The role of MSI1 C-terminal disordered domain and the linker between RRM domains in the interaction with RNA.

The truncated version of MSI1 used in this study lacked most of the C-terminal disordered domain. This was a result of the challenges that posed the purification of the full-length protein. Therefore, an important and active segment of the protein for the interaction was missing. Performing interaction experiments with the full-length protein could provide deeper insights of RBP binding mechanisms. This is not only due to the suggested direct involvement of the C-terminal domain interactions with RNA, but also because of its potential role in interacting with other proteins that could also regulate the interaction with RNA. The presence of the C-terminal domain would allow the characterization of the formation of cellular condensates and further explore its impact on multivalent binding mechanism.

Apart from the C-terminal domain, the linker between both RRM domains has been involved in regulating RNA-protein interactions. It has been described as a disordered and flexible structure that allows the protein to interact with spatially separated RNA motifs. As part of the RNAct ITN, there were several secondary aims focused on the linker that could not be fulfilled. First, it was planned to substitute key residues of the linker to create more rigid structures and study the impact on the interaction, especially on the affinity. Second, new binding motifs could have been introduced in the linker to make it interact with small molecules. This would have added an additional regulatory layer that could have been used for the development of synthetic biology applications, such as biosensors, or as a testbed for the development of therapeutic molecules.

Impact of longer spacer sequences in RNA

All RNA sequences used in this study derived from endogenous MSI1 interaction partners, and have a relatively short distance between binding motifs, which could have had an impact on the bivalent interactions. The study of

different separation sequences between RNA motifs would allow to find the maximum distance that the linker allows, as well as the maximum distance where bivalency is kept as a preferred binding mechanism.

Thermodynamic study of the binding

Binding processes are influenced by two major components, enthalpy and entropy. SPR based biosensors allow the study of interaction kinetics at different temperatures and therefore to analyse if the affinity of the binding is driven by the enthalpic component or the entropic component. Depending on which factor is influencing the affinity different strategies can be applied, for example, to the design of protein mutants with affinity for different sequences or to the design of inhibitors with better affinity.

Real-Time study of CD44v6 receptor expression

The fourth manuscript set the base for a more detailed study of the MSI1 regulation of CD44v6 receptors in real-time. This study can be expanded by monitoring receptors and characterizing other pathways. Additionally, modified cell lines could be included, where MSI1 expression is induced by a small molecule, allowing real-time observation of changes in receptor expression patterns.

Popular Summary

In 2020, the word RNA started to gain widespread attention, mainly due to its role as the main component of the COVID-19 vaccines. Although RNA, and its interactions with other biomolecules, have been studied for decades, recent years have seen a surge of interest in this molecule. This renewed fascination is not just explained by the appearance of a new class of RNA-based therapeutics, but also by the diverse and often critical functions, that RNA performs in our cells.

This thesis focuses on the characterization of some of these functions that relate how RNA interacts with other molecules, forming complexes that exert biological effects. In this study, RNA-protein interactions were specifically examined. A specific class of proteins, known as RNA binding proteins, can bind RNA and modify its activity in protein synthesis. These interactions have an impact in a wide range of cellular functions, including cell growth, stress response or cancer. Therefore, it is of great importance to understand the underlying mechanisms of these interactions.

There are many ways to characterize biological interactions. Here, we mainly studied both how complexes of proteins with RNA are formed, and how they break up, or dissociate, over time. This process is known as real-time interaction analysis and reveals the kinetics and affinity, or binding strength, of interacting molecules. For this, we used an RNA binding protein called Musashi-1 (MSI1) and different RNA strands. Some of these strands were able to fold themselves into a rigid shape, such as hairpins (because the shape looks like a hairpin) that can influence the interaction characteristic.

In **Paper I**, we characterized the interaction kinetics and affinity of the MSI1 protein to linear and hairpin-like RNA. Due to the complexity of the data, we developed a new method to estimate one the parameters that defined the interactions, the equilibrium dissociation constant (affinity) of the bivalent interaction, $K_{D2.}$

In **Paper II**, we designed and characterized variations of the MSI1 protein that were able to recognize different RNA than the original protein. To achieve this, we used a new software, **RRMScorer**, that can predict the interactions

between mutated proteins and RNA. The possibility to design new proteins that bind to different sequences opened the door to introducing non-native behaviours in some cells, as well as to better understand the nature of the interactions. Additionally, a structural characterization of the interaction was performed.

In **Paper III**, a new method to study RNA-protein interactions in living bacteria grown on agar was designed. The interaction was detected by changes in the production of a fluorescence protein that was coded on the same RNA strand as the RNA motif that potentially is bound by MSI1. The method proved to be effective and comparable with other approaches that used bacteria in suspension. In contrast to bacteria grown in suspension, this new method was also able to monitor how the interaction was modulated by oleic acid, a compound that showed to inhibit the interaction between MSI1 and RNA.

In **Paper IV**, effects of MSI1 inhibition in the expression of a cancer marker were investigated. The cellular receptor CD44 is present on the cell surface and its abundancy is related with metastasis. MSI1 is known to regulate CD44 expression, and we used different inhibitors, oleic acid and luteolin to assess the impact on CD44 receptor levels. While some variations were observed, further studies are needed to confirm whether these effects are directly caused by MSI1 inhibition.

This thesis represents a significant step forward in understanding the molecular mechanisms governing the interaction between RNA-binding proteins and RNA. This knowledge can be translated into new screening strategies for the development of drugs targeting RNA biology as well as in new tools for synthetic biology applications

Resumen Divulgativo

En 2020 el acrónimo ARN comenzó a popularizarse entre la población debido al papel de esta molécula como uno de los principales componentes en las vacunas del COVID-19. Aunque tanto el ARN como sus interacciones con otras biomoléculas han sido estudiados durante décadas, en los últimos años ha aumentado el interés por todo lo que rodea a esta molécula. Esto no solo se explica por la aparición de una nueva clase de terapias basadas en ARN, o que tienen a esta molécula como su diana, sino también por las diversas y, a menudo fundamentales, funciones que el ARN desempeña en nuestras células.

Esta tesis se centra en la caracterización de la interacción del ARN con otras moléculas, con las que forma complejos que condicionan su actividad biológica. En este estudio, investigamos, en particular, la interacción entre el ARN y las proteínas. Una clase específica de proteínas, conocidas como RNA-binding proteins (o proteínas de unión al ARN), son, como su nombre indica, capaces de interaccionar con el ARN. Estas interacciones regulan una amplia gama de funciones celulares, incluyendo el crecimiento de las células, la respuesta al estrés y el cáncer. Por lo tanto, es de gran importancia comprender los mecanismos subyacentes de estas interacciones.

Existen muchas formas de caracterizar las interacciones biológicas. El foco de este estudio consiste en discernir cómo se forman los complejos de proteínas con ARN y cómo se disocian a lo largo del tiempo, proceso conocido como cinética de interacción (o binding kinetics). Para ello, utilizamos una proteína de unión al ARN llamada Musashi-1 (MSI1) y diferentes secuencias de ARN. Algunas de estas secuencias son capaces de formar estructuras secundarias llamadas horquillas (debido a su parecido con horquillas para el pelo), las cuales pueden influir en las interacciones que se establezcan.

En el **Artículo I**, se caracterizó la cinética de interacción de la proteína con ARN lineal y en forma de horquilla. Debido a la complejidad de los datos, desarrollamos un nuevo método para estimar uno de los parámetros que definen las interacciones: la constante de disociación en equilibrio de la interacción bivalente (afinidad), K_{D2}.

En el **Artículo II**, se diseñaron y caracterizaron diferentes variantes de la proteína MSI1 capaces de reconocer ARN distintos de los que reconocía la proteína original. Para ello, utilizamos un nuevo software, RRMScorer, que puede predecir si habrá o no interacción entre diferentes proteínas y ARN. La posibilidad de diseñar nuevas proteínas que se unan a diferentes secuencias abre la puerta a la introducción de comportamientos exógenos en algunas células, así como a una mejor comprensión de la naturaleza de estas interacciones. Para complementar este estudio, se realizó una caracterización estructural de la interacción usando resonancia magnética nuclear.

En el **Artículo III**, se diseñó un nuevo método para estudiar interacciones ARN-proteína en células bacterianas vivas cultivadas en un soporte sólido. La interacción se detectó mediante cambios en la producción de una proteína fluorescente; si la interacción ocurría con éxito, la expresión de la proteína se inhibía y la fluorescencia disminuía. El método demostró ser eficaz y comparable con otros enfoques que utilizan bacterias en suspensión. La interacción se moduló utilizando ácido oleico, un compuesto que demostró inhibir la interacción entre MSI1 y el ARN.

En el **Artículo IV**, se investigan los efectos de la inhibición de MSI1 en la expresión de un marcador de cáncer. CD44 está presente en la superficie celular y se relaciona con la metástasis. Se sabe que MSI1 regula la expresión de CD44, por lo que utilizamos diferentes inhibidores, como el ácido oleico y la luteolina, para evaluar su impacto en los niveles del receptor CD44. Aunque se observaron algunas variaciones, se necesitan más estudios para confirmar si estos efectos son causados directamente por la inhibición de MSI1.

Esta tesis representa un paso adelante significativo en la comprensión de los mecanismos moleculares que rigen la interacción entre proteínas y RNA. Este conocimiento se puede aplicar en la creación de nuevas estrategias para el desarrollo de terapias que tengan como diana la biología del RNA, así como en nuevas herramientas en el campo de la biología sintética.

Populär sammanfattning

Det var år 2020 som ordet RNA började få stor uppmärksamhet, främst på grund av dess roll som en nyckelkomponent i COVID-19. Även om RNA och dess interaktioner med andra biomolekyler har studerats i årtionden, har de senaste åren sett ett ökat intresse för denna molekyl. Denna förnyade fascination förklaras inte bara av framväxten av en ny klass av RNA-baserade terapier, utan också av de många och ofta kritiska roller som RNA spelar i våra celler.

Denna avhandling fokuserar på att karakterisera några av dessa funktioner på interaktionsnivå. RNA interagerar med, eller binder till, andra molekyler och bildar komplex som utövar biologiska effekter. I denna specifika studie undersökte vi särskilt interaktionen mellan RNA och proteiner. En specifik klass av proteiner, kända som RNA-bindande proteiner, kan binda till RNA och modifiera dess aktivitet. Dessa interaktioner påverkar en rad cellulära funktioner, inklusive celltillväxt, stressrespons och cancer. Därför är det av stor vikt att förstå de underliggande mekanismerna bakom dessa interaktioner.

Det finns många sätt att karakterisera biologiska interaktioner. Här studerade vi huvudsakligen hur proteinkomplex med RNA bildas och hur de dissocierar över tid, en process känd som interaktionskinetik. För detta använde vi ett RNA-bindande protein vid namn Musashi-1 (MSI1) och olika RNA-strängar. Vissa av dessa strängar kunde bilda sekundära strukturer, kallade hårnålar (eftersom de liknar hårnålar), som kan påverka biomolekylära interaktioner.

I **Artikel I** karakteriserade vi interaktionskinetiken mellan proteinet och både linjärt samt hårnålsformat RNA. På grund av datakomplexiteten utvecklade vi en ny metod för att uppskatta en av de parametrar som definierar interaktionerna: den ekvivalenta dissociationskonstanten för den bivalenta interaktionen, K_{D2}.

I **Artikel II** designade och karakteriserade vi olika varianter av MSI1-proteinet som kunde känna igen andra RNA-sekvenser än det ursprungliga proteinet. För att uppnå detta använde vi en ny programvara, RRMScorer, som kan förutsäga om en interaktion mellan olika proteiner och RNA kommer att ske eller inte. Möjligheten att designa nya proteiner som binder till olika sekvenser öppnar dörren för att introducera icke-nativa funktioner i vissa celler samt för

en bättre förståelse av interaktionernas natur. Dessutom genomfördes en strukturell karakterisering av interaktionen.

I **Artikel III** designade vi en ny metod för att studera RNA-proteininteraktioner i levande bakterieceller odlade på ett fast underlag. Interaktionen detekterades genom förändringar i produktionen av ett fluorescerande protein; om interaktionen lyckades hämmandes proteinuttrycket och fluorescensen minskade. Metoden visade sig vara effektiv och jämförbar med andra metoder som använde bakterier i suspension. Interaktionen modulerades genom användning av syran ”oleic acid”, en förening som visade sig hämma interaktionen mellan MSI1 och RNA.

I **Artikel IV** undersökte vi effekterna av MSI1-hämning på uttrycket av en cancermarkör. CD44 finns på cellytan och är kopplat till metastas. MSI1 är känt för att reglera CD44-uttrycket, och vi använde olika hämmare, såsom oleic acid och luteolin, för att försöka bedöma deras påverkan på CD44-receptornivåerna. Även om vissa variationer observerades krävs ytterligare studier för att bekräfta om dessa effekter är direkt orsakade av MSI1-hämning.

Denna avhandling representerar ett viktigt steg framåt för att förstå de molekylära mekanismerna som styr interaktionen mellan RNA-bindande proteiner och RNA. Denna kunskap kan översättas till nya screeningstrategier för utveckling av läkemedel som riktar sig mot RNA-biologi samt i nya verktyg för tillämpningar inom syntetisk biologi.

Acknowledgements

A doctoral thesis can only be understood as a collective work. As such, I would like to thank everybody who contributed to this work in any possible way during these past years.

I would like to thank my supervisors, **Jos**, **Helena**, and **Ylva**. **Jos**, I cannot thank you enough for your constant support and care throughout these years. I can see my progression as a scientist since I arrived at Ridgeview five years ago, and you have been a key influence. I will always remember your patience when explaining to me all the complex concepts of the thesis. **Helena**, I am grateful for all the discussions about science, especially about interaction kinetics. Your advice was invaluable, offering me a broader understanding of how academic research works. To **Ylva**, you were always willing to help and solve any doubts I had.

To all Ridgeview employees who supported this learning journey: **Sina**, **Anna**, **John**, **Kalle**, **Hadis**, **Sara**, **Piero**, **Qinqtiang**, and **Paolo**. You all created a fantastic work environment that contributed to the completion of this thesis. I want to thank especially **João**, you guided me in the beginning and helped me think as a PhD student, and **Hanna**, for your willingness to discuss scientific challenges. **Daniel**, for your patience when clarifying my math doubts and for your guidance through the last stages of this project.

To all members of RNAct, who contributed to this work with enthusiasm, time, and endless scientific discussions: **Anna K.**, **Hrishi**, **Anahí**, **Jose**, **Nikki**, **Stefano**, and **Aitor**. I want to especially thank **Anna P.**, for all the time that we spent together struggling with Musashi, trying to get the projects finished (now we are almost done!). I would also like to thank **Roswitha** for your patience while teaching me the secrets of synthetic biology and for your guidance during my research stay at Rodrigo's group. **Joel**, thank you for solving all my doubts on the computational side and for all the great conversations we had. And finally, **Luca**, for your careful review of the NMR section of this thesis, best of luck on your defense!

To the members of the Kemi-BMC department and the Danielson Lab. Especially to **Gun** and **Mostafa** who were always keen to help me with any doubts

or struggles. **Nadine**, **Aimiliani**, **Max**, thank you for being my main support in the department and for guiding me through all the routines. **Andrei** and **Ioanna**, thank you for revising this text, for your help with the figures, and for all the fikas we had during these years. **Helena** and **Eldar**, thank you for helping me set up the SPR experiments and for sharing your expertise with the different biosensor-based techniques that we have in the department.

To all the Rudbecklaboratoriet people who helped me over the years, especially to **Tabassom** and **Tianqi** for always helping me resolve experimental doubts and cell lab issues, and to **Anzhelika**, who always listened and gave me a fresh perspective.

To all the friends I made during this time in Uppsala and who gave me the daily strength to continue working on this project. Such relationships are the most valuable thing I got during the time of my PhD.

To all my friends, especially in Spain, who listened to me and supported me during these years. Without you this would have been a much more difficult task, if not impossible. Gracias por haber estado siempre ahí y espero que puedamos seguir apoyándonos mutuamente por muchos años.

To my family, especially to my parents **Ángela** y **Emilio**, and my sister **Laura**, thank you for being with me, listening and supporting me independently of how steep the hill was. Esta tesis no existiría sin vosotros y por tanto podéis estar tan orgullosos de ella como yo lo estoy.

References

1. Vane, J. R. & Botting, R. M. The mechanism of action of aspirin. in *Thrombosis Research* **110,** 255–258 (2003).
2. Sheu, S.-Y., Yang, D.-Y., Selzle, H. L. & Schlag, E. W. Energetics of Hydrogen Bonds in Peptides. **100**, *Proc Natl Acad Sci U S A* (2003).
3. Roth, C. M., Neal, B. L. & Lenhoff, A. M. Van der Waals interactions involving proteins. *Biophys J* **70**, 977–987 (1996).
4. Corley, M., Burns, M. C. & Yeo, G. W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell* **78** 9–29 (2020).
5. Yu, Q., Ye, W., Jiang, C., Luo, R. & Chen, H. F. Specific recognition mechanism between RNA and the KH3 domain of Nova-2 protein. *Journal of Physical Chemistry B* **118**, 12426–12434 (2014).
6. Abraham Punnoose, J. *et al.* High-throughput single-molecule quantification of individual base stacking energies in nucleic acids. *Nat Commun* **14**, (2023).
7. Fu, X. D. Non-coding RNA: A new frontier in regulatory biology. *National Science Review* **1** 190–204 (2014).
8. Li, J. & Liu, C. Coding or noncoding, the converging concepts of RNAs. *Frontiers in Genetics* **10**, (2019).
9. Vicens, Q. & Kieft, J. S. Thoughts on how to think (and talk) about RNA structure *Proc Natl Acad Sci U S A* **119**, (2022).
10. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **20** 474–489 (2019).
11. Cao, X., Zhang, Y., Ding, Y. & Wan, Y. Identification of RNA structures and their roles in RNA functions. *Nature Reviews Molecular Cell Biology* **25**, 784–801 (2024).
12. Watakabe, A., Inoue, K., Sakamoto, H. & Shimura, Y. A Secondary Structure at the 3' Splice Site Affects Tle in Vidr Splicing Reaction of Mouse Immunoglobulin A Chain Pre-MRNAs. *Nucleic Acids Research* **17**, (1989).
13. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
14. Aw, J. G. A. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol Cell* **62**, 603–617 (2016).
15. Gonzalez, I., Buonomo, S. B. C., Nasmyth, K. & Von Ahsen, U. ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation. *Current Biology* **9**, 337–340 (1999).
16. Van Treeck, B. *et al.* RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc Natl Acad Sci U S A* **115**, 2734–2739 (2018).
17. Erckes, V. & Steuer, C. A story of peptides, lipophilicity and chromatography - back and forth in time. *RSC Medicinal Chemistry* **13**, 676–687 (2022).

18. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829–845 (2014).
19. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* **70**, 854-867.e9 (2018).
20. Mackereth, C. D. & Sattler, M. Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology* **22**, 287–3 (2012).
21. Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology* **19**, 327–341 (2018).
22. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell* **60**, 208–219 (2015).
23. O'Connell, L. C. *et al.* Intrinsically disordered regions and RNA binding domains contribute to protein enrichment in biomolecular condensates in Xenopus oocytes. *Sci Rep* **14**, (2024).
24. Liao, J. Y. *et al.* RBPWorld for exploring functions and disease associations of RNA-binding proteins across species. *Nucleic Acids Res* **53**, 220–232 (2025).
25. Roca-Martínez, J., Dhondge, H., Sattler, M. & Vranken, W. F. Deciphering the RRM-RNA recognition code: A computational analysis. *PLoS Comput Biol* **19**, (2023).
26. Ray, D. *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* **27**, 667–670 (2009).
27. Cléry, A., Blatter, M. & Allain, F. H. T. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology* **18** 290–298 (2008).
28. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal* **272**, 2118–2131 (2005).
29. Gupta, A. & Gribskov, M. The role of RNA sequence and structure in RNA-protein interactions. *J Mol Biol* **409**, 574–587 (2011).
30. Wilson, K. A., Holland, D. J. & Wetmore, S. D. Topology of RNA-protein nucleobase-amino acid π-π interactions and comparison to analogous DNA-protein π-π contacts. *RNA* **22**, 696–708 (2016).
31. Ohyama, T. *et al.* Structure of Musashi1 in a complex with target RNA: The role of aromatic stacking interactions. *Nucleic Acids Res* **40**, 3218–3231 (2012).
32. Ruth Zearfoss, N. *et al.* A conserved three-nucleotide core motif defines musashi RNA binding specificity. *Journal of Biological Chemistry* **289**, 35530–35541 (2014).
33. Imai, T. *et al.* The Neural RNA-Binding Protein Musashi1 Translationally Regulates Mammalian numb Gene Expression by Interacting with Its mRNA *Mol Cell Biol* **21**, 3888–3900 (2001).
34. Iwaoka, R. *et al.* Structural Insight into the Recognition of r(UAG) by Musashi-1 RBD2, and Construction of a Model of Musashi-1 RBD1-2 Bound to the Minimum Target RNA. *Molecules* **22**, (2017).
35. Chen, Y. *et al.* Short C-terminal Musashi-1 proteins regulate pluripotency states in embryonic stem cells. *Cell Rep* **42**, (2023).
36. Chen, T. C. & Huang, J. R. Musashi-1: An Example of How Polyalanine Tracts Contribute to Self-Association in the Intrinsically Disordered Regions of RNA-Binding Proteins. *Int J Mol Sci* **21**, (2020).

37. Cragle, C. E. *et al.* Musashi interaction with poly(A)-binding protein is required for activation of target mRNA translation. *Journal of Biological Chemistry* **294**, 10969–10986 (2019).
38. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
39. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **18** 285–298 (2017).
40. Zhao, B. *et al.* Intrinsic Disorder in Human RNA-Binding Proteins. *J Mol Biol* **433**, (2021).
41. Strulson, C. A., Molden, R. C., Keating, C. D. & Bevilacqua, P. C. RNA catalysis through compartmentalization. *Nat Chem* **4**, 941–946 (2012).
42. Nikpour, P. *et al.* The RNA binding protein Musashi1 regulates apoptosis, gene expression and stress granule formation in urothelial carcinoma cells. *J Cell Mol Med* **15**, 1210–1224 (2011).
43. Clingman, C. C. *et al.* Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. *Elife* **2014**, 1–26 (2014).
44. Yi, C. *et al.* Luteolin inhibits Musashi1 binding to RNA and disrupts cancer phenotypes in glioblastoma cells. *RNA Biol* **15**, 1420–1432 (2018).
45. Minuesa, G. *et al.* Small-molecule targeting of MUSASHI RNA-binding activity in acute myeloid leukemia. *Nat Commun* **10**, 1–15 (2019).
46. Nakamura, M., Okano, H., Blendy, J. A. & Montell, C. Musashi, a Neural RNA-Binding Protein Required for Drosophila Adult External Sensory Organ Development. *Neuron* **13**, (1994).
47. Sakakibara, S.-I. *et al.* Mouse-Musashi-1, a Neural RNA-Binding Protein Highly Enriched in the Mammalian CNS Stem Cell. *Developmental Biol* **176**, (1996).
48. Battelli, C., Nikopoulos, G. N., Mitchell, J. G. & Verdi, J. M. The RNA-binding protein Musashi-1 regulates neural development through the translational repression of p21 WAF-1. *Molecular and cellular neurosciences* **31**, 85–96 (2006).
49. Muto, J. *et al.* RNA-binding protein musashi1 modulates glioma cell growth through the post-transcriptional regulation of notch and PI 3 kinase/Akt signaling pathways. *PLoS One* **7**, (2012).
50. Pastó, A. *et al.* NOTCH3 signaling regulates MUSASHI-1 expression in metastatic colorectal cancer cells. *Cancer Res* **74**, 2106–2118 (2014).
51. Wang, X.-Y. *et al.* Musashi1 Modulates Mammary Progenitor Cell Expansion through Proliferin-Mediated Activation of the Wnt and Notch Pathways. *Mol Cell Biol* **28**, 3589–3599 (2008).
52. Rezza, A. *et al.* The overexpression of the putative gut stem cell marker Musashi-1 induces tumorigenesis through Wnt and Notch activation. *J Cell Sci* **123**, 3256–3265 (2010).
53. Luo, Z. *et al.* NUMB enhances Notch signaling by repressing ubiquitination of NOTCH1 intracellular domain. *J Mol Cell Biol* **12**, 345–358 (2020).
54. Zhou, B. *et al.* Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduction and Targeted Therapy* **7**, (2022).
55. Jadhav, S. *et al.* RNA-binding Protein Musashi Homologue 1 Regulates Kidney Fibrosis by Translational Inhibition of p21 and Numb mRNA. *Journal of Biological Chemistry* **291**, 14085–14094 (2016).
56. Glazer, R. I., Wang, X.-Y., Yuan, H. & Yin, Y. Musashi1: A Stem Cell Marker No Longer in Search of a Function *Cell Cycle* **17**, 2635–2639 (2008).

57. Charlesworth, A., Wilczynska, A., Thampi, P., Cox, L. L. & MacNicol, A. M. Musashi regulates the temporal order of mRNA translation during Xenopus oocyte maturation. *EMBO Journal* **25**, 2792–2801 (2006).
58. Barber, C. N. & Raben, D. M. Lipid metabolism crosstalk in the brain: Glia and neurons. *Frontiers in Cellular Neuroscience* **13**, (2019).
59. Li, N. *et al.* The Msi Family of RNA-Binding Proteins Function Redundantly as Intestinal Oncoproteins Article The Msi Family of RNA-Binding Proteins Function Redundantly as Intestinal Oncoproteins. *CellReports* **13**, 2440–2455 (2015).
60. Lachinani, L. *et al.* The oncogene Musashi1 encodes novel miRNAs in breast cancer. *Sci Rep* **13**, (2023).
61. Bley, N. *et al.* Musashi-1—a stemness RBP for cancer therapy? *Biology* **10**, (2021).
62. Chiou, G. Y. *et al.* Musashi-1 promotes a cancer stem cell lineage and chemoresistance in colorectal cancer cells. *Sci Rep* **7**, 1–13 (2017).
63. Lin, J. C., Tsai, J. T., Chao, T. Y., Ma, H. I. & Liu, W. H. Musashi-1 Enhances Glioblastoma Migration by Promoting ICAM1 Translation. *Neoplasia (United States)* **21**, 459–468 (2019).
64. Chen, H. Y. *et al.* Musashi-1 promotes chemoresistant granule formation by PKR/eIF2α signalling cascade in refractory glioblastoma. *Biochim Biophys Acta Mol Basis Dis* **1864**, 1850–1861 (2018).
65. Chen, H., Lin, L., Wang, M. & Laurent, B. Musashi-1 Enhances Glioblastoma Cell Migration and Cytoskeletal Dynamics through Translational Inhibition of Tensin3. *Sci Rep* **7**, 1–14 (2017).
66. Lagadec, C. *et al.* The RNA-binding protein musashi-1 regulates proteasome subunit expression in breast cancerand glioma-initiating cells. *Stem Cells* **32**, 135–144 (2014).
67. Nahas, G. R. *et al.* The RNA-binding protein Musashi 1 stabilizes the oncotachykinin 1 mRNA in breast cancer cells to promote cell growth. *FASEB Journal* **30**, 149–159 (2016).
68. Troschel, F. M. *et al.* Knockdown of Musashi RNA binding proteins decreases radioresistance but enhances cell motility and invasion in triple-negative breast cancer. *Int J Mol Sci* **21**, (2020).
69. Shou, Z. *et al.* Overexpression of Musashi-1 protein is associated with progression and poor prognosis of gastric cancer. *Oncol Lett* **13**, 3556–3566 (2017).
70. Kuang, R. G. *et al.* Expression and significance of Musashi-1 in gastric cancer and precancerous lesions. *World J Gastroenterol* **19**, 6637–6644 (2013).
71. Orian-Rousseau, V. & Sleeman, J. CD44 is a Multidomain Signaling Platform that Integrates Extracellular Matrix Cues with Growth Factor and Cytokine Signals. *Adv Cancer Res* **123**, 231–254 (2014).
72. Senbanjo, L. T. & Chellaiah, M. A. CD44: A multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front Cell Dev Biol* **5**, (2017).
73. Mesrati, M. H., Syafruddin, S. E., Mohtar, M. A. & Syahir, A. CD44: A multifunctional mediator of cancer progression. *Biomolecules* **11**, (2021).
74. Chen, C., Zhao, S., Karnad, A. & Freeman, J. W. The biology and role of CD44 in cancer progression: Therapeutic implications. *Journal of Hematology and Oncology* **11**, (2018).
75. Li, C. *et al.* RNA-binding protein LSM7 facilitates breast cancer metastasis through mediating alternative splicing of CD44. *Life Sci* **356**, (2024).
76. Liu, S. *et al.* CD44 is a potential immunotherapeutic target and affects macrophage infiltration leading to poor prognosis. *Sci Rep* **13**, (2023).

77. Xu, H., Niu, M., Yuan, X., Wu, K. & Liu, A. CD44 as a tumor biomarker and therapeutic target. *Experimental Hematology and Oncology* **9**, (2020).
78. Pötschke, R. *et al.* MSI1 promotes the expression of the GBM stem cell marker CD44 by impairing miRNA-dependent degradation. *Cancers (Basel)* **12**, 1–18 (2020).
79. Montalbano, M. *et al.* RNA-binding proteins Musashi and tau soluble aggregates initiate nuclear dysfunction. *Nat Commun* **11**, 1–16 (2020).
80. Sengupta, U. *et al.* Formation of Toxic Oligomeric Assemblies of RNA-binding Protein: Musashi in Alzheimer's disease. *Acta Neuropathol Commun* **6**, 113 (2018).
81. Montalbano, M. *et al.* Tau oligomers mediate aggregation of RNA-binding proteins Musashi1 and Musashi2 inducing Lamin alteration. *Aging Cell* **18**, (2019).
82. Chiu, S. H., Ho, W. L., Sun, Y. C., Kuo, J. C. & Huang, J. rong. Phase separation driven by interchangeable properties in the intrinsically disordered regions of protein paralogs. *Commun Biol* **5**, (2022).
83. Morelli, C. *et al.* RNA modulates hnRNPA1A amyloid formation mediated by biomolecular condensates. *Nat Chem* **16**, (2024).
84. Musso, D. & Gubler, D. J. Zika virus. *Clin Microbiol Rev* **29**, 487–524 (2016).
85. Chen, X. *et al.* Zika virus RNA structure controls its unique neurotropism by bipartite binding to Musashi-1. *Nat Commun* **14**, (2023).
86. Jarmoskaite, I., Alsadhan, I., Vaidyanathan, P. P. & Herschlag, D. How to measure and evaluate binding affinities. *Elife* **9**, 1–34 (2020).
87. Schreiber, G., Haran, G. & Zhou, H. X. Fundamental aspects of protein - Protein association kinetics. *Chemical Reviews* **109**, 839–860 (2009).
88. Lu, H. & Tonge, P. J. Drug-target residence time: Critical information for lead optimization. *Current Opinion in Chemical Biology* **14**, 467–474 (2010).
89. Kruse, A. C. *et al.* Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482**, 552–556 (2012).
90. Yin, N., Pei, J. & Lai, L. A comprehensive analysis of the influence of drug binding kinetics on drug action at molecular and systems levels. *Mol Biosyst* **9**, 1381–1389 (2013).
91. Tonge, P. J. Drug-Target Kinetics in Drug Discovery. *ACS Chemical Neuroscience* **9**, 29–39 (2018).
92. Copeland, R., Pompliano, D. & Meek, T. Drug–target residence time and its implications for lead optimization. *Nat Rev Drug Discov* **5**, 730–739 (2006).
93. Tummino, P. J. & Copeland, R. A. Residence time of receptor - Ligand complexes and its effect on biological function. *Biochemistry* **47**, 5481–5492 (2008).
94. Schoop, A. & Dey, F. On-rate based optimization of structure-kinetic relationship - Surfing the kinetic map. *Drug Discovery Today: Technologies* **17**, 9–15 (2015).
95. Dahl, G. & Akerud, T. Pharmacokinetics and the drug-target residence time concept. *Drug Discov Today* **18**, 697–707 (2013).
96. Sohrabi-Jahromi, S. & Söding, J. Thermodynamic modeling reveals widespread multivalent binding by RNA-binding proteins. *Bioinformatics* **37**, I308–I316 (2021).
97. Schneider, T. *et al.* Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nat Commun* **10**, (2019).
98. Falkenberg, C. V., Carson, J. H. & Blinov, M. L. Multivalent Molecules as Modulators of RNA Granule Size and Composition. *Biophys J* **113**, 235–245 (2017).

99. Grese, Z. R. *et al.* Specific RNA interactions promote TDP-43 multivalent phase separation and maintain liquid properties. *EMBO Rep* **22**, (2021).
100. Cabral, S. E., Otis, J. P. & Mowry, K. L. Multivalent interactions with RNA drive recruitment and dynamics in biomolecular condensates in Xenopus oocytes. *iScience* **25**, 104811 (2022).
101. Campos-Melo, D., Hawley, Z. C. E., Droppelmann, C. A. & Strong, M. J. The Integral Role of RNA in Stress Granule Formation and Function. *Frontiers in Cell and Developmental Biology* **9**, (2021).
102. Johnsson, B. *et al.* Comparison of Methods for Immobilization to Carboxymethyl Dextran Sensor Surfaces by Analysis of the Specific Activity of Monoclonal Antibodies. *Journal of Molecular Recognition* **8**, (1995).
103. Liedberg, B. O., Nylander, C. & Lundstrom, I. Surface Plasmon Resonance for Gas Detection and Biosensing. *Sensors and Actuators* **4**, (1983).
104. Johnsson, B. *et al.* Immobilization of Proteins to a Carboxymethyldextran-Modified Gold Surface for Biospecific Interaction Analysis in Surface Plasmon Resonance Sensors. *Analytical Biochemistry* **198**, (1991).
105. Stenberg, E., Roos, H. & Urbaniczky, C. Quantitative Determination of Surface Concentration of Protein with Surface Plasmon Resonance Using Radiolabeled Proteins. *Journal of Colloid and Interface Science* **143**, (1991).
106. Katsamba, P. S., Park, S. & Laird-offringa, I. A. Kinetic studies of RNA – protein interactions using surface plasmon resonance. *Methods* **26**, 95–104 (2002).
107. Heber, S. *et al.* Staufen2-mediated RNA recognition and localization requires combinatorial action of multiple domains. *Nat Commun* **10**, (2019).
108. Pérez-Ropero, G., Pérez-Ràfols, A., Martelli, T., Danielson, U. H. & Buijs, J. Unraveling the Bivalent and Rapid Interactions Between a Multivalent RNA Recognition Motif and RNA: A Kinetic Approach. *Biochemistry* (2024).
109. Koutcher, J. A. & Burt, C. T. The rising number of biomedical applications of nu. 101–111 (1984).
110. Roberts, A. E. L., Kragh, K. N., Bjarnsholt, T. & Diggle, S. P. The Limitations of in Vitro Experimentation in Understanding Biofilms and Chronic Infection. *Journal of Molecular Biology* **427**, 3646–3661 (2015).
111. Pina, A. S., Lowe, C. R. & Roque, A. C. A. Challenges and opportunities in the purification of recombinant tagged proteins. *Biotechnology Advances* **32**, 366–381 (2014).
112. England, W. E. *et al.* An atlas of posttranslational modifications on RNA binding proteins. *Nucleic Acids Res* **50**, 4329–4339 (2022).
113. Olmsted, I. R., Kussrow, A. & Bornhop, D. J. Comparison of free-solution and surface-immobilized molecular interactions using a single platform. *Anal Chem* **84**, 10817–10822 (2012).
114. Bondza, S. *et al.* Real-time characterization of antibody binding to receptors on living immune cells. *Front Immunol* **8**, 1–11 (2017).
115. Önell, A. & Andersson, K. Kinetic determinations of molecular interactions using Biacore--minimum data requirements for efficient experimental design. *J Mol Recognit* **18**, 307–317 (2005).
116. Encarnação, J. C., Schulte, T., Achour, A., Björkelund, H. & Andersson, K. Detecting ligand interactions in real time on living bacterial cells. *Appl Microbiol Biotechnol* **102**, 4193–4201 (2018).
117. Markossian, S., Grossman, A. & Arkin, M. Analyzing Kinetic Binding Data. *Advancing Translational Sciences* (2021).
118. Langmuir, I. The adsorption of gases on plane surfaces of glass, mica and platinum. *J Am Chem Soc* **40**, 1361–1403 (1918).

119. Schuck, P. & Minton, A. P. Analysis of Mass Transport-Limited Binding Kinetics in Evanescent Wave Biosensors. *Analytical Biochemistry* **240**, (1996).
120. Antigen-Antibody Binding and Mass Transport by Convection and Diffusion to a Surface.
121. Myszka, D. G., He, X., Dembo, M., Morton, T. A. & Goldstein, B. Extending the Range of Rate Constants Available from BIACORE: Interpreting Mass Transport-Influenced Binding Data. *Biophysical Journal* **75**, (1998).
122. Karlsson, R. & F'ilt, A. Experimental Design for Kinetic Analysis of Protein-Protein Interactions with Surface Plasmon Resonance Biosensors *Journal of Immunological Methods* **200**, (1997).
123. Weber, J. *et al.* Modelling ligand depletion for simultaneous affinity and binding site quantification on cells and tissue. *Sci Rep* **13**, (2023).
124. Gakamsky, D. M. *et al.* Kinetic evidence for a ligand-binding-induced conformational transition in the T cell receptor. *Proc Natl Acad Sci U S A* **104**, 16639–16644 (2007).
125. Vauquelin, G., Van Liefde, I. & Swinney, D. C. On the different experimental manifestations of two-state 'induced-fit' binding of drugs to their cellular targets. *British Journal of Pharmacology* **173**, 1268–1285 (2016).
126. Vauquelin, G. & Charlton, S. J. Exploring avidity: Understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br J Pharmacol* **168**, 1771–1785 (2013).
127. Nguyen, K. *et al.* Parameter estimation and identifiability analysis for a bivalent analyte model of monoclonal antibody-antigen binding. *Anal Biochem* **679**, (2023).
128. Altschuh, D. *et al.* Deciphering complex protein interaction kinetics using Interaction Map. *Biochem Biophys Res Commun* **428**, 74–79 (2012).
129. Gompertz, B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* **115**, (1825).
130. Winsor, C. P. The Gompertz Curve as a growth curve. *Proc Natl Acad Sci U S A* **18**, (1932).
131. Laird, A. K. Dynamics of Tumour Growth. *British journal of cancer* **13**, (1964).
132. Zwietering, M. H., Jongenburger, I., Rombouts, F. M., Van ' K. & Riet, T. Modeling of the Bacterial Growth Curve. *Applied and environmental microbiology* **56**, (1990).
133. Hou, C. Energetic cost of biosynthesis is a missing link between growth and longevity in mammals. *Proc Natl Acad Sci U S A* **121**, (2024).
134. Tjørve, K. M. C. & Tjørve, E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS One* **12**, 1–17 (2017).
135. O'shannessy, D. J. & Winzor, D. J. Interpretation of Deviations from Pseudo-First-Order Kinetic Behavior in the Characterization of Ligand Binding by Biosensor Technology. *Analytical Biochemistry* **236**, (1996).
136. Myszka, D. G. Improving biosensor analysis. *Journal of Molecular Recognition* **12**, 279–284 (1999).
137. Katsamba, P. S. *et al.* Kinetic analysis of a high-affinity antibody/antigen interaction performed by multiple Biacore users. *Anal Biochem* **352**, 208–221 (2006).
138. Svoboda, P. & Di Cara, A. Hairpin RNA: A secondary structure of primary importance. *Cellular and Molecular Life Sciences* **63**, 901–918 (2006).

139. Dolcemascolo, R. *et al.* Repurposing the mammalian RNA-binding protein Musashi-1 as an allosteric translation repressor in bacteria. *Elife* **12**, (2024).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2508

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-552069

ACTA UNIVERSITATIS
UPSALIENSIS
2025