

Fluorescence Studies of Influenza RNA and RNA Polymerase



Alexandra Iulia Tomescu

A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy at the University of Oxford.

Balliol College, University of Oxford
17 January 2014

To my grandparents
Mami & Em, Mama Cuta & Tata Pilu
for guiding my curiosity in my early explorations of the world.

To my parents
Ada & Bogdan
for teaching me about unconditional love and true passion.

To my brother
Vlad
for all his amazingness.

Acknowledgements

I would have not managed to travel this far, were it not for the outstanding people I have met along the way, who believing in me, my ideas and my pursuits, mentored, encouraged, supported and inspired me, sometimes without even realising it. Thank you!

I start this acknowledgement by thanking Prof Achillefs Kapanidis, whose constant scientific buzz and care towards the members of his group were an inspiration. My time in the lab wouldn't have been the same without the members of the Kapanidis research group (past and present), who have contributed to a scientifically-stimulating, fun and friendly working environment.

I am grateful to Dr Nicole Robb for all the discussions, suggestions and practical support that have undoubtedly helped the progression of my work on influenza RNA polymerase – promoter interaction; to Marko Šuštaršič and Dr Timothy Craggs for their advice on 3D modelling of the RNA promoter; to Gabriel Borvak, Narin Hengrung, Ashley York, Jane Sharps, Dr Edward Hutchinson, Dr Frank Vreede and Prof Ervin Fodor from the Fodor research group (Dunn School of Pathology, University of Oxford) for their useful discussions and support with all my influenza-related work and to Dr Afaf El-Sagheer and Prof Tom Brown from the Brown research group (Department of Chemistry, University of Oxford) for their guidance and support with the click chemistry project.

A big 'thank you' also goes to Anne Plochowietz for her help with all things physics, and to Dr Louise Aigrain and Dr David Bauer who, alongside Dr Nicole Robb, have taken the time to review and proofread my thesis, and provided me with useful comments.

A special word of thanks goes to: Joe, Athan (Kris), Denise, Rita, Konrad, Erato, Cristiana, Anca, Louise WJ and Wander for their friendship, stimulating discussions, inspiration and support in times of doubt.

My work was funded by the Medical Research Council through a Doctoral Training Award from the Department of Physics, a one-year Ratiu Foundation Scholarship and a research grant from the European Research Council awarded to A. Kapanidis.

Fluorescence Studies of Influenza RNA and RNA Polymerase

Alexandra Iulia Tomescu

Balliol College

A thesis submitted for the degree of Doctor of Philosophy

University of Oxford, Michaelmas Term 2013

Abstract

The influenza A virus genome consists of eight single-stranded segments of negative-sense viral RNA (vRNA) with highly conserved, partially complementary termini. These termini associate in a double-stranded RNA structure, known as a panhandle, which is bound by the viral RNA-dependent RNA polymerase and can serve as a promoter in both viral transcription and replication.

In part A of this thesis, I use a combination of classical biochemistry techniques and fluorescence techniques (both at the ensemble and single-molecule level) for a quantitative investigation of the interaction between purified influenza A RNA polymerase and the individual 5' and 3' conserved termini of the vRNA segments, as well as the double-stranded vRNA promoter. Furthermore, I report the first direct, real-time observation of the promoter changing its structure when bound by the polymerase and show that the structure assumed agrees best with the corkscrew model.

In part B of this thesis, I use fluorescence to detect RNA: I design and test a single-molecule biosensor aimed at probing the presence of influenza A RNA in a sample, on the one hand, and I use click-chemistry to fluorescently label very shorty RNAs (3-25nt) that have been generated in an *in vitro* transcription reaction, on the other. The biosensing assay I propose can be further developed for diagnostic purposed, while click-chemistry labelling of short RNAs can be optimised and extended such that it becomes a reliable alternative to the use of radiolabels.

Outline

While traditionally doctorates are concerned with only one problem, early on in my degree, I asked whether it would be possible to have more variety: this resulted in working on the three different projects presented in this thesis, all of which have as underlying themes the use of fluorescence for investigating or reporting on biological processes, on the one hand, and RNA biology, on the other.

A short introduction to fluorescence and single-molecule fluorescence makes up the first section of this thesis.

To present my findings in the best possible manner, I have opted to divide my thesis into two main parts:

- Part A is entitled “Investigating influenza A RNA polymerase – promoter binding using fluorescence techniques” and consists of three chapters.
 - Chapter 1 gives an overview of the influenza RNA polymerase, focusing on its expression, purification and characterisation;
 - Chapter 2 provides insight as to how influenza RNA polymerase interacts with the single-stranded conserved ends of the viral genome. The interaction was considered from a quantitative point of view, which allowed for determination of binding constants.
 - Chapter 3 extends the study of the interaction between the RNA polymerase and the conserved viral termini to the viral double-stranded RNA promoter. The interaction was considered from both a quantitative and a structural point of view: a binding constant for the interaction was determined and the structure assumed by the promoter when bound by the polymerase was further clarified in view of proposed models. Finally, the experimental data generated was used to generate a 3D model of the bound, double-stranded promoter.
- Part B is entitled “Fluorescence-based assays for RNA detection” and consists of two independent chapters.

- Chapter 4 discusses the development of a fluorescence-based biosensing assay targeted at detecting influenza RNA (the positive-sense viral mRNA and cRNA) in a sample of total extracted RNA from eukaryotic cells infected with the virus.
- Chapter 5 investigates whether copper-free click chemistry can be used to fluorescently label short RNA strands (with lengths varying between 3 and 25 nucleotides).

Each chapter in this thesis consists of five sections, starting with an introduction, followed by the materials and methods section, which is in turn followed by sections in which the main findings are presented and discussed. At the end of the chapter, I have acknowledged those who have contributed to the work presented (by providing me with reagents or data analysis assistance) and listed the publications that have been cited. Where appropriate, a section detailing further work (“Perspectives”) has also been included.

Contents

Acknowledgements	ii
Abstract.....	iii
Outline	iv
Introduction to single-molecule fluorescence.....	1
1. What is fluorescence?.....	1
2. Fluorescence resonance energy transfer (FRET)	4
3. Single-molecule FRET (smFRET).....	7
4. Microscopy	8
5. Alternating laser excitation (ALEX)	10
6. From apparent to accurate FRET	12
7. Bibliography.....	14
Part A: Investigating influenza A RNA polymerase – promoter binding using fluorescence techniques	16
Chapter 1: Influenza RNA polymerase: expression, purification and characterisation.....	17
1.1 INTRODUCTION.....	17
1.2 MATERIALS AND METHODS.....	22
1.2.1 Cloning of mOrange2 in the PB2—TAP pcDNA vector	22
1.2.2 Reverse plating PCR (to determine the orientation of the mOrange2 gene in the PB2-mOrange2-TAP construct).....	23
1.2.3 Plasmid production (PB2-mOrange2-TAP pcDNA vector)	24
1.2.4 H1N1 RNA polymerase expression and purification.....	24
1.2.5 H3N2 RNA polymerase expression and purification.....	26
1.2.6 Assessing the purity of purified protein samples	26
1.2.7 Activity assay (<i>in vitro</i> transcription)	27
1.2.8 Binding assay (single-molecule fluorescence spectroscopy)	28
1.2.9 Protein quantitation (SyproRuby stain).....	29
1.2.10 Protein quantitation (ensemble fluorescence).....	29
1.2.11 Protein quantitation (single-molecule fluorescence)	30
1.3 RESULTS	31
1.3.1 Expression and purification of unlabelled and GFP-tagged H1N1 RNA polymerase	31

1.3.2	Generating the PB2- mOrange2-TAP plasmid	32
1.3.3	Expression, purification and functional characterisation of the mOrange2-labelled H1N1 RNA polymerase	34
1.3.4	Protein concentration determination.....	37
1.3.5	Characterisation of the H3N2 RNA polymerase	40
1.4	DISCUSSION	42
1.5	CONTRIBUTIONS	43
1.6	BIBLIOGRAPHY	43
Chapter 2: Studying the interaction between RNA polymerase and single stranded RNAs		47
2.1	INTRODUCTION.....	47
2.2	MATERIALS AND METHODS.....	50
2.2.1	RNA polymerase	50
2.2.2	RNA strands	50
2.2.3	RNA labelling and purification	51
2.2.4	Determining labelled RNA concentration and labelling efficiency.....	52
2.2.5	Electrophoretic Mobility Shift Assay (EMSA).....	53
2.2.6	Protein Induced Fluorescence Enhancement (single-molecule; on surface)	53
2.2.7	Fluorescence anisotropy.....	54
2.3	RESULTS	55
2.3.1	Investigating polymerase binding to the 5' end of the vRNA in solution	55
2.3.2	Investigating polymerase binding to the 5' end of the vRNA on the surface	58
2.3.3	Investigating polymerase binding to the 3' end of the vRNA in solution	62
2.3.4	Investigating polymerase binding to the 3' end of the vRNA on the surface	66
2.3.5	Polymerase binding to a non-influenza ssRNA	69
2.3.6	Investigating the polymerase binding sequentially to the 5' and 3' ends	70
2.3.7	Investigating polymerase binding to the cRNA promoter ends	74
2.4	DISCUSSION	75
2.5	CONTRIBUTIONS	78
2.6	BIBLIOGRAPHY	79
Chapter 3: Studying the RNA polymerase – promoter interaction		81
3.1	INTRODUCTION.....	81
3.2	MATERIALS AND METHODS.....	84

3.2.1	RNA polymerase	84
3.2.2	RNA constructs	84
3.2.3	Electrophoretic Mobility Shift Assay.....	85
3.2.4	Polymerase activity assay (<i>in vitro</i> transcription).....	86
3.2.5	Binding assay (single-molecule fluorescence spectroscopy)	86
3.2.6	RNA secondary structure prediction	87
3.2.7	2D to 3D conversion	88
3.2.8	Dye positioning	88
3.2.9	Determining distances in the 3D RNA structures	89
3.3	RESULTS	90
3.3.1	Is binding influenced by the fluorophore position on the dsRNA construct?	90
3.3.2	Qualitative investigation of the polymerase interacting with the dsRNA promoter.....	91
3.3.3	Structural investigation of the polymerase-dsRNA interaction.....	94
3.3.4	K_D determination	101
3.3.5	Analysis of binding to a double-stranded vRNA promoter using single molecule quenchable FRET (quFRET)	102
3.3.6	Distance determination and 3D modelling.....	103
3.3.7	RNA polymerase binding to the dsRNA construct in the presence of ApG dinucleotide	110
3.4	DISCUSSION	112
3.5	CONTRIBUTIONS	116
3.6	BIBLIOGRAPHY	116
Part B:	Fluorescence-based assays for RNA detection.....	119
Chapter 4:	A biosensor for influenza RNA detection.....	120
4.1	INTRODUCTION.....	120
	Proposed detection assay.....	122
4.2	MATERIALS AND METHODS.....	126
4.2.1	Probes	126
4.2.2	Eukaryotic cell culture	127
4.2.3	Virus purification	128
4.2.4	Plaque assay.....	129
4.2.5	Viral RNA preparation.....	130

4.2.6	Primer extension assay	132
4.2.7	Biosensing assay	132
4.2.8	Binding troubleshooting: PAGE electrophoresis	133
4.2.9	Binding troubleshooting: digestion assay	133
4.3	RESULTS	134
4.3.1	Initial attempts.....	134
4.3.2	Troubleshooting.....	140
4.3.2.1	RNA availability in the sample	140
4.3.2.2	The viral RNA extracted is rich in secondary structure.....	145
4.3.2.3	Binding conditions – capture probe	146
4.3.2.4	Binding conditions – detection probe	148
4.4	DISCUSSION	154
4.5	PERSPECTIVES	156
4.6	BIBLIOGRAPHY	158
Chapter 5: Labelling short RNAs using click-chemistry.....		160
5.1	INTRODUCTION.....	160
5.1.1	Motivation	160
5.1.2	Bacterial transcription	163
5.1.3	Fluorescently labelling short RNAs	166
5.1.3.1	Enzymatic labelling	167
5.1.3.2	Click chemistry strategies	168
5.2	MATERIALS AND METHODS.....	170
5.2.1	Nucleic acid strands	170
5.2.2	Specific click chemistry reagents	171
5.2.3	Click chemistry labelling of standard RNAs	172
5.2.4	Open complex formation.....	173
5.2.5	<i>In vitro</i> transcription (IVT) reaction	174
5.2.6	Click chemistry labelling of IVT RNA products and clean up	174
5.2.7	Silver staining and Sybr Green II staining	175
5.3	RESULTS AND DISCUSSION	176
5.3.1	Transcription profiles: unmodified ApA vs. 5' azide-modified ApA.....	176
5.3.2	Click chemistry labelling of RNA abortive products post electrophoresis.....	179

5.3.3	Click chemistry labelling of RNA abortive products during electrophoresis	181
5.3.4	Click chemistry labelling of RNA abortive products before electrophoresis	185
5.4	PERSPECTIVES	195
5.5	CONTRIBUTIONS	197
5.6	BIBLIOGRAPHY	197
Appendix: List of presentations and publications		201

Introduction to single-molecule fluorescence

1. What is fluorescence?

The underlying thread of the work presented in this thesis is the use of fluorescent techniques for investigation of biological processes. Hence, I believe it is not unreasonable to start by asking what fluorescence is.

Fluorescence is the process through which a molecule relaxes from an electronically excited state to a more stable ground state by emission of a photon. The different electronic states of a fluorescent molecule and the fluctuation of energy between them are best described by the Jablonski diagram (fig. 1). When light is shone on a fluorophore, the molecule is able to absorb a high energy photon (shorter wavelength) which will cause an electron from the ground state (S_0) to jump to an electronically excited state (S_1). Once in the excited state, the electron will undergo vibrational relaxation until it reaches the lowest vibration level of the excited state; from here, to return to the S_0 state, a molecule can take different routes: fluorescence, phosphorescence and quenching (Lackowicz 2006).

Fluorescence is the mechanism by which the molecule returns to the ground state via emission of a higher wavelength photon (than the one absorbed initially), a mechanism that occurs with a lifetime of 10^{-8} seconds. The difference in wavelength between the absorbed and the emitted photons is known as the Stokes shift. The Stokes shift is a key feature of fluorescence and has practical implications, allowing for more of the emitted

light to be collected and separated from the light used to excite the fluorophore in the first place (Lackowicz 2006).

Phosphorescence takes place when the excited molecule undergoes intersystem crossing (a process by which the molecule undergoes spin conversion from the excited S_1 state to the first triplet state T_1), after which it emits a photon to return to the ground state; this process is much slower than fluorescence, having a lifetime of 10^{-3} seconds. The formation of the triplet state also plays a role in the photophysics of the fluorophore, which can undergo blinking (when the molecule enters a reversible dark state) or even photobleaching (when the molecule is no longer able to go through the excitation-emission cycle and fluoresce). The occurrence of these two phenomena is known to be affected by the presence of oxygen in the solution the fluorophore is in; indeed, photobleaching is specifically known to increase due to photooxidation (Turro et al. 2010). Hence, many of the measurements involving fluorescence are taken in the presence of oxygen scavengers, such as Gloxy – a mix of glucose oxidase, catalase and glucose (Benesch & Benesch 1953) – which is also used in the work presented here. Yet, depleting the buffer solution of too much oxygen negatively influences the ability of the fluorophore to emit photons via fluorescence, biasing it towards formation of the triple state, which is why, in addition to the oxygen scavenger system used, the buffers employed in this research were supplemented with 6-hydroxyl-2,5,7,8-tetramethychroma-2-carboxylic acid (Trolox; Rasnik et al. 2006; Cordes et al. 2009), a triple state quencher.

Finally, quenching of fluorescence can occur through collision with other molecules (collisional or dynamic quenching) or through formation of non-fluorescent complexes between the fluorescent molecule and the quencher (static quenching) (Lackowicz 2006).

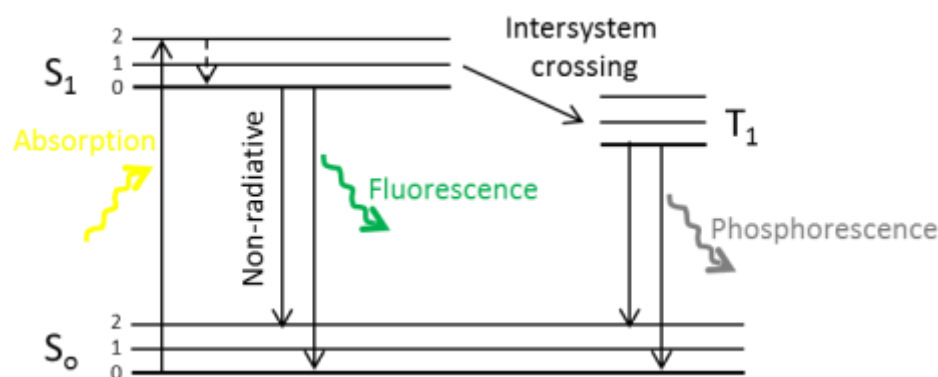


Fig. 1 – Jablonski diagram, showing two different electronic states of a fluorophore: the singlet ground state (S_0) and the first (excited) state (S_1), both of which can have a number of vibrational energy levels (0, 1, 2 in this figure). The diagram shows how the excited molecule returns from a high vibrational energy level to the lowest vibrational energy level of the excited state via vibrational relaxation (dashed arrow), and then to the singlet ground state S_0 , either through a non-radiative process or a radiative process (fluorescence); alternatively, the molecule can enter a triplet state via intersystem crossing, from where it can return to the ground state via phosphorescence (diagram adapted from (Lackowicz 2006)).

Fluorescence is further dependent on two characteristics of the fluorescent molecule – quantum yield and fluorescence lifetime. Quantum yield (ϕ) is a measure of the efficiency of fluorescence, being defined as the ration between the number of emitted photons and the total number of photons absorbed; it is, in other words, a measure of the brightness of the fluorophore. A quantum yield value of ~ 1 suggests that the main route by which the excited fluorescent molecule returns to the ground state is through photon emission. Fluorescence lifetime (τ) is a measure of the average time the fluorophore spends in the excited state S_1 .

2. Fluorescence resonance energy transfer (FRET)

Rather than emitting a photon, a fluorophore can return to the ground state by the non-radiative transfer of energy to a second molecule, in a process known as Förster Resonance Energy Transfer (FRET; after Theodor Förster, who first described it in 1948). This type of energy transfer is only possible if the emission spectrum of the first molecule overlaps with the excitation spectrum of the second one, with the first molecule said to act as a donor, while the second, as an acceptor (Clegg 1992). When the acceptor (A) is a fluorescent molecule, the transfer of energy can lead to photon emission (fig. 2A and B). The efficiency of FRET is directly dependent on the distance between the donor and the acceptor molecule, as the interaction between the two is a dipole-dipole interaction. This allows the technique to be used as a spectroscopic ruler, effective at reporting distances that range between 2 and 10 nm (fig. 2C; Streyer & Haugland 1967).

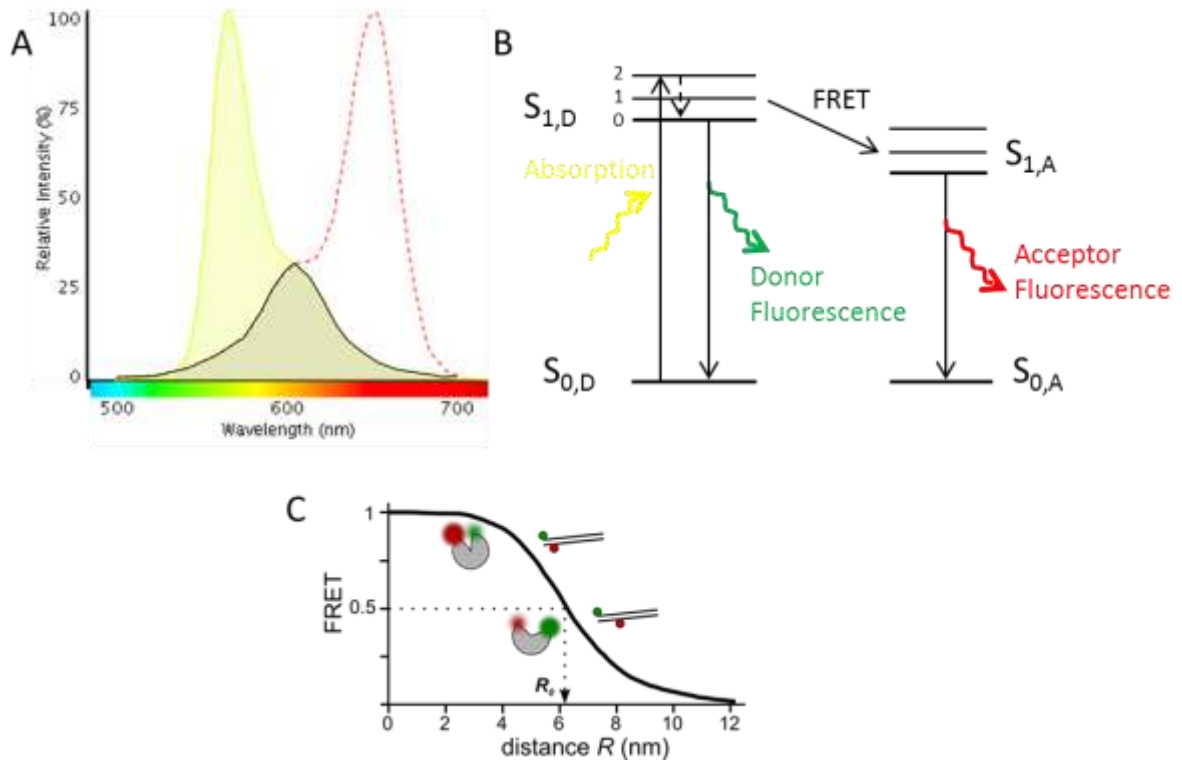


Fig. 2 – A) A normalised emission/excitation spectrum for the Cy3-Alexa647 FRET pair: energy transfer between the two is possible when the emission spectra of the donor (Cy3, in yellow) overlaps with the excitation spectrum of an acceptor (Alexa647, dotted red line). The overlapping region of the spectra is an important determining factor for the Förster radius, which is the distance at which the FRET efficiency is 50%. B) A simplified Jablonski diagram showing the path of energy transfer between an excited donor molecule (D) and an acceptor molecule (A). C) An illustration of the distance-dependence of FRET efficiency, showing that the closest the donor and acceptor are, the higher the energy transfer between them.

The rate of energy transfer between the donor and the acceptor is given by:

$$k_T = \frac{1}{\tau_D} \left(\frac{R_0}{r} \right)^6$$

τ_D = fluorescence lifetime of the donor

R_0 = Förster radius at which the D-A energy transfer is 50%

r = distance between D and A

The Förster radius at which the energy transfer between D and A is 50% is specific to each donor-acceptor pair and can be calculated using the following formula:

$$R_0^6 = \frac{9000 \ln(10) \Phi_D \kappa^2 J}{128 \pi^5 n^4 N_A}$$

Φ_D = quantum yield of D (in the absence of A)

κ^2 = relative dipole orientation factor for the transition of D emission and A absorption

transition dipoles (κ^2 can vary between 0 and 4 as it is dependent on the rotational freedom of the fluorophores; for the purpose of the work in this thesis, $\kappa^2 = 2/3$ which

assumes unrestricted rotational freedom of at least one of the dipoles)

n = refractive index of the medium ($n = 1.4$ for aqueous solutions)

N_A = Avogadro's number

J = integral of the spectral overlap between D emission and A absorption

The J integral is defined as:

$$J = \int_0^{\infty} F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda$$

$F_D(\lambda)$ = the normalised fluorescence emission spectrum of the donor

ϵ_A = the molar extinction coefficient

λ = wavelength

Efficiency of the D-A energy transfer is a measure of the ratio of photons absorbed by the donor that are actually transferred to the acceptor and given by:

$$E = \frac{1}{1 + \left(\frac{r}{R_o}\right)^6}$$

This equation reinforces the distance dependency of FRET efficiency to the distance between the two fluorophores that make up the donor-acceptor pair (fig. 2C). The highest value E can take is 1, which denotes 100% energy transfer efficiency; however, experimentally-determined FRET values are considered to be most reliable when E takes values between 0.1 and 0.9, which is in the (more or less) linear region of the FRET efficiency curve.

Another way of determining the FRET efficiency is by taking the ratio of acceptor fluorescence to total fluorescence under donor excitation:

$$E = \frac{F_D^A}{F_D^D + F_D^A}$$

F_D^A = fluorescence intensity of A under D excitation

F_D^D = fluorescence intensity of D under D excitation

3. Single-molecule FRET (smFRET)

While providing insight into many biological systems, ensemble fluorescence measurements return an averaged FRET value that arises through observation of many molecules, exhibiting different D-A distances, at one time. Yet, it is possible to employ single-molecule FRET (smFRET; Weiss 1999), which allows not only the observation of individual molecules, but also differentiation between them in real time: so powerful is

smFRET, that it allows differentiation between molecules that exhibit static heterogeneity (such as two independent molecules that are both labelled, but on which the D-A fluorophores are at different distances) and dynamic heterogeneity (such as one molecule that undergoes a conformational change which also causes a change in the D-A pair distance) (Lackowicz 2006). This particular strength of single-molecule fluorescence is important for some of the work described in this thesis, especially for the experiments described in Chapter 3.

Besides being so powerful, the technique is also very versatile, allowing observation of freely diffusing molecules in solution (by using confocal microscopy), as well as immobilised molecules on a surface (by using total internal reflection microscopy), both of which have been used to generate the experiments described in this work.

4. Microscopy

Diffusion-based confocal microscopy was employed to observe freely diffusing molecules in solution: diluted to picomolar concentrations, fluorescently-tagged molecules diffuse via Brownian motion through a femtolitre confocal spot (created by focused light), giving rise to 'bursts' of fluorescence (Shera et al. 2000). The fluorescence emitted by the passing molecules is detected by avalanche photodiodes sensitive enough to detect single photons. Sensitivity of detection is important, as molecules can only be observed for a limited amount of time (which is equivalent to how long it takes for the observed molecule to diffuse through the confocal spot, ~1ms); however, when they are observed,

they are free of any interactions with the surface, which is one of the main advantages of using confocal microscopy (Nie et al. 1994).

When experiments are aimed at studying immobilised molecules on a surface, total internal reflection fluorescence (TIRF) microscopy can be used (Thompson et al. 1981; Schneckenburger 2005). In TIRF microscopy, illumination of the sample is achieved by focusing the incoming laser light to the back focal plane of the objective at a higher angle than the critical angle of the beam, which gives rise to an evanescent wave capable of exciting molecules that are within $\sim 100\text{nm}$ from the surface of the microscopy slide. The intensity of the wave decreases from the surface of the coverslip onto which fluorescent molecules are immobilised, up, which presents one of the advantages of TIRF microscopy, namely a high signal-to-noise ratio, as only the molecules attached to the surface are observed.

While in confocal microscopy, the sample observed is essentially one fluorescent molecule diffusing through the confocal spot, in TIRF microscopy, several molecules can be observed at the same time and for longer, which is another advantage of this method. Of course, the method also presents disadvantages – such as non-specific absorption of molecules on the surface, which leads to a higher background signal.

For a detailed description of the microscopy setups used in the work presented in this thesis, the reader is referred to the doctoral thesis written by Crawford (2011).

5. Alternating laser excitation (ALEX)

In the majority of the experiments presented in this thesis, confocal or TIRF microscopy were combined with two colour alternating laser excitation (ALEX; Kapanidis et al. 2004), which significantly improves single-molecule FRET experiments by independently probing for the presence of both the donor and the acceptor. This is achieved by using two excitation lasers, one capable of exciting the donor and the other the acceptor, which are pulsed, and hence illuminate the sample in alternation. During donor excitation, both donor (F_D^D) and acceptor (F_D^A) emissions are recorded and FRET efficiency can be calculated as shown above; during acceptor excitation, acceptor emission (F_A^A) is monitored and a second parameter can be calculated – the relative fluorophore stoichiometry, S :

$$S = \frac{F_D^D + F_D^A}{F_D^D + F_D^A + F_A^A}$$

This additional observable can be defined as the ratio of total fluorescence recorded under donor excitation to the total fluorescence observed under acceptor excitation. As in the case of FRET efficiency, S can take values between 0 and 1: $S=1$ is indicative of a population of molecules that is donor-only (no acceptors present; labelled 'D only'), while $S=0$ is the mark of an acceptor-only ('A only') population; when both a donor and an acceptor are present on the same molecule (labelled 'DA'), the stoichiometry will be ~ 0.5 . Essentially, ALEX provides a better characterisation of the system under study. The observables E and S can be plotted together on a two-dimensional E/S histogram (fig. 3), which shows how many molecules are characterised by specific values that are particular

to the states or the interactions observed. Indeed, much of the work shown in this thesis relies on the use of E/S histograms and the legend to fig. 3 explains how to interpret the data contained in such a histogram.

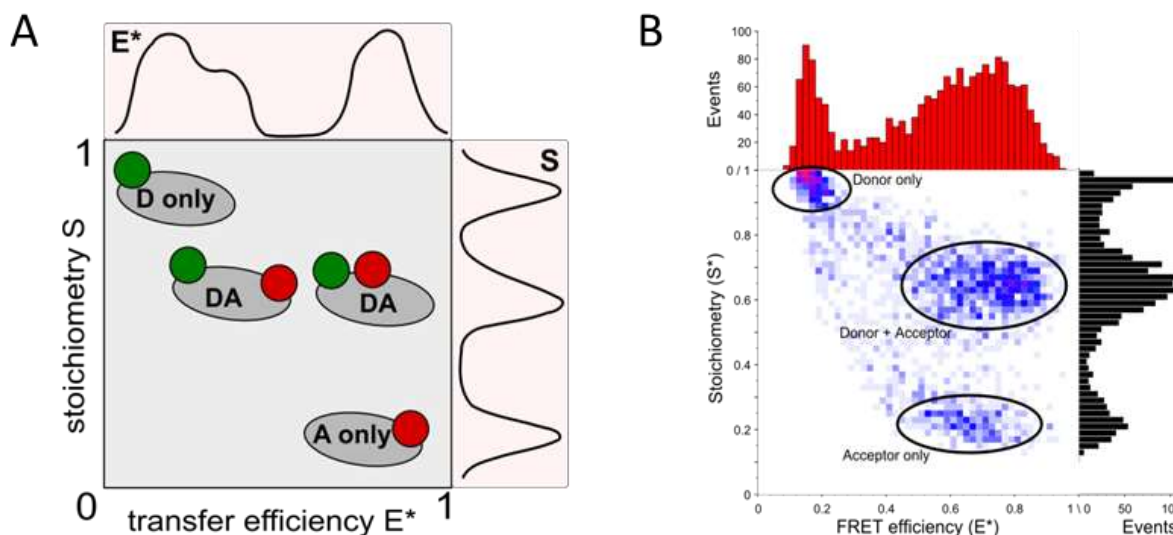


Fig. 3 – A) A schematic representation of an E/S histogram, with the E values represented on the horizontal axes and the S values on the vertical. On the histogram, donor-only molecules are shown on the top left, while acceptor-only molecules are plotted on the bottom right; as the histogram is 2-dimensional, each of these populations is also characterised by a FRET efficiency value which is meaningless, since only half of the FRET pair is present. When the molecule is labelled with both a donor and an acceptor, $S=0.5$, and the E value offers insight as to the relative distance between the two fluorophores. B) An experimentally determined E/S histogram onto which the D only, A only and DA populations are highlighted.

Generally, these histograms are plotted using values for E and S that were determined using raw photon counts, which are denoted E^* and S_{raw} . E^* is an expression of relative distance changes between the two fluorophores in the FRET pair as the photon counts taken into consideration are not corrected for the influence of the background, of the crosstalk between the donor and acceptor channels and the detection efficiencies of the dyes. However, taking into account all these factors allows for determination of accurate

FRET values which can be used in structural studies, as it is the case with the work shown in Chapter 3.

6. From apparent to accurate FRET

Resolving accurate FRET from apparent FRET (E^*) can be achieved in three steps (Hohlbein et al. 2014):

- Step 1 – correcting for the background: in confocal microscopy, impurities, inelastic Raman scattering of water molecules, as well as dark counts in the detectors contribute to background intensity which influences the photon counts detected under either donor or acceptor excitation in each of the two channels. To correct for it, this intensity is determined by taking the mean count rate in each of the channels and multiplying it with the duration of the fluorescence burst; the resulting value is then subtracted from each of the photon streams to give the following proximity ratio:

$$E_{PR}^{raw} = \frac{{}_cF_D^A}{{}_cF_D^D + {}_cF_D^A}$$

where the subscript ‘c’ indicates that background is accounted for.

- Step 2 – correcting for crosstalk, which can come from direct excitation of the acceptor by the laser used to specifically excite the donor fluorophore (in short, direct excitation – Dir) and from detected donor emission in the acceptor channel (in short, leakage – Lk). Using the donor-only population from the background-corrected E/S histogram, Lk can be determined; using the acceptor-only population from the same background-adjusted histogram, Dir can be determined.

Once these quantities are known, the following equation can be used to determine the number of photons emitted by the acceptor due to energy transfer from the donor (F^{FRET}):

$${}_cF_D^A = Lk + Dir + F^{FRET}$$

This can be calculated for each individual burst and then further used to determine the cross-talk corrected proximity ratio, E_{PR} :

$$E_{PR} = \frac{F^{FRET}}{{}_cF_D^D + F^{FRET}}$$

- Step 3 – correcting for the detection efficiency of the dyes: this is achieved by determining the correction factor γ , which is a ratio between the quantum yields of the dyes and the detection efficiency of each of the dyes in the FRET pair in the absence of the other. Different E_{PR} and S values from two samples, characterised by different FRET values, but tagged with the same FRET pair and corrected for both background and cross-talk, are plotted as $1/S$ versus E_{PR} , giving a straight line – its intercept on the y axis (Ω) and its slope (Σ) are used to find the correction factor γ :

$$\gamma = \frac{\Omega - 1}{\Omega + \Sigma - 1}$$

Ω = intercept on the y axis of the $1/S$ vs E_{PR} plot

Σ = slope of the $1/S$ vs E_{PR} plot

Finally, accurate FRET is given by:

$$E = \frac{F^{FRET}}{\gamma_c F_D^D + F^{FRET}}$$

The corrections for background and cross-talk can be similarly applied to stoichiometry.

Once E was found, molecular distances can be determined by the rearranging the FRET efficiency equation to:

$$R = R_0 \left(\left(1 - \frac{1}{E} \right) - 1 \right)^{1/6}$$

The correction from apparent to accurate FRET is used in Chapter 3, which aims to establish the structure of the influenza viral RNA promoter when bound by the polymerase.

7. Bibliography

- Benesch, R.E. & Benesch, R., 1953. Enzymatic Removal of Oxygen for Polarography and Related Methods. *Science*, 118(3068), pp.447–448.
- Clegg, R.M., 1992. Fluorescence resonance energy transfer and nucleic acids. *Methods in enzymology*, 211, pp.353–88.
- Cordes, T., Vogelsang, J. & Tinnefeld, P., 2009. On the Mechanism of Trolox as Antiblinking and Antibleaching Reagent. , (iv), pp.5018–5019.
- Crawford, R., 2011. *Single-molecule DNA Sensors and Cages For Transcription Factors In Vitro and In Vivo*. University of Oxford.
- Hohlbein, J., Craggs, T.D. & Cordes, T., 2014. Alternating-laser excitation: single-molecule FRET and beyond. *Chemical Society reviews*, 43(4), pp.1156–71.
- Kapanidis, A.N. et al., 2004. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 101(24), pp.8936–41.
- Lackowicz, J.R., 2006. *Principles of Fluorescence Spectroscopy* 3rd Editio., Springer Science & Business Media.
- Nie, S., Chiu, D.T. & Zare, R.N., 1994. Probing Individual Molecules with Confocal Fluorescence Microscopy. *Science*, 266(5187), pp.1018–1021.
- Rasnik, I., McKinney, S. a & Ha, T., 2006. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature methods*, 3(11), pp.891–3.

- Schneckenburger, H., 2005. Total internal reflection fluorescence microscopy: technical innovations and novel applications. *Current opinion in biotechnology*, 16(1), pp.13–8.
- Shera, E.B. et al., 2000. Detection of single fluorescent molecules. , 174(6), pp.553–557.
- Streyer, L. & Haugland, R.P., 1967. Energy transfer: a spectroscopic ruler. *Proceedings of the National Academy of Sciences*, 58(2), pp.719–726.
- Thompson, N.L., Burghardt, T.P. & Axelrod, D., 1981. Measuring Surface Dynamics of Biomolecules by Total Internal Reflection Fluorescence with Photobleaching Recovery or Correlation Spectroscopy. *Biophysical journal*, 33(March), pp.435–454.
- Turro, T.J., Ramamurthy, V. & Scaiano, J.C., 2010. *Modern Molecular Photochemistry of Organic Molecules*, University Science Books, Sausalito, CA.
- Weiss, S., 1999. Fluorescence Spectroscopy of Single Biomolecules. *Science*, 283(5408), pp.1676–1683.

Part A: Investigating influenza A RNA polymerase – promoter binding using fluorescence techniques

Chapter 1: Influenza RNA polymerase: expression, purification and characterisation

1.1 INTRODUCTION

The influenza A virus is a human and animal pathogen known to cause respiratory illness by infecting both the upper respiratory tract (nose and throat) and the lower respiratory tract (lungs) alike (fig. 1A offers a schematic of the virus life cycle). Most frequently, influenza is responsible for seasonal (winter) epidemics; occasionally, the virus is known to give rise to severe pandemics, such as the “Spanish flu” pandemic in 1918-1919, believed to have claimed over 50 million lives worldwide (Taubenberger & Morens 2006). Part of the *Orthomyxoviridae* family, the influenza virus is characterised by a genome consisting of eight linear, single-stranded, negative-polarity RNA segments. Each RNA segment associates with a viral RNA polymerase and several molecules of nucleoprotein (NP) in a structure known as the ribonucleoprotein (RNP; fig. 1B). While the NP protein binds single-stranded RNA, the RNA polymerase specifically binds the double-stranded structure formed by the 5′ and 3′ ends of each of the viral genes, an interaction that seems to be paramount for genome packaging. The ends of the viral RNA consist of two conserved sequences, 13 nucleotides long on the 5′ end and 12 nucleotides long on the 3′ end, which show partial, inverted complementarity: this feature of the influenza genome, allows the two ends to form a double-stranded structure known as a panhandle.

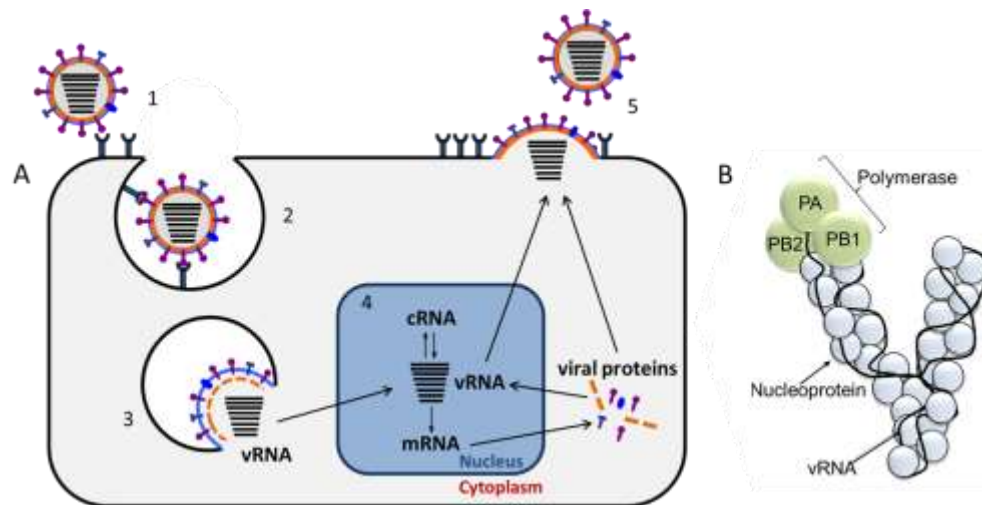


Fig. 1 – A) The virus life cycle. The first step of the viral life cycle is attachment of the virus to the exterior of the host cell (1), by recognition of specific sialic acid linkages via the viral haemagglutinin proteins found on the exterior of the virions. Once attached, the virus is internalised via endocytosis (2) and upon endosome acidification the viral RNPs are released into the cell (4); the low pH is required for fusion and uncoating of the virus (3). At this point, RNA synthesis can begin: as the virus relies on the same protein to preform both transcription and replication and on using the viral RNA (vRNA) as a template in both of these steps, it follows that these two processes cannot take place at the same time. Indeed, accumulation of viral messenger RNAs (mRNAs) in the first hours post-infection suggests that transcription is prevalent at this stage, with its levels falling sharply afterwards and the main RNA synthesis activity becoming vRNA production, via the replication intermediate complementary RNA (cRNA). The influenza vRNA is of negative polarity, while the mRNA and the synthesis intermediate cRNA are of positive polarity, with the only difference between the two being at the ends: the mRNA carries a methylated guanosine cap (and an additional 10-13 nucleotides of the host's pre-mRNA that was processed in the initial stage of transcription) at its 5' end and a poly(A) tail at its 3' end, mimicking the cellular mRNA in morphology so that it is recognised by the cell's translational machinery; the cRNA ends are fully complementary to and an exact copy of the vRNA ends. Once the virus has replicated its genetic material and produce enough viral proteins to give rise to new RNPs, they are exported from the nucleus and targeted to the apical membrane of the host cell where the viral envelope proteins have been assembled and are ready to initiate budding and release of the virus (5). B) A cartoon representation of the ribonucleoprotein. (Diagrams adapted from Fields 2007)

Besides being an RNA binding protein, the influenza A polymerase is a versatile RNA-dependent RNA polymerase capable of functioning as both a transcriptase (producing mRNA) and a replicase (producing vRNA in a two-step process). Each of the three subunits associating to form the polymerase, a protein complex weighing ~250kDa, plays a role in both of these two processes.

The best characterised polymerase subunit is PB1 (polymerase basic protein 1, MW = 86.5 kDa), harbouring the catalytic activity of the polymerase complex (Braam et al. 1983); indeed, this subunit contains conserved motifs commonly found in other RNA-dependent RNA polymerases (Argos 1988; Poch et al. 1989; Biswas & Nayak 1994). The subunit on its own is capable of binding both the vRNA promoter (González & Ortín 1999a; González & Ortín 1999b) and the cRNA promoter (González & Ortín 1999b) in a specific manner.

The PB2 (polymerase basic protein 2, MW = 85.7 kDa) subunit recognises and binds cellular pre-mRNAs in transcription (Li et al. 2001; Guilligay et al. 2008; Honda et al. 1999; Fechter et al. 2003). The function of this subunit in replication is still largely unknown, but studies have shown its presence to be a requirement for *de novo* initiation (Perales & Ortín 1997; Lee et al. 2002; Gastaminza et al. 2003; Deng et al. 2005).

The PA (polymerase acidic protein, MW = 84.2 kDa) subunit was initially characterised as a serine protease (Sanz-Ezquerro et al. 1995; Sanz-Ezquerro et al. 1996; Hara et al. 2001), but this activity of the protein is not critical for its function as a polymerase (Fodor et al. 2002), suggesting that the main function of the protein lies elsewhere. Later, it was determined that, in transcription, the PA subunit acts as an endonuclease involved in cap-snatching (Fodor et al. 2002; Dias et al. 2009; Yuan et al. 2009); in replication, its function

has not yet been elucidated, but some studies suggest it is required for binding of the polymerase complex to the cRNA promoter (Maier et al. 2008) and also for cRNA synthesis (Perales et al. 2000). Furthermore, PA seems to play an important role in assembly and release of virions (Regan et al. 2006).

The requirement for and the function of the PB2 and PA subunits in transcription demonstrate how the virus evolved to take advantage of the hosts' translation machinery. Specifically, these subunits are implicated in 'disguising' the viral mRNA in such a way that the cell host can treat it as if it were a genuine eukaryotic mRNA. Eukaryotic mRNAs to be translated carry a 7-methylguanosine at their 5' end (known as a 5' cap) and an adenosine repeat (poly(A) tail) at their 3' end, features that have to be replicated by the virus to be able to use the cell's intrinsic translational machinery. As described above, PB2 is able to recognise and bind a capped cellular mRNA; the cap, alongside a 10-13nt long RNA fragment, is cleaved by the PA subunit and transferred to the PB1 polymerase subunit which can use it to start synthesis of the mRNA. The poly(A) tail is generated when the polymerase tries to copy a 5-8 U-rich sequence situated downstream of the 5' end of the vRNA; due to steric hindrance caused by the fact that the 5' end is closely bound to the polymerase, the protein stutters, generating an A-rich RNA sequence .

The polymerase subunits can associate to form the trimeric complex even in the absence of the nucleoprotein (NP) and RNA (Akkina et al. 1987). The complex assembles in a linear fashion, with PB1 acting as a central scaffold – PB1's N-terminal end is thought to interact with PA's C-terminus, while its C-terminus is believed to bind to the PB2 N-terminal

sequence (Toyoda et al. 1996; Honda et al. 2002). This ordered interaction is supported by crystal structures of the interacting terminal sequences of the proteins in question (He et al. 2008; Obayashi et al. 2008; Sugiyama et al. 2009).

PA-PB1 and PB1-PB2 dimers were detected alongside the heterotrimeric complex; surprisingly, these partial polymerase complexes harboured specialised functions – PA-PB1 being functionally analogous to the RNA polymerase complex in replication and PB1-PB2, in transcription (Honda et al. 2002). On the contrary, it was believed that PA and PB2 do not directly interact; yet, recently, by employing a fluorescence-based assay, a direct observation of this protein pair was recorded (Hemerka et al. 2009).

Intriguingly, studies have shown that the origin of these two subunits is important for efficient virus replication, suggesting a second criterion for the success of cross-species infection, besides the type of sialic acid linkage recognised by haemagglutinin (HA). In fact, in all viral subtypes shown to have caused pandemics in the last century – H1N1, responsible for the ‘Spanish flu’ in 1918; H2N2, the ‘Asian flu’ in 1957; H3N2, the ‘Russian flu’ in 1968 – have had their RNA polymerase carrying the PB2 and PA subunits from the same virus type. Recently, it was shown that co-incorporation of these two subunits from the same strain might be necessary for generation of viable reassortment viruses (Hara et al. 2013).

In this chapter, I describe my efforts to express and purify H1N1 RNAP, both unlabelled and tagged with either a GFP or an mOrange2 fluorescent protein, using an expression/purification system relying on mammalian cells, as well as to characterise and quantitate the resulting protein preparations. Due to the drawbacks of this system

discussed at the end of this chapter, a baculovirus expression system coupled with an affinity purification step, followed by a size-exclusion one, was alternatively employed.

1.2 MATERIALS AND METHODS

1.2.1 Cloning of mOrange2 in the PB2—TAP pcDNA vector

The mOrange2 pcDNA vector (kindly provided by Roger Tsien, University of California - San Diego) was amplified by PCR; the amplification reaction was carried out using 10ng of mOrange2 plasmid, 500nM each of forward and reverse primers, 200μM of each dATP, dCTP, dGTP and dTTP, 1.5U Pfu polymerase (Stratagene) in 1x Pfu reaction buffer, and incubated in a thermocycler at 95°C for 30 seconds, 57°C for 30 seconds and 72°C for 30 seconds, for a total of 24 cycles, followed by an extra 2 minutes at 72°C.

The primers used had the following sequence:

(Forward primer): 5' CTACGCGGCCGCATGGTGAGCAAGG 3'

(Reverse primer): 5' CTACGCGGCCGCCTTGTACAGCTCGTCCATGC 3'

Post amplification, the mOrange2 PCR product was mixed with 5x Green GoTaq Flexi Buffer (Promega) and loaded onto a 1% agarose gel in 1 x TAE buffer (90mM Tris, 90mM acetic acid, 2.5mM EDTA), which was run in 1x TAE with ethidium bromide at 120 V for 40 minutes. The band was excised and gel purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions.

The original vector (mOrange2 pcDNA, amplified) and the destination vector (PB2-GFP-TAP pcDNA) were independently digested with 1U *NotI* restriction enzyme (New England Biolabs) in the supplied NEB buffer with 0.1μg BSA in a total reaction volume of 40μl

(mOrange2 plasmid) or 20ul (GFP plasmid); the digestion reaction was incubated at 37°C for 1 hour.

Post digestion, the PB2-GFP-TAP pcDNA sample was treated with 1µl calf intestinal alkaline phosphatase (CIP, New England Biolabs) in 1x CIP buffer for 1 hour at 37°C.

Both digestion products were analysed on a 1% agarose gel; the bands of interest corresponding to the mOrange2 gene and the PB2-TAP vector were excised and purified as above.

The purified mOrange2 insert and PB2-TAP vector were ligated at a molar ratio of 1:1 (insert:vector) using 40U T4 DNA ligase in 1 x ligation buffer in a total reaction volume of 2µl. The whole ligation reaction was used for transformation of 50µl electro-competent *E.coli* NM554 cells, via electroporation. The cells were allowed to recover in 120ul S.O.C. medium (Invitrogen) with shaking at 37°C for 1 hour. Finally, the bacteria were spread onto agar plates containing 100µg/ml ampicillin and incubated overnight at 37°C.

1.2.2 Reverse plating PCR (to determine the orientation of the mOrange2 gene in the PB2-mOrange2-TAP construct)

From the plate cultured above, 6 colonies exhibiting the 'small' morphology and 6 colonies exhibiting the 'big' morphology were picked and streaked onto a fresh agar+100µg/ml ampicillin plate (divided into squares numbered 1-13, each corresponding to one PCR reaction) and then individually added to a PCR reaction containing 1µM each of forward and reverse primers, 100µM of each dNTP, 2U Taq polymerase (Promega) in 1x Taq reaction buffer, and incubated in a thermocycler at 95°C for 30 seconds, 57°C for 30 seconds and 72°C for 1 minute, for a total of 30 cycles, followed by an extra 5 minutes at

72°C; samples were further analysed on a 1% agarose gel (as previously described), where the mobility of the products allow us to determine the length of the PCR product and hence the direction of insertion.

The primers used had the following sequence:

(Forward primer): 5' CTACGCGGCCGCATGGTGAGCAAGG 3'

(Reverse primer): 5' GATCAGCCTCGACTGTGCCTTCTAGTTGCCA 3'

1.2.3 Plasmid production (PB2-mOrange2-TAP pcDNA vector)

The colonies identified as carrying the mOrange2 gene in the correct orientation were picked and used to inoculate 2ml cultures of LB medium supplemented with 100µg/ml ampicillin and incubated with shaking at 37°C overnight, after which they were pelleted and the plasmid DNA extracted using a QIAprep spin miniprep kit (Qiagen) according to the manufacturer's instructions. Purified plasmid DNA was sequenced by Geneservice, Oxford (Source Bioscience Plc).

Alternatively, 100ml cultures of LB medium supplemented with 100µg/ml ampicillin were inoculated with 1ml overnight culture and incubated overnight at 37°C, with shaking, after which the plasmid DNA was extracted using a QIAfilter maxiprep kit (Qiagen) as instructed by the manufacturer.

1.2.4 H1N1 RNA polymerase expression and purification

For expression and reconstitution of the H1N1 (A/WSN/33) RNAP, the following plasmids were co-expressed in human embryonic kidney cells (HEK 293T), grown and maintained in Dulbecco's Modified Eagle's Medium: pcDNA-PA, pcDNA-PB1 and pcDNA-PB2-TAP (for

unlabelled protein complex), pcDNA-PA, pcDNA-PB1 and pcDNA-PB2-GFP-TAP (for the GFP-fused protein) or pcDNA-PA, pcDNA-PB1 and pcDNA-PB2-mOrange2-TAP (for the mOrange2-fused variant).

293T cells (cultured in 10cm Petri dishes, 8.8×10^6 cells/dish when confluent) were transfected with 5 μ g of the relevant plasmids using Lipofectamine2000 (Invitrogen) and then incubated at 37°C for 24-48 hours, washed in 1x Phosphate Buffer Saline (PBS, Sigma), harvested and re-suspended in lysis buffer (50mM HEPES-NaOH (Sigma) pH 8.0, 200mM NaCl, 25% glycerol, 0.5% tert-octylphenoxypolyoxyethylene ethanol (IGEPAL, Sigma) or nonylphenoxypolyethoxyethanol (Tergitol-type NP40, Sigma), 1mM β -mercaptoethanol, 1x complete EDTA-free protease inhibitor (Roche)). Samples were incubated on ice for 30 minutes and centrifuged for 5 minutes (4°C and 13000 rpm); the cell lysate was centrifuged and the supernatant added to Immunoglobulin G (IgG) Sepharose (GE Healthcare). The lysate-IgG Sepharose mix was incubated at 4°C for 2-2.5 hours with rocking. Further, the sample was centrifuged for 1 minute (1000 rpm) and the pellet was washed twice with wash buffer (5 minutes, 4°C, with rocking in 10mM HEPES-NaOH pH 8.0, 150mM NaCl, 10% glycerol, 0.1% IGEPAL CA-630 or Tergitol-type NP40, 1x complete EDTA-free protease inhibitor), re-suspended in cleavage buffer (wash buffer supplemented with 1mM dithiothreitol (DTT)) and further centrifuged (10 minutes, 4°C, 2000 rpm). Finally, the sample was incubated with elution buffer (tobacco etch virus protease(AcTEV, Invitrogen) in cleavage buffer) at 4°C, with rocking for 2 hours, and then briefly spun down (10 seconds, 2000 rpm) to collect the supernatant containing the purified protein, which was stored in 30% glycerol at -80°C.

1.2.5 H3N2 RNA polymerase expression and purification

H3N2 (A/NT/60/68) RNAP with a PB2-TAP subunit was expressed using the Multibac system in Sf9 insect cells in suspension. Cells (density: 1.2×10^6 cells/ml, in a 1l culture), at were infected with recombinant baculovirus (MOI=0.5–1.0), incubated at 27°C for 72 hours, harvested by centrifugation and lysed by sonication in lysis buffer (50mM HEPES-NaOH pH 7.5, 500mM NaCl, 10% glycerol, 0.05% octylthioglucoside, 1× complete EDTA-free protease inhibitor, 100µg/mL RNaseA); the suspension was centrifuged and the supernatant was added to IgG Sepharose beads (~2ml bead slurry per liter of original culture) and incubated at 4 °C for 3-4 hours. After incubation, the beads were washed with 50mM HEPES-NaOH pH 7.5, 500mM NaCl, 10% glycerol and 0.05% octylthioglucoside and the polymerase was released by incubation with AcTEV protease (150 U/ml of IgG Sepharose) in wash buffer supplemented with 2.5mM reduced glutathione at 4 °C, overnight. Finally, the polymerase was purified on a HiLoad™ 16/60 Superdex S200 column (GE Healthcare) at 4°C, using 50mM HEPES-NaOH pH 7.5, 500mM NaCl, 10% glycerol, 1mM MgCl₂, and 0.1mM MnCl₂.

Protein concentration was assessed via NanoDrop and was found to be 4.5mg/ml; taking into account the molecular weight of the protein, concentration was determined to be 18µM.

1.2.6 Assessing the purity of purified protein samples

The purity of the protein preparation was assessed by electrophoresis on a discontinuous polyacrylamide gel system consisting of a 4% stacking and an 8% resolving SDS-PAGE gel (29:1 acrylamide:bis-acrylamide). The protein samples were mixed with 1x sample volume

of loading buffer (100mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 3mM β -mercaptoethanol and 0.04% bromophenol blue), heated at 95°C for 5 minutes and loaded on the gel. The gel was run in Tris-glycine-SDS buffer (25mM Tris, 250mM glycine and 0.1% SDS) for 1h at 200V.

The protein bands were visualised using the SilverXpress (Invitrogen) silver staining kit according to the manufacturer's instructions for the 293T-expressed protein or by staining with Coomassie Brilliant Blue (BioRad) for 1 hour at room temperature (with shaking), followed by destaining in H₂O overnight, for the Sf9-expressed protein.

1.2.7 Activity assay (*in vitro* transcription)

In order to determine whether the purified polymerase is active *in vitro*, the capacity of the protein to transcribe RNA in an ApG-assay (Brownlee & Sharps, 2002; Bouloy, 1978; Pritlove, 1998) was tested: 1.5ul RNAP was incubated with 10nM double-stranded vRNA promoter (final concentration) in 10mM MgCl₂ and 2mM DTT supplemented with 1.87U RNasin (40U/ μ l, Promega), 1mM ApG and 50uM:25uM:25uM ATP:CTP:UTP mix, in the presence of 0.3 μ Ci/ μ l α -³²P GTP (10 μ Ci/ μ l, Perkin Elmer) and incubated for 2h at 30°C. The transcription reaction was stopped by addition of 10ul loading dye (80% formamide, 10mM EDTA, 0.04% bromophenol blue and 0.04% xylene cyanol FF) and denatured by heating at 95°C for 5 minutes, then electrophoresed on a 6M urea/25% polyacrylamide (19:1 acrylamide:bis-acrylamide) sequencing gel in 1x TBE (89mM Tris-borate, 2mM EDTA pH 8.3; Sigma Aldrich) for 3h at 1200V. The products of the reaction were visualised by autoradiography.

1.2.8 Binding assay (single-molecule fluorescence spectroscopy)

The measurement was performed on a custom-built confocal microscope allowing alternating-laser excitation of donor and acceptor fluorophores, at average excitation intensities of 250 μ W at 532nm and 60 μ W at 635nm. A dsRNA construct, labelled with ATTO647N and Cy3B diluted in PBS (supplemented with 15mM MgCl₂) to a final concentration of 100pM was incubated either alone or with 30 μ l of purified RNA polymerase in storage buffer, for 10 minutes at room temperature, after which data was acquired. Alternatively, a dsRNA construct, labelled with ATTO647N, was diluted in the same manner and incubated with 30 μ l of purified mOrange2-RNAP.

A custom software package was used to assign fluorescence photons to either donor or acceptor-based excitation depending on their arrival time (Seneca – Kristofer Gryte, University of Oxford). Fluorophore stoichiometry and the FRET efficiency (E^*) were calculated for each burst above the minDA=30 threshold (where the threshold ensures that the analysis only takes into account those bursts in which the minimum number of photons emitted by the acceptor after donor excitation – and hence energy transfer – is 30), yielding a two-dimensional E^*/S histogram; one-dimensional FRET distributions characteristic for donor-acceptor species were obtained by applying a $0.4 < S < 0.8$ threshold, while the mean E^* value was calculate by fitting the FRET distributions with a Gaussian function.

1.2.9 Protein quantitation (SyproRuby stain)

Standard samples were prepared by dilution from 2mg/ml BSA (Sigma) in H₂O and mixed with 1x sample volume of protein loading buffer; the RNAP sample was prepared as previously described. Samples were heated for 10 minutes at 95°C and immediately loaded a 4% stacking/8% separating gel (as previously described), which was ran at 150V until the electrophoretic front reached the bottom of the plates. The gel was stained with SyproRuby following the manufacturer's instructions and visualised on a BioRad Molecular Imager PharosFX system equipped with a 532nm laser (excitation intensity: 25mW). Densitometry was employed for quantitating the protein vs the standards by analysing the image acquired and using the standard 'gel analyser' module in the ImageJ software (Rasband, NIH, 1997-2014).

1.2.10 Protein quantitation (ensemble fluorescence)

A 425nm-600nm fluorescence emission scan was performed on an L-format fluorescence spectrophotometer (PTI), upon excitation at 488nm. Standards were prepared by diluting 1mg/ml GFP (kindly provided by Anthony Walsh, University of Oxford) in PBS (supplemented with 100µg/ml BSA) to a final concentration of 10nM, 50nM and 100nM; PBS+100µg/ml BSA was used to determine the fluorescence baseline. The peak area of the GFP peaks of interest (deemed to be between 488nm and 524nm) was calculated using the standard 'integrate' function available in the PTI standard software supplied with the machine.

1.2.11 Protein quantitation (single-molecule fluorescence)

The measurement was performed on a custom-built confocal microscope as described previously. A dsDNA labelled with ATTO647N and Cy3B, showing a mean FRET value $E=0.47$ (construct T1B17, previously characterised in Holden et al. 2010) was diluted to a final concentration of 100pM and 25pM in PBS and used as a reference to which the sample of unknown protein concentration was compared; mOrange2-labelled RNA polymerase in storage buffer was diluted 100x, 200x and 400x in PBS and visualized. In a 400x dilution, the sample was diluted enough to be observed at the single-molecule level. Custom written software was used to register and evaluate the detected signal, as described previously. For both the reference and the unknown concentration sample, the same number of fluorescence events were recorded and the number of molecules detected in green and the red channel, respectively, for both were obtained, these values were then used to determine the concentration of the unknown sample as follows:

- $[\text{number of molecules detected in the G channel}]_{\text{sample}} : [\text{number of molecules detected in the G channel}]_{\text{reference}} = x$
- $[\text{number of molecules detected in the R channel}]_{\text{sample}} : [\text{number of molecules detected in the R channel}]_{\text{reference}} = y$
- the final concentration of the unknown sample was calculated using the formula:

$$conc = \frac{x+y}{2} D_f$$

D_f = dilution factor

1.3 RESULTS

1.3.1 Expression and purification of unlabelled and GFP-tagged H1N1 RNA polymerase

Expression of the influenza RNA polymerase was achieved by transfection of mammalian cells (HEK293T) with pcDNA vectors carrying segments corresponding to the PA, PB1 and PB2 polymerase subunits of the A/WSN/33 (H1N1) virus isolate (for the native polymerase complex) or PA, PB1 and PB2-GFP (for the GFP-tagged polymerase), as described by Deng et al. 2005; Pleschka et al. 1996 and in the *Materials and methods*, section 1.2.4. Once expressed, the native RNA polymerase was purified using the tandem affinity purification method (Rigaut et al. 1999; Puig et al. 2001 and *Materials and methods* section 1.2.4).

In brief, this purification strategy relies on fusing a TAP tag to the protein of interest; such a tag consists of three modules, whose order dictates the succession of purification steps: the immunoglobulin-G binding protein A module (used in the first affinity step of the purification), the tobacco etch virus (TEV) protease specific cleavage site and a calmodulin binding protein (CBP) module (used in the second affinity step of the purification procedure). Since the influenza polymerase heterotrimer is reconstituted as it is expressed in cell culture, rather than post-purification, there is no need to have the TAP tag fused to all the three subunits of the complex; rather, it is fused to the PB2 subunit, where it was shown not to interfere formation and activity of the heterotrimer (E. Fodor, personal communication). In our specific protein preparation, the second affinity step is omitted, as the purity of the product after the TEV cleavage is higher than 95% (as per

SDS-PAGE analysis, fig. 1B, 'unlabelled RNAP' lane). An identical approach was used for purification of the GFP-tagged RNA polymerase, which showed similar purity as the native protein preparation (fig. 1B, 'GFP-RNAP' lane).

Both these RNA polymerase preparations were tested for activity (as described in the *Materials and methods* section): compared to the unlabelled RNAP, the GFP-labelled polymerase is less active (fig. 1C, compare 'unlabelled RNAP' lane with the 'GFP-RNAP' lane).

1.3.2 Generating the PB2- mOrange2-TAP plasmid

As the RNA polymerase purified here is to be used for single-molecule assays, it is important to ensure that the label with which the protein is tagged is compatible with our available microscopy set-up. With a beta barrel structure, the GFP protein emits green fluorescence (emission peak = 509nm) when it is excited with blue light (excitation peak = 488nm); however, this is not optimal for our instruments, which are equipped with a green laser (532nm), rather than a blue one (445 nm). Given this, a more suitable fluorescent protein for our application is mOrange2, with an excitation maximum at 549nm and an emission maximum at 565nm.

As both the pcDNA-PB2-GFP-TAP plasmid and the pcDNA-mOrange2 plasmid carry *NotI* sites, this restriction enzyme was employed to independently excise the GFP gene from the PB2-encoding vector and the mOrange2 gene from the pcDNA vector in which it was originally supplied. Post-excision, to prevent the re-circularization of the pcDNA-PB2-TAP vector, the DNA was dephosphorylated using calf intestinal phosphatase. Next, the mOrange2 gene and the PB2-TAP vector were gel purified and ligated. After ligation, the

PB2-mOrange2-TAP vector was used to transform electro-competent *E.coli* NM554 cells; the plated cells, upon growing overnight, produced colonies with two different morphologies, indicating that the mOrange2 gene might have inserted both in the correct and the inverse orientation. Alternatively, the morphology could also be indicative of cells with the desired construct, as well as cells carrying a PB2-TAP plasmid that went through re-ligation before insertion of fluorescent protein gene.

To determine which of the morphologies is characteristic of cells carrying the pcDNA vector with the mOrange2 in the desired orientation, a reverse plating PCR approach was adopted (Section 1.2.2 in *Materials and methods*): using a forward primer designed for the mOrange2 gene and a reverse primer specific for the vector, DNA from several colonies showing the same morphology was amplified and analysed on an agarose gel, allowing determination of the length of the PCR product and hence, whether the mOrange2 gene is present and, if it is the case, its direction of insertion.

Some of the tested colonies (those that initially showed the small morphology) returned positive PCR results for the PB2-mOrange2-TAP construct of the correct length; as further confirmation of this, the plasmid was sequenced. Finally, larger quantities of the vector with the correct orientation insert were produced and purified, to be used further for transfection in mammalian cells.

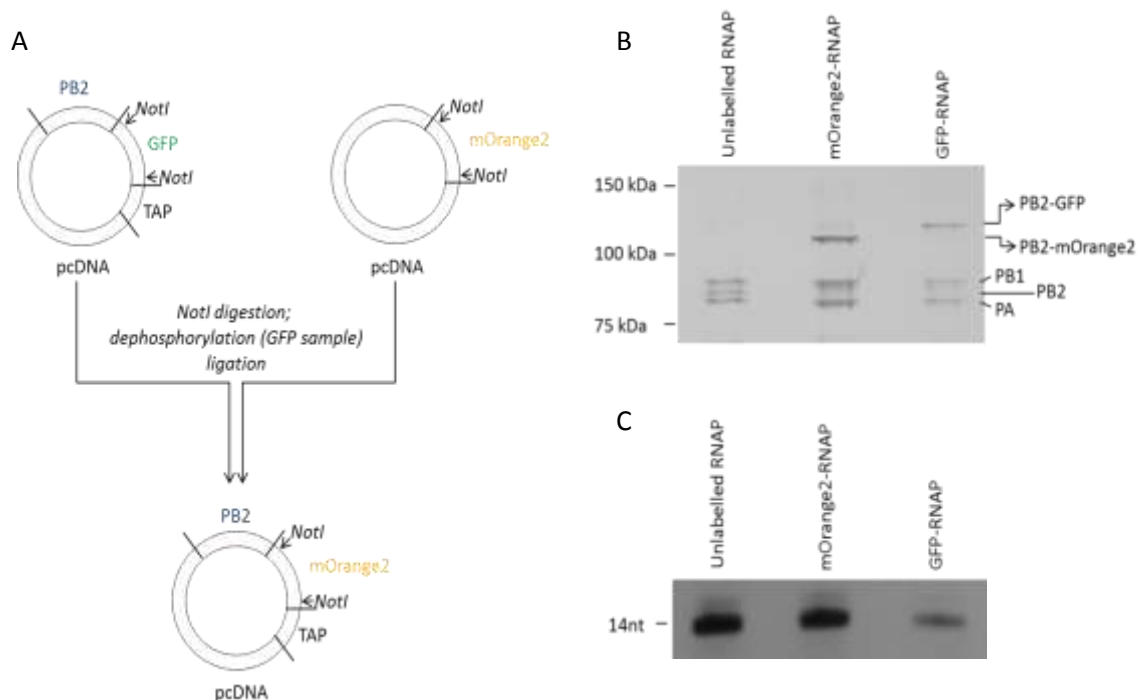


Fig. 2 – A) Cloning strategy for generating the mOrange2-fused PB2 pcDNA vector. B) Silver stained SDS-PAGE analysis of unlabelled and mOrange2- or GFP-labelled RNA polymerase complexes, purified using the TAP method. C) Radiolabelled products generated in an in vitro transcription assay using the purified polymerase complexes shown in fig. 1B.

1.3.3 Expression, purification and functional characterisation of the mOrange2-labelled H1N1 RNA polymerase

Having managed to successfully clone the mOrange2 fluorescent protein gene in the plasmid carrying the PB2-TAP, the next step was to co-express this subunit alongside the other two polymerase subunits, PB1 and PA and then purify the complex using the same TAP-affinity purification strategy as for the native RNA polymerase. Checking our protein sample on an SDS-PAGE gel showed that it is equally pure as the other two samples discussed previously (fig. 2B, 'mOrange2-RNAP' versus the other two lanes shown).

Testing activity in an *in vitro* transcription assay, showed that all the protein samples under study (non-labelled RNAP, mOrange2-tagged RNAP and GFP-labelled RNAP) are active and capable of producing RNA (fig. 2C); it is not possible to compare the activity of the samples directly as their concentrations were not known at this point (determining the protein concentrations of these samples will be discussed in detail in section 1.3.4 of this chapter).

Finally, the mOrange2-tagged protein was used in an experiment aiming to investigate the binding of the protein to the dsRNA influenza promoter. In this experiment, I first investigated an ATTO647N/Cy3b labelled, double-stranded promoter on its own; based on the distance between the dyes, the mean FRET value calculated for the construct was $E \sim 0.5$, with which the observed experimental value agrees well (fig. 3A). Next, I used the unlabelled RNAP to test binding, expecting to see an increase in FRET as the distance between the dyes shortens due to the promoter assuming a corkscrew structure (fig. 3; Flick & Hobom 1999); this was not the case and the apparent FRET value observed remained $E \sim 0.5$ (fig. 3B). Finally, using the mOrange2-tagged polymerase, I repeated the experiment, using a singly-labelled ATTO647N dsRNA promoter; as the mOrange2 is attached to the PB2 subunit, known to be specifically involved in binding the 3' end of the promoter (Jung & Brownlee 2006), the hope is that these two fluorescent molecules would be in close proximity and would interact as a FRET pair (in which the mOrange2 would act as a donor and ATTO647N as an acceptor molecule). In this case, no FRET was observed between the two fluorescent species (fig. 3C); moreover, judging from the stoichiometry of the observed species, we can clearly distinguish a donor-only population

(corresponding to the mOrange2 RNAP, $S > 0.8$) and an acceptor-only population (the RNA, $S < 0.4$).

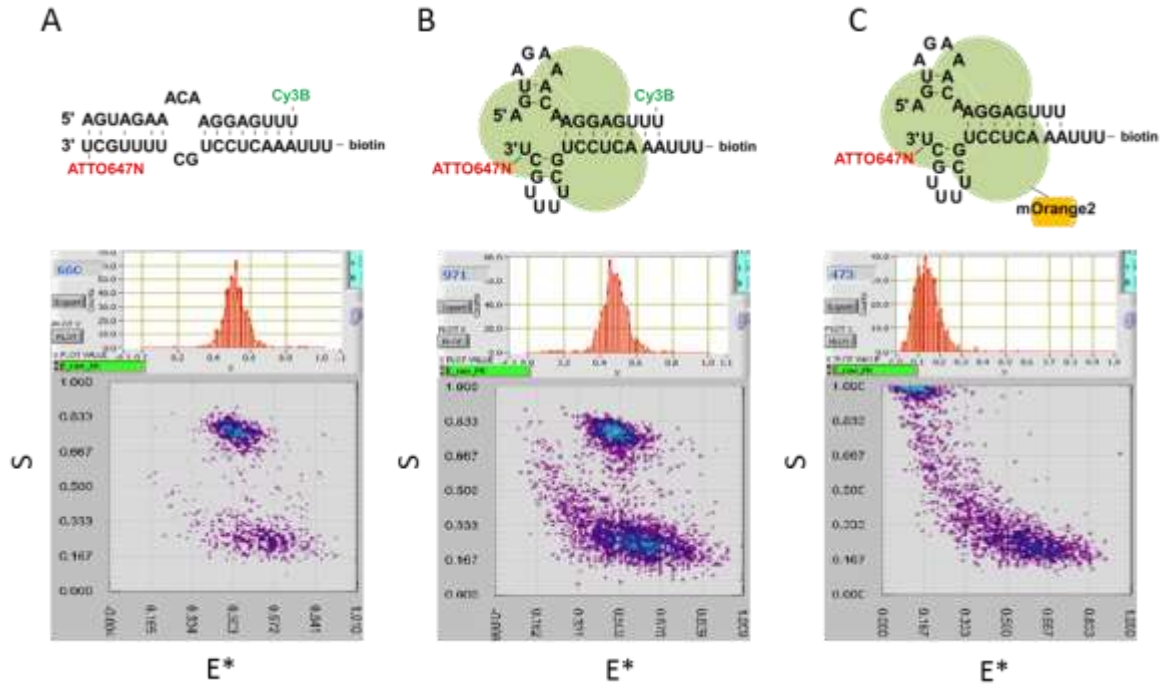


Fig. 3 – A) Investigating the dsRNA influenza promoter, labelled with ATTO647N (on the 3' end) and Cy3B (on the 5' end), in solution using confocal microscopy: the observed FRET efficiency on the E/S histogram for this construct is $E^* \sim 0.5$. B) Investigating the same promoter as in A), but in the presence of unlabelled RNAP: the observed apparent FRET is still $E^* \sim 0.5$, suggesting that the polymerase does not bind to the promoter. A large proportion of acceptor-only species is also detectable ($S < 0.4$). C) Investigating the binding of the mOrange2-RNAP to the dsRNA promoter, labelled with ATTO647N on its 3' end. The two-dimensional E^*/S histogram indicates that the species present are mostly donor-only ($S > 0.8$) and acceptor-only ($S < 0.4$), indicating that no or very few donor-acceptor species are present. For all A-C, the top panel of the figure indicate what is expected to occur when the polymerase binds the promoter, and not what is observed.

Intrigued by these results, which, on the one hand, suggest that the polymerase is active in transcription and implicitly, binds the RNA (fig. 2C) and those which show that the polymerase is not able to bind to the RNA (fig. 3B&C), I decided to determine the

concentration of the protein preparation, believing that the low concentration of polymerase might explain these results.

1.3.4 Protein concentration determination

Routinely, the protein purified as described in this chapter is used for qualitative studies of the RNA polymerase interacting with other viral or host proteins; the first step of this study, though, is to investigate the polymerase-vRNA promoter interaction from a quantitative point of view. Hence, it was necessary to determine the concentration of purified polymerase obtained at the end of the purification procedure.

In this sense, I employed several approaches:

- A classical protein quantitation assay, such as the Bradford assay (Bradford 1976; Kruger 1994), did not yield any conclusive results (not shown). This was an initial indication that the protein concentration in the sample is quite low, possibly below the minimal detection limit of this assay (given as 1µg/ml).
- A quantitative, fluorescent protein stain, SyproRuby (Simpson 2006), which is able to detect PAGE-separated proteins in amounts as low as 0.25ng (fig. 4). The unlabelled RNA polymerase used in this assay had the TAP-tag attached to the PA protein; this has was not considered to be a problem, as the yield of protein expression and purification was expected to be similar regardless of whether the TAP tag is on the PA or the PB2 protein (experimental observation from the Fodor lab). Using image analysis software (ImageJ), the intensity of the protein bands was determined (for both the BSA standards and the RNA polymerase sample) and

the concentration of the purified, unlabelled RNAP sample was determined to be 13nM (the analysis was run for the PB1 subunit).

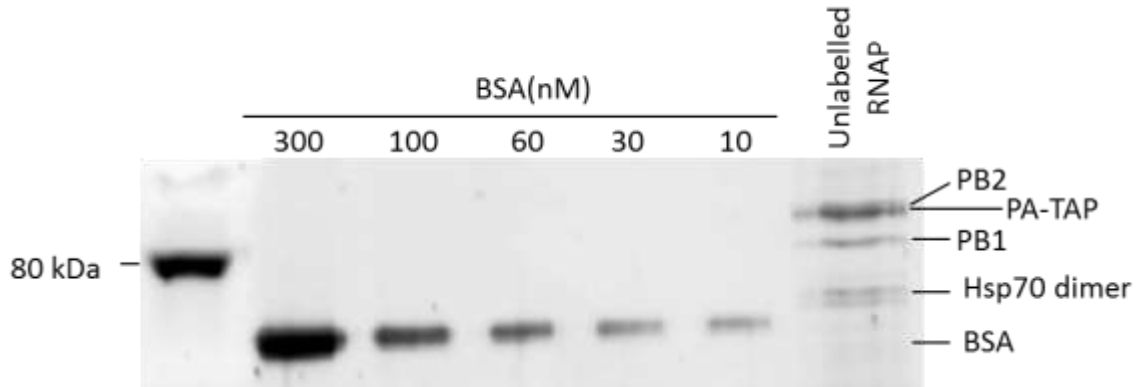


Fig. 4 – Protein concentration determination using the quantitative gel stain SyproRuby: known concentrations of BSA protein were run on the gel alongside our unlabelled RNA polymerase sample. Using the BSA as a standard, I determined that the concentration of the protein in my sample is 13nM.

- An ensemble fluorescence-based protein quantitation assay – detecting the fluorescence of the GFP-tagged protein and comparing it against a standard curve (fig. 5). A fluorescence scan at 425nm-600nm upon excitation at 488nm, recorded two peaks: the one centred at 450nm is characteristic for the PBS buffer (supplemented with 100µg/ml BSA), while the one centred at around 510nm is the GFP emission peak (fig. 5A). The GFP emission peaks (equating to the total emission from the GFP molecule within the 488-524nm interval) corresponding to known concentrations of the protein, together with the ones for the unknown concentration sample, were selected and had their area quantitated – this value plotted against protein concentration gave a standard curve, which was used to determine the concentration of two elution samples resulting from the last step of the protein purification procedure: the GFP-RNAP E2 sample (~47nM) and of the

GFP-RNAP E3 sample (~65nM) (fig. 5B). These two samples were chosen as they had the strongest colouration among all of the elution fractions collected, indicating that they are most likely to contain the GFP-tagged protein of interest.

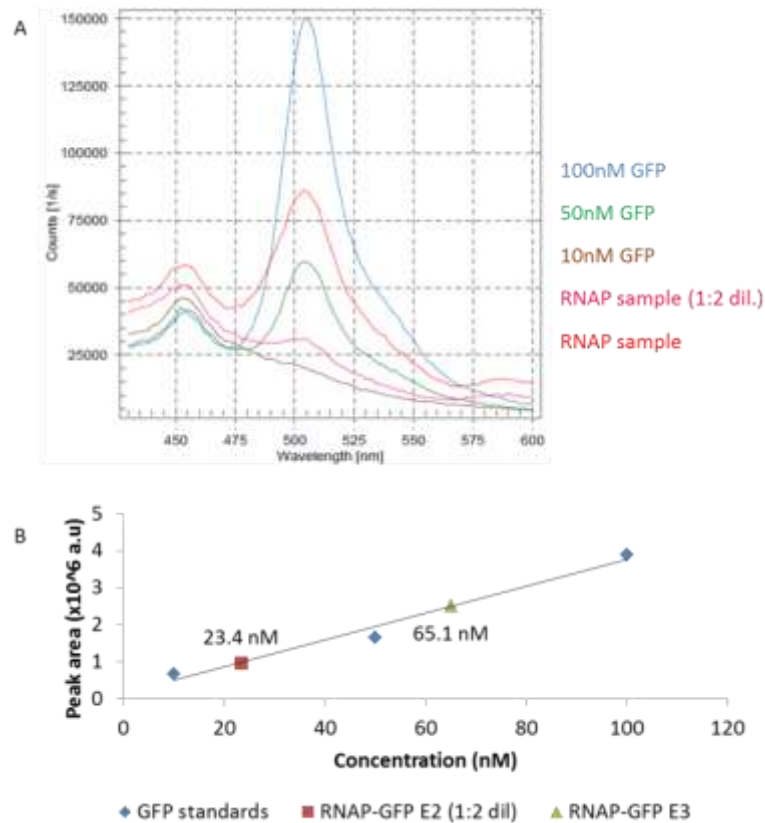


Fig. 5 – A) Emission peaks of known concentrations of GFP and of GFP-tagged RNA polymerase (around 500nm wavelength mark); the peak area was integrated and plotted against the concentration (B), allowing for determination of the protein concentration in a sample that is a 1:2 dilution of the elution sample 2 (from the purification procedure described above) – red square and the original elution sample 3 – green triangle. The values determined for the concentration of each sample are given on the graph: 23.4nM (1:2 dilution of E2) and 65.1nM (E3).

- A single-molecule fluorescence-based protein quantitation assay – detecting the fluorescence of the mOrange2-labelled protein and comparing it with standard of known concentration using a custom built confocal microscope. The fluorescence

of a dsDNA standard (FRET efficiency = 0.47) at different concentration was recorded (fig. 6A shows a 25pM dilution in PBS); alternatively, fluorescence of several dilutions of mOrange2-RNAP was recorded (fig. 6B shows a 1:400 dilution). The two sets of data were compared and the 1:400 mOrange2-RNAP dilution was approximated as 50pM; considering the dilution factor, the mOrange2-RNAP sample was calculated to be ~20nM.

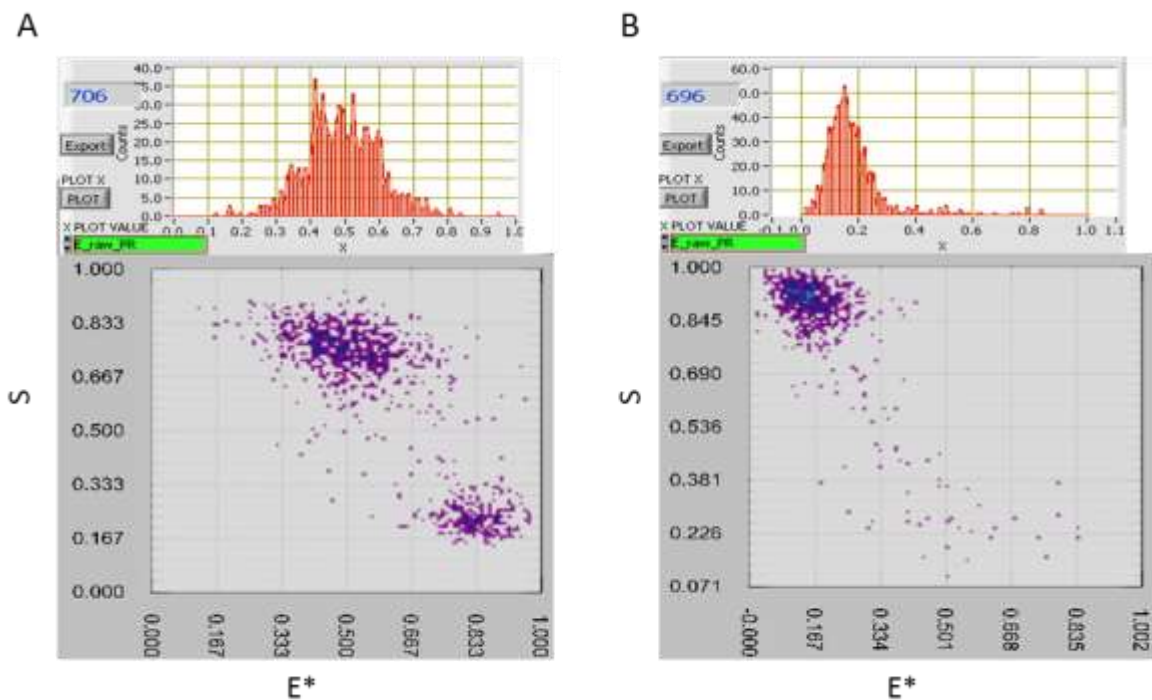


Fig. 6 – A) Histogram of 25pM of doubly-labelled dsDNA construct (18bp between the dyes, expected FRET efficiency: 0.5; number of events detected ~700). B) Histogram of an unknown concentration sample of mOrange2-RNAP (1:400 dilution; number of events detected ~700).

1.3.5 Characterisation of the H3N2 RNA polymerase

Given the low protein concentration determined for the RNAP expressed and purified from mammalian cells, it was necessary to find a different protein expression/purification system, one that would allow for increased yields. While scaling up the expression and

purification of the protein in mammalian cells is possible, given the yields determined here, it would be impractical to scale the cell culture enough so that micromolar concentrations of protein would be obtained.

The baculovirus-expression system in insect cells is known to yield higher amounts of protein than the plasmid-expression system in mammalian cells. One of the benefits of a greater yield of protein after the affinity purification step (first, TAP-purification) is that it allows for the protein to be subjected to a second one: size exclusion chromatography allows for separation of the intact, native RNAP complex and the complex charged with an RNA strand, on the one hand, and partially formed complexes, containing only two of the three subunits, on the other (fig. 7B).

The H3N2 influenza RNA polymerase was expressed and purified as described above (Narin Hengrung, University of Oxford); the protein samples obtained via this method show >95% purity (fig. 7A) and are active in a transcription-based activity assay (fig. 7C).

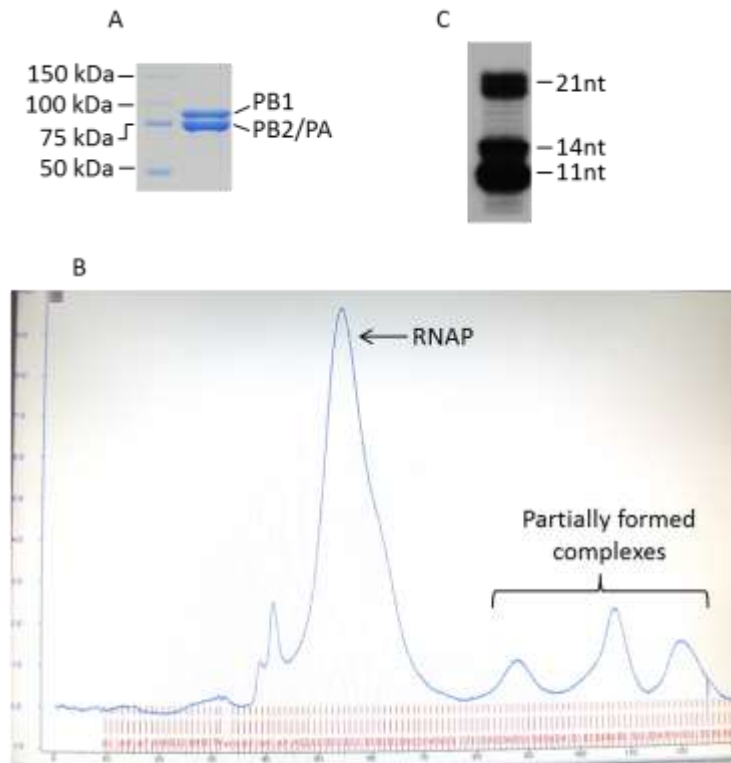


Fig. 7 – A) SDS-PAGE analysis of the H3N2 RNA polymerase (Coomassie stained). B) Gel filtration chromatogram showing the different fractions that can be separated: the major peak, corresponds to the RNA polymerase fraction (this is the sample that was used in all subsequent experiments), while the three minor peaks to the right of the RNAP one are corresponding to fractions that contain partially formed RNAP complexes (by association of only two of the subunits) and possibly, single subunits. (Narin Hengrung). C) Transcription products resulting from an *in vitro* assay, assessing the activity of the RNA polymerase.

1.4 DISCUSSION

The protein obtained through expression in mammalian cells, although pure and active, is produced with a low yield. While this might be sufficient for qualitative studies of the polymerase, it is not for a quantitative investigation of the protein's behaviour in the presence of the viral promoter.

Through their adherent nature and the fact that they grow in a monolayer, HEK 293T cells pose a well-defined limit to the scalability of the culture, and implicitly, on the maximal yield of protein produced in this expression system. On the contrary, the Sf9 insect cells grow in suspension, allowing the culture to be down- or upsized according to needs.

The protein obtained using the baculovirus-expression system is equally pure and active as the one purified from mammalian cells; the advantage of this protein preparation is that it allows for separation of the RNA-charged and uncharged RNA polymerase complexes, which is beneficial for the experiments which aim to determine binding constants for the protein binding to different species of viral RNA (single-stranded RNAs and the dsRNA promoter), as discussed in Chapters 2 and 3.

1.5 CONTRIBUTIONS

The H3N2 recombinant RNA polymerase was expressed and purified as described in section 1.2.5 by Narin Hengrung (University of Oxford).

1.6 BIBLIOGRAPHY

- Akkina, R.K. et al., 1987. Intracellular localization of the viral polymerase proteins in cells infected with influenza virus and cells expressing PB1 protein from cloned cDNA. *Journal of virology*, 61(7), pp.2217–2224.
- Argos, P., 1988. A sequence motif in many polymerases. *Nucleic acids research*, 16(21), pp.9909–16.
- Biswas, S.K. & Nayak, D.P., 1994. influenza A virus polymerase basic protein 1 . Mutational Analysis of the Conserved Motifs of Influenza A Virus Polymerase Basic Protein 1. , 68(3).
- Braam, J., Ulmanen, I. & Krug, R.M., 1983. Molecular model of a eucaryotic transcription complex: Functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell*, 34(2), pp.609–618.

- Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1-2), pp.248–254.
- Deng, T. et al., 2005. In Vitro Assembly of PB2 with a PB1-PA Dimer Supports a New Model of Assembly of Influenza A Virus Polymerase Subunits into a Functional Trimeric Complex. , 79(13), pp.8669–8674.
- Dias, A. et al., 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458(7240), pp.914–8.
- Fechter, P. et al., 2003. Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *The Journal of biological chemistry*, 278(22), pp.20381–8.
- Fields, B.N., 2007. *Fields' Virology* 5th Editio. D. M. Knipe & P. M. Howley, eds., Lippincott Williams & Wilkins.
- Flick, R. & Hobom, G., 1999. Interaction of influenza virus polymerase with viral RNA in the “corkscrew” conformation. *The Journal of general virology*, 80 (Pt 10), pp.2565–72.
- Fodor, E. et al., 2002. A Single Amino Acid Mutation in the PA Subunit of the Influenza Virus RNA Polymerase Inhibits Endonucleolytic Cleavage of Capped RNAs. , 76(18), pp.8989–9001.
- Gastaminza, P. et al., 2003. Mutations in the N-Terminal Region of Influenza Virus PB2 Protein Affect Virus RNA Replication but Not Transcription Mutations in the N-Terminal Region of Influenza Virus PB2 Protein Affect Virus RNA Replication but Not Transcription.
- González, S. & Ortín, J., 1999a. Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. *Journal of virology*, 73(1), pp.631–637.
- González, S. & Ortín, J., 1999b. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *The EMBO journal*, 18(13), pp.3767–75.
- Guilligay, D. et al., 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature structural & molecular biology*, 15(5), pp.500–6.
- Hara, K. et al., 2013. Co-incorporation of the PB2 and PA polymerase subunits from human H3N2 influenza virus is a critical determinant of the replication of reassortant ribonucleoprotein complexes. *The Journal of general virology*, 94(Pt 11), pp.2406–16.
- Hara, K. et al., 2001. Influenza virus RNA polymerase PA subunit is a novel serine protease with Ser624 at the active site. *Genes to cells : devoted to molecular & cellular mechanisms*, 6(2), pp.87–97.
- He, X. et al., 2008. Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature*, 454(7208), pp.1123–6.

- Hemerka, J.N. et al., 2009. Detection and characterization of influenza A virus PA-PB2 interaction through a bimolecular fluorescence complementation assay. *Journal of virology*, 83(8), pp.3944–55.
- Holden, S.J. et al., 2010. Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophysical Journal*, 99(9), pp.3102–3111.
- Honda, A., Mizumoto, K. & Ishihama, A., 2002. Minimum molecular architectures for transcription. , 99(20), pp.1–6.
- Honda, A., Mizumoto, K. & Ishihama, A., 1999. Two separate sequences of PB2 subunit constitute the RNA cap-binding site of influenza virus RNA polymerase. *Genes to Cells*, 4(8), pp.475–485.
- Jung, T.E. & Brownlee, G.G., 2006. A new promoter-binding site in the PB1 subunit of the influenza A virus polymerase. *The Journal of general virology*, 87(Pt 3), pp.679–88.
- Kruger, N.J., 1994. The Bradford method for protein quantitation. *Methods in molecular biology (Clifton, N.J.)*, 32, pp.9–15.
- Lee, M.T.M. et al., 2002. Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic acids research*, 30(2), pp.429–38.
- Li, M.L., Rao, P. & Krug, R.M., 2001. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *The EMBO journal*, 20(8), pp.2078–86.
- Maier, H.J. et al., 2008. Differential role of the influenza A virus polymerase PA subunit for vRNA and cRNA promoter binding. *Virology*, 370(1), pp.194–204.
- Obayashi, E. et al., 2008. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature*, 454(7208), pp.1127–31.
- Perales, B. et al., 2000. The replication activity of influenza virus polymerase is linked to the capacity of the PA subunit to induce proteolysis. *Journal of virology*, 74(3), pp.1307–12.
- Perales, B. & Ortín, J., 1997. The influenza A virus PB2 polymerase subunit is required for the replication of viral RNA. *Journal of virology*, 71(2), pp.1381–5.
- Pleschka, S. et al., 1996. A plasmid-based reverse genetics system for influenza A virus . These include : A Plasmid-Based Reverse Genetics System for Influenza A Virus. , 70(6).
- Poch, O. et al., 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *The EMBO journal*, 8(12), pp.3867–74.
- Puig, O. et al., 2001. The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. *Methods*, 24(3), pp.218–229.

Rasband, W.S., 1997. ImageJ.

Regan, J.F., Liang, Y. & Parslow, T.G., 2006. Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *Journal of virology*, 80(1), pp.252–61.

Rigaut, G. et al., 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nature biotechnology*, 17(10), pp.1030–2.

Sanz-Ezquerro, J.J. et al., 1995. Individual expression of influenza virus PA protein induces degradation of coexpressed proteins. *Journal of virology*, 69(4), pp.2420–6.

Sanz-Ezquerro, J.J. et al., 1996. The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis. *Journal of virology*, 70(3), pp.1905–11.

Simpson, R.J., 2006. Fluorescent Staining of Proteins with SYPRO Ruby. *CSH protocols*, 2006(5), p.pdb.prot4567–.

Sugiyama, K. et al., 2009. Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *The EMBO journal*, 28(12), pp.1803–11.

Taubenberger, J. & Morens, D., 2006. 1918 influenza: the mother of all pandemics.

Toyoda, T. et al., 1996. Molecular Assembly of the Influenza Virus RNA Polymerase: Determination of the Subunit-Subunit Contact Sites. *Journal of General Virology*, 77(9), pp.2149–2157.

Yuan, P. et al., 2009. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature*, 458(7240), pp.909–13.

Chapter 2: Studying the interaction between RNA polymerase and single stranded RNAs

2.1 INTRODUCTION

Besides its function as a replicase, transcriptase and poly(A) polymerase (Fodor 2013), the influenza RNA polymerase is also an RNA binding protein, associating tightly with the viral genome. Each of the eight influenza genomic RNA segments (Mcgeoch et al. 1976) feature conserved and partially complementary sequences at their 3' and 5' ends (Robertson 1979; Desselberger et al. 1980). The 5' end sequence is 13nt long, while the 3' end one consists of 12nt; two sequences were proposed for the 3' end – they are identical, except for the residue found at position 4 which can be either a U or a C (Desselberger et al. 1980). Invariably, for type A viruses, the segments encoding for the three polymerase proteins have a 3' C4 end (Desselberger et al. 1980). This heterogeneity at position 4 on the 3' end is believed to play an important role in regulation of transcription, replication and pathogenicity, all of which have been explained through the decreased binding affinity for the C4 promoter by the polymerase (Lee & Seong 1998; de Wit et al. 2004; Jiang et al. 2010; Sun et al. 2014).

The interaction between the RNA polymerase and the conserved gene ends has been extensively studied. Two main directions of study can be distinguished, with one investigating the individual termini (5' or 3' end) binding to the polymerase, while the other is concerned with investigating the polymerase interacting with the double

stranded panhandle (formed when the 5' and 3' ends base-pair through their complementary regions).

When considering the ends separately, it was shown that both of them can be bound by the polymerase independently, especially the 5' end, which also binds with a higher affinity than the 3' end (Fodor et al. 1994; Tiley et al. 1994; González & Ortín 1999a). Moreover, when the 5' end is bound individually by the protein, the complex is capable of binding a single-stranded 3' end, but is not able to bind a double-stranded structure such as the promoter; this observation led to the proposal that sequential binding of the 5' and 3' ends might influence activation of the protein for transcription (Tiley et al. 1994; Li et al. 1998). While binding of the two ends was initially believed to occur at two different binding sites, each of them specific for one of the viral ends (Li et al. 1998), it was later shown that the PB1 subunit of the polymerase has two binding sites for the 5' end and that 3' end binding to the polymerase is mediated through 3' base-pairing to the already bound 5' end (Jung & Brownlee 2006); moreover, this same study suggested that the PB2 subunit of the polymerase is specifically involved in binding the 3' viral terminus. Several studies established PB1 to be the most important protein for 5' end binding (Li et al. 1998; González & Ortín 1999a), but evidence that all of three of the polymerase subunits contribute was also put forward (Tiley et al. 1994; Hagen et al. 1994).

Further investigation on whether the binding of these individual RNA structures is structure-dependent or sequence-dependent revealed that several residues within the vRNA conserved promoter are paramount for establishing contacts with the polymerase. Specifically, residues A1, G2, U3 and C9 on the 5' end and G9, U10, C11 and C12 on the 3'

end are considered instrumental for binding and mutating them severely affects complex formation; alternatively, mutating residues A7, A8 and A11 on the 5' end and U7 and U8 on the 3' end also decreases binding, but to a lesser extent (Fodor et al. 1993; Fodor et al. 1994). These studies suggest that in establishing the polymerase-vRNA complex, the sequence of the RNA to be bound is important.

The interaction between the polymerase and the vRNAs seems to have implications for the function and activity of the protein. For example, it was found that in transcription, mRNA cap binding is dependent on the presence of a bound 5' end (Cianci et al. 1995), while endonucleolytic cleavage, the next step in viral transcription, is conditioned by both of the viral termini being present (Hagen et al. 1994; Cianci et al. 1995).

Since the influenza polymerase functions both as a transcriptase and a replicase, and replication of vRNA occurs via an intermediary, complementary RNA (cRNA), it follows that the polymerase needs to be able to bind to cRNA, too. This positive-sense intermediate is also able to form a promoter, but the interaction between the protein and the individual ends of the cRNA segments or even the double-stranded promoter was not extensively studied. Yet it was shown that the viral polymerase binds the 5' end and 3' end of cRNA independently, just as it is capable of binding their vRNA correspondents (González & Ortín 1999b). This same study further suggested that while binding is still largely carried out by the PB1 subunit, the binding sites for the vRNA and the cRNA are distinct – with implications for recognition of the two RNA species.

As much of the previous published work was carried out with polymerase from nuclear extracts or that has been immunoprecipitated, some of the contradicting results that

have been previously published could be explained by the fact that the protein might sometimes have been charged with one of the vRNA ends (most possibly, the 5' end).

In this chapter, I discuss results that show further proof of the polymerase's differential affinity for the two terminal sequences of the viral genes; I attempt to investigate whether the polymerase binding affinity for the 3' end changes when the 5' end is present and, finally, I briefly look at the polymerase binding to the conserved ends of the cRNA promoter.

2.2 MATERIALS AND METHODS

2.2.1 RNA polymerase

The recombinant A/Northern Territories/60/68 RNA polymerase (H3N2) used in the experiments described here was expressed and purified from Sf9 cells as described in Chapter 1.

2.2.2 RNA strands

RNA strands of various lengths corresponding to the 3' end and 5' end conserved ends of the single-stranded viral RNA segment encoding for the neuraminidase gene were custom synthesised (IBA, Germany) to carry an NH₂-U residue and a biotin, as indicated (table 1); alternatively, the RNAs were fluorescently labelled by the manufacturer (table 1).

Strand	Sequence (5'→3'; modified residue in bold)
5' end	AGUAGAAACAAGGAGUU(U-NH₂)
5' end	AGUAGAAACAAGGAGUU(U-Cy3)UUUGAACAAACUACUU-biotin
5' end (mutated)	AAUAGAUAAAAGGAGUU(U-NH₂)
3' end	biotin-UUUAAACUCCUGCUUUUGC(U-NH₂)

3' end	biotin-UUUAAACUCCUGCUUU(U-NH₂)GCU
3' end	(ATTO647N-U)AAACUCCUGCUUUCGCU
3' end (mutated)	biotin-UUUAAACUCCUCCAUU(U-NH₂)GUU

Table 1 – vRNA strands used in this study

RNA strands corresponding to the 3' and 5' ends of the conserved complementary RNA (cRNA) promoter of the neuraminidase gene (table 2) were also custom synthesized and labelled (IBA).

Strand	Sequence (5'->3'; modified residue in bold)
5' end	AGCAAAAGCAGGAGUU(U-ATTO647N)
3' end	(ATTO647N-U)AAACUCCUUGUUUCUACU

Table 2 – cRNA strands used in this study

2.2.3 RNA labelling and purification

50pmol NH₂-modified RNAs were labelled by reacting them with 50nmol Cy3 NHS-ester (GEHealthcare) or with 50nmol ATTO647N NHS-ester (Sigma) in 100mM sodium tetraborate at 37°C, 400rpm shaking overnight. The RNA was ethanol precipitated, re-suspended in 1x sample volume 100% formamide, denatured at 95°C for 5 minutes and electrophoresed on a 6M urea/20% polyacrylamide gel at 500V until the indicator dye (bromophenol blue – loaded separately) has reached the bottom. To identify the position of the bands of interest, the gel was visualized by 1) using a BioRad Molecular Imager PharosFX system equipped with a 532nm laser (excitation intensity: 25mW; for the strands labelled with Cy3) and a 635nm laser (excitation: 10mW; for the strands carrying an ATTO647N label) and 2) UV shadowing, after which the desired products were excised. Labelled RNA was recovered using the “crush and soak” method in 500µl TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0) at room temperature, 400rpm shaking,

overnight. Samples were then centrifuged and the labelled RNA-containing supernatant was recovered; fractions were concentrated in a vacuum concentrator and the resulting pellets re-suspended in 20-50µl TE buffer. Finally, the RNA samples were desalted by size exclusion chromatography on a Micro Bio-Spin 6 column (BioRad) in TE buffer.

2.2.4 Determining labelled RNA concentration and labelling efficiency

The absorbance of the labelled RNA samples was measured at 260nm (for the RNA) and at 550nm (for the Cy3 dye) or 647nm (for the ATTO647N dye); the A_{RNA}/A_{dye} allowed for determination of the crosstalk between the two (crosstalk is a measure of the contribution of the dye to the A_{RNA} value). Labelled RNA absorption was determined using the formula:

$$A_{RNA(labelled)} = A_{RNA} - crosstalk * A_{dye}$$

Applying the Beer-Lambert law, $A = \epsilon cl$, the concentration (c) of the labelled RNA and of the dye in the sample can be determined; ϵ (which stands for molar extinction coefficient, a measure of how strongly a substance absorbs light of a given wavelength) is given by the manufacturer for each RNA strand and is equal to $150,000 \text{ M}^{-1}\text{cm}^{-1}$ for both Cy3 and ATTO647N; $l=1\text{cm}$ (this is the path length over which the sample is measured).

Finally, the labelling efficiency allows for determination of how much of the RNA is labelled and/or whether any free dye is still present in the sample; this is calculated using the following formula:

$$X = \frac{C_{RNA(labelled)} - C_{dye}}{C_{RNA(labelled)}} * 100$$

When $X=100\%$ it signifies that the labelling reaction was 100% efficiency (all the RNA in the sample is labelled and there is no excess dye); if $X<100\%$, the difference to 100 signifies the percentage of RNA in the sample that is unlabeled, while if $X>100\%$, the excess over 100, signifies the percentage of free dye still present in the sample.

2.2.5 Electrophoretic Mobility Shift Assay (EMSA)

20nM of either Cy3-labelled or ATTO647N-labelled ssRNA was incubated on its own or in the presence of increasing concentrations of protein in T8 buffer (50mM Tris-HCl pH 8.0, 100mM KCl, 10mM $MgCl_2$, 100 μ g/ μ l BSA, 1mM DTT and 5% glycerol) supplemented with 500mM NaCl and 2U RNasin (40U/ μ l, Promega) and incubated at 28°C for 15 minutes. Reactions were stopped by addition of 1x reaction volume of 50% glycerol, after which samples were loaded on a 1% TBE-agarose gel which was run for 2 minutes at 150V and then for 1 hour at 70V. The gel was analysed on a BioRad Molecular Imager PharosFX system as described previously in section 2.2.2.

2.2.6 Protein Induced Fluorescence Enhancement (single-molecule; on surface)

Glass slide cleaning and immobilisation of biotinylated nucleic acids have been previously described (Margeat et al, 2006, Biophys J); in brief, a cleaned glass microscopy slide onto which a 4-well gasket had been sealed, previously treated with PEG:PEG-biotin (100:1.25 ratio), was incubated with 0.5mg/ml neutravidin for 5 minutes at room temperature. 50pM of biotinylated Cy3-labelled ssRNA was added to the well and incubated for 2 minutes at room temperature, after which the buffer was replaced with RNA imaging

buffer [T8 buffer supplemented with 1% glucose, 165 U/ml glucose oxidase, 2170 U/ml catalase, and 1mM Trolox; (Roy et al. 2008)] and sealed with a glass coverslip before imaging.

For measurements taken in the presence of the protein,

The measurements were taken on a custom-built total internal reflection microscope (TIRF), under 561nm laser excitation (continuous wave), 1mW intensity and with a 100ms temporal resolution. For each sample, a total of 20 movies were collected, each movie being 5 minutes long. Analysis of the collected data was performed using custom-written software (TwoTone: Holden et al. 2010, Seneca: Kristofer Gryte, University of Oxford, unpublished).

2.2.7 Fluorescence anisotropy

A two-fold serial dilution of polymerase (starting concentration: 1 μ M; lowest concentration: 0.02nM) was incubated with 1nM ATTO647N-labelled RNA in T8 buffer supplemented with 0.5M NaCl, for 15 minutes at 28°C. The measurement was performed in a quartz cuvette, in an L-format scanning fluorimeter (Photon Technology International), equipped with Glan-Thomson polarizers; the sample was excited at 640 nm and its emission measured at 670 nm for 1 minute at 5 seconds intervals. The data was plotted in OriginPro8 and a K_D value was extracted by fitting with the following function:

$$y = A_0 + (A_f - A_0) \left(\frac{(x + D_t + K_D) - \sqrt{(x + D_t + K_D)^2 - 4D_t x}}{2D_t} \right)$$

A_0 = initial anisotropy value

A_f = final anisotropy value, fixed at 0.37a.u

D_t = RNA concentration, fixed at 1nM

K_D = binding constant

2.3 RESULTS

2.3.1 Investigating polymerase binding to the 5' end of the vRNA in solution

Initially, the interaction between the polymerase and the 5' RNA conserved end was investigated using an electrophoretic mobility shift (EMSA) approach, a commonly used assay for determining the properties of proteins interacting with nucleic acids (Hellman & Fried 2007). Conceptually simple, the experiment proved to be technically challenging, since the protein tended to aggregate at high nanomolar concentrations (Chi et al. 2003). Nevertheless, complex formation could still be detected (fig. 1A).

Repeating the assay several times confirmed that the observed behaviour is consistent, suggesting that it might be possible to analyse the result quantitatively. Direct inspection of the EMSA result (fig. 1A) suggests that the protein concentration at which half of the RNA is bound (i.e. the K_D value) is between 20-30nM. To verify this, intensities of the observed bands (both for the bound and the unbound species) were determined and the obtained values were used to calculate how much of the RNA is bound ('% bound');

plotting this against the protein concentration led a to plot, that when fitted, suggests that the binding constant for this interaction is much lower than the quantity of RNA used in the experiment (fig. 1B).

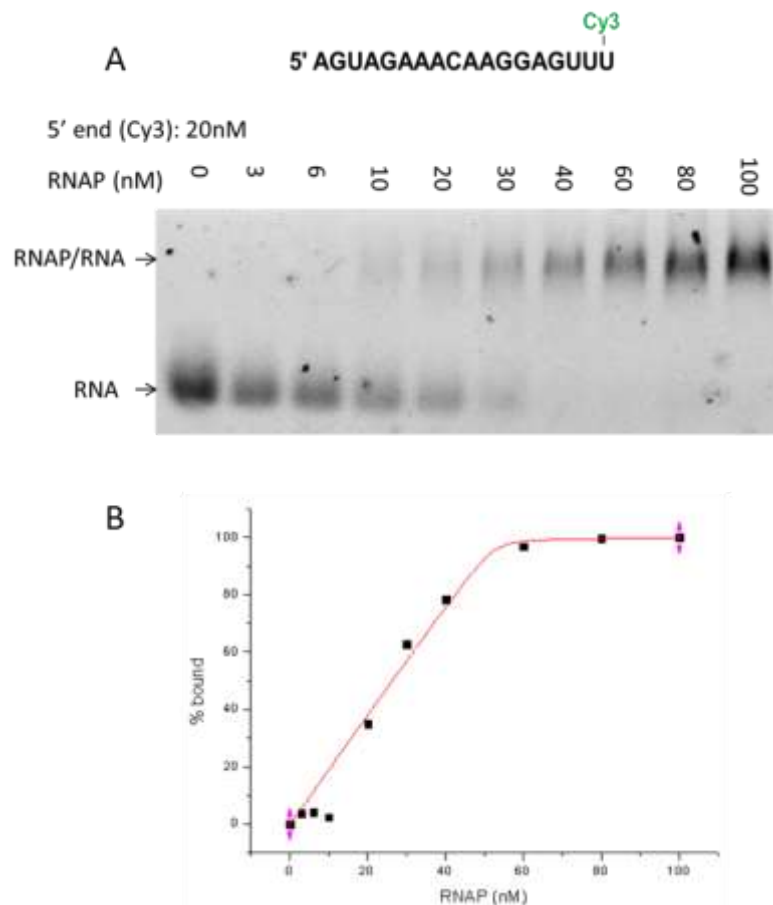


Fig. 1 – A) EMSA analysis of 3-100nM polymerase binding to 20nM Cy3-labelled 5' end RNA (the sequence and label position is depicted above the gel image). The separated products are the RNAP/RNA complex (upper bands) and the free RNA (lower bands). B) Plot of the bound RNA fraction (in % to the total RNA detected) vs. protein concentration, as determined from analysis of the image in 1A using the 'gel analyzer' module in the ImageJ software; the data was plotted in OriginPro8 and fitted using a quadratic function as described in the Materials and Methods section.

However, caution needs to be used when interpreting the results of this experiment in a quantitative manner, since the EMSA is not a true equilibrium method; it is, nevertheless, a good indication towards the individual binding behaviour of the polymerase with respect to the given 5' end substrate. Hence, to determine a true binding constant for the interaction between the polymerase and the 5' terminal sequence of the vRNA genome, a different approach was employed.

A more sensitive technique to investigate such interactions is fluorescence anisotropy, which monitors the degree by which light emitted by a fluorophore attached to a diffusing ssRNA in solution is polarised (the more polarised emitted light, the higher the anisotropy). If the investigated RNA is free from protein, it will rotate faster than the time required for light emission, causing most of the emitted light to be depolarised; on the contrary, when the RNA is bound by a protein, the complex will move much slower, sometimes not even changing orientation between fluorophore excitation and emission, in which case the emitted light will be largely polarised. Theoretically, the maximum anisotropy value possible is 0.4a.u; however, in practice, since the dye dipole can still rotate during the lifetime of the excited state, some depolarization occurs and the anisotropy rarely takes values above 0.37a.u (which is the value used as a maximum limit when fitting experimental data presented here).

As expected, titrating increasing amounts of the polymerase to the wild type 5' end promoter strand (fig. 2A), leads to an initial increase in anisotropy corresponding to the ssRNA—protein complex being formed. At higher protein concentrations (30-1000nM), the anisotropy doesn't change and we equate this to the system having reached

saturation. The data set used to determine the binding constant is an average of three independent results, fitted as described in *Materials and Methods*, section 2.2.7 to reveal a $K_D=2.2\pm0.6\text{nM}$ for this interaction. As a control, a 5' end polymerase binding mutant strand (G2->A, A7->U, C9->A) was used as a substrate (fig. 2B), showing that binding was greatly reduced ($K_D>1\mu\text{M}$).

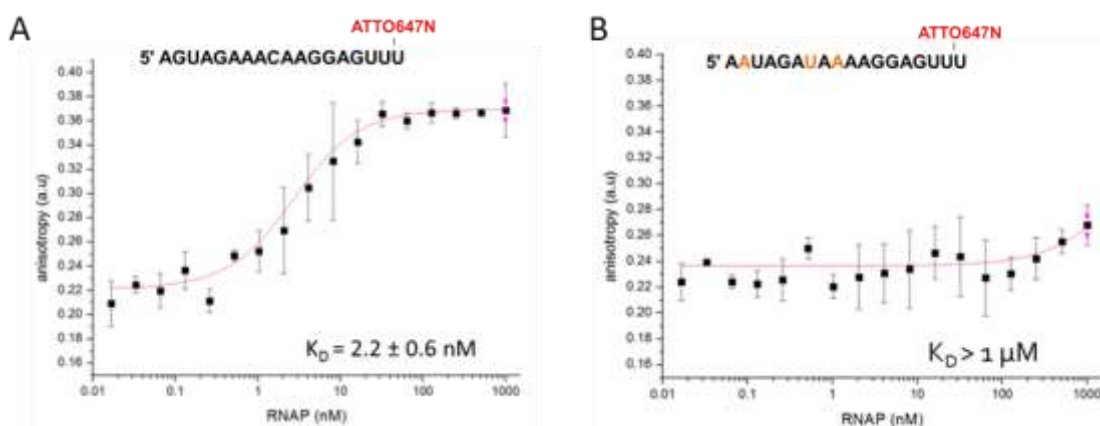


Fig. 2 – A) Fluorescence anisotropy curve resulting from incubation of a serial dilution of RNAP with 1nM ATTO647N-labelled wild type 5'end. The data was plotted in OriginPro8 and fitted using a quadratic function as described in the Materials and methods section; the determined K_D for this interaction was $2.2\pm0.6\text{nM}$. B) Fluorescence anisotropy curve resulting from incubation of a serial dilution of RNAP with 1nM ATTO647N-labelled mutated 5'end (mutated residues are shown in orange). The data was fitted as above, but since the titration did not reach saturation, no binding constant could be determined.

2.3.2 Investigating polymerase binding to the 5' end of the vRNA on the surface

It was recently shown that dyes undergoing *cis-trans* isomerisation can have their fluorescence enhanced if their local environment changes, for example when they are in the proximity of a protein; this is due to reducing the rate of isomerisation to the

photoinactive *cis* isomer and stabilising the *trans* state. The process is referred to as protein induced fluorescence enhancement (PIFE; Hwang et al, 2011) and can also be used to monitor polymerase binding to RNA.

In brief, an RNA strand, which is labelled with both a biotin and the organic fluorophore Cy3, known to be amenable to PIFE, is immobilised on a glass surface and imaged. Studying the RNA in the absence of protein allows for determination of the fluorophore intensity when attached to the RNA (for the purpose of these experiments, this value will indicate the fluorescence baseline), as well as a confirmation that there are no intensity fluctuations in the absence of the protein (fig. 3C). Of the 578 RNA particles detected, about 80% are characterised by an average intensity of $\sim 1750 \text{ a.u.}$ (ranging between 1000 and 2500 a.u.). For the rest 20% of molecules, the dye seems to be affected by photophysical effects, such as blinking and bleaching. Upon addition of 500 nM RNA polymerase, an increase in fluorescence intensity can be almost immediately detected even by looking at the field of view (fig. 3B, compare the 'RNA only' panel with the other two shown). In the presence of the protein, a total of 591 molecules were detected, of which approximately 5% exhibited PIFE, characterised as an increase in intensity from the baseline established before to $\sim 4000 \text{ a.u.}$ (signalling a 2.2x increase in fluorescence intensity, on average). Of the detected bound molecules, only a small fraction were characterised by multiple binding events; this might be explained by surface effects such as the RNA or the protein binding non-specifically to the glass surface and hence, becoming unavailable for binding. Should more single-molecules detected had shown multiple binding, it would have been possible to perform a Hidden Markov Modelling

analysis, which would have yielded dwell times for the RNA in the unbound and bound states, leading to determining a binding constant.

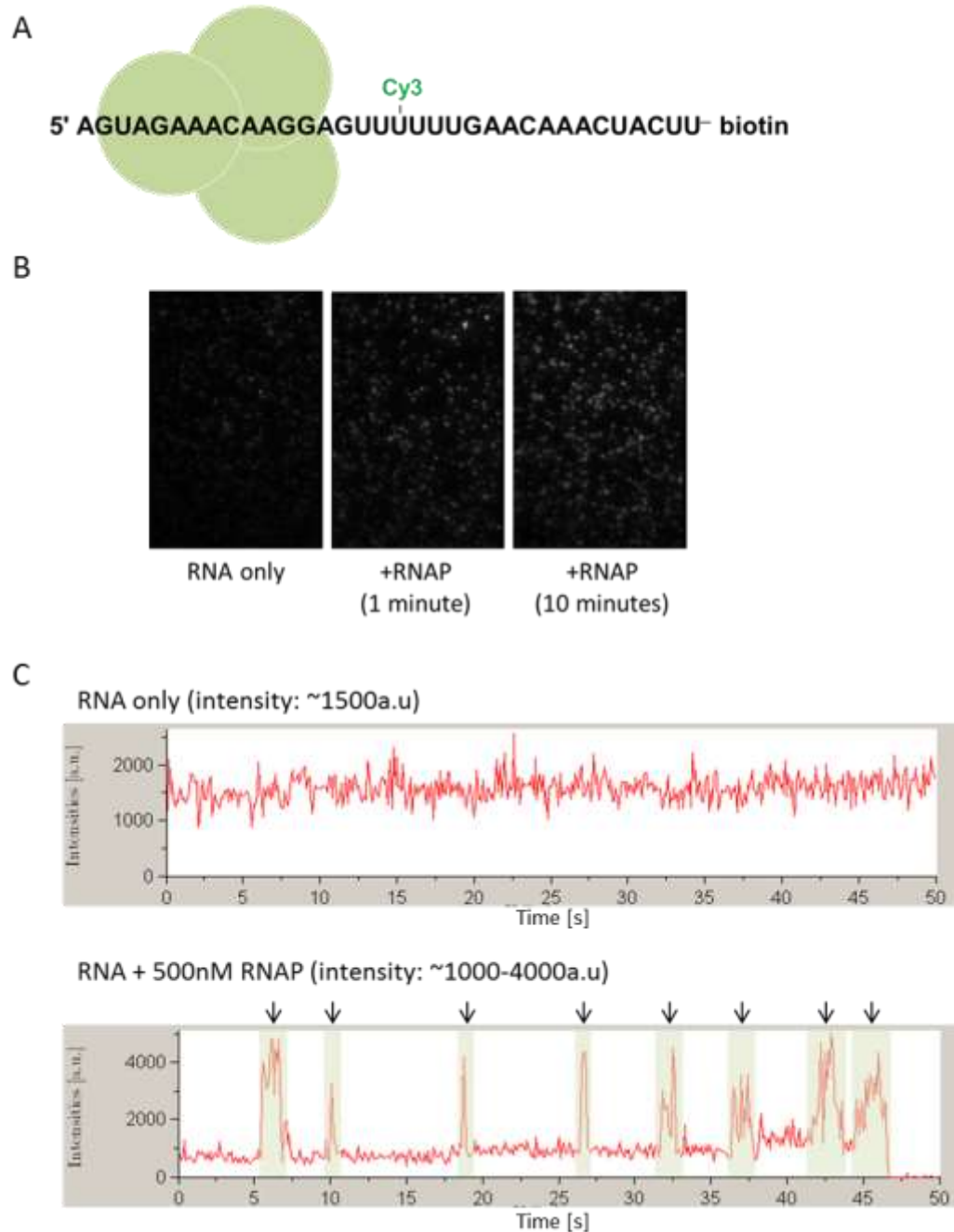


Fig. 3 – A) 5' end construct used for the experiments on the surface, with a Cy3-label placed at position 18 (same as previously) and a biotin group attached at position 34. B) Representative field of view captured during the experimental observation. Panel 'RNA only' shows a still frame from a movie of immobilised Cy3-labelled 5' end RNA (50pM concentration); panel '+RNAP (1 minute)' shows a still frame from a movie taken 1 minute after addition of RNA polymerase; finally, panel '+RNAP (10 minute)' shows a still frame from a recording taken 10 minutes after the

polymerase was added to the well. C) A representative single-molecule time trace for an unbound RNA molecule, showing a recorded intensity of $\sim 1500\text{a.u.}$; below, a time trace for an RNA molecule that is bound repeatedly by RNA polymerase (added at a final concentration of 500nM) – the intensity fluctuates between $\sim 1000\text{a.u.}$ (attributed to the unbound RNA) and $\sim 4000\text{a.u.}$ (attributed to the bound RNA exhibiting PIFE; arrows and highlighting indicate increases in detected fluorescence).

2.3.3 Investigating polymerase binding to the 3' end of the vRNA in solution

Similarly as for the 5' end, binding to the 3' end was initially investigated using the EMSA method. Binding was first tested by using a ssRNA construct that harboured the fluorescent dye at position 1 on the 3' end (fig. 4); compared to the protein-RNA complex detected for the 5' end, no significant shift could be detected, but a rather slight one, albeit at polymerase concentrations higher than 100nM . This slight shift was attributed to the RNAP/RNA complex; this observed mobility raises the possibility that the interaction is not as strong as for the other terminal sequence and the complex dissociates during the electrophoretic analysis. Alternatively, the position of the dye could be interfering with the polymerase binding stably to this RNA construct.

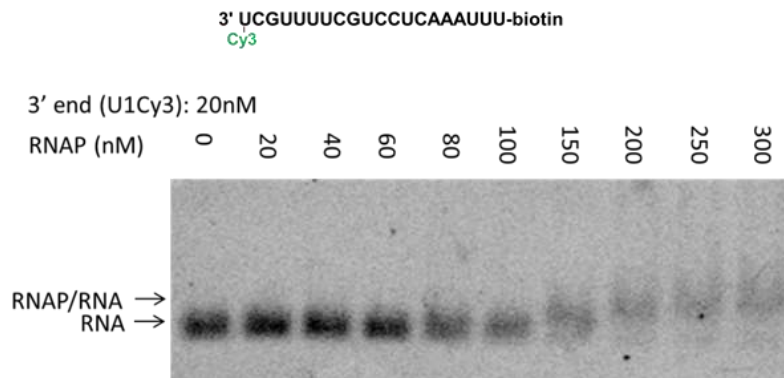


Fig. 4 – EMSA of an increasing concentration of polymerase (20-300nM) titrated to 20nM 3' end strand labelled at position 1 with Cy3.

To test whether this might be the case, an identical construct in sequence, but carrying a fluorophore at position 4 was employed; upon titration of 10-100nM polymerase, formation of the RNA polymerase – RNA complex was observed (fig. 5A). Using the intensity of the detected bands, the percentage of bound RNA fraction was calculated and plotted against protein concentration (fig. 5B): as expected, this kind of analysis could not offer a precise value for the binding constant.

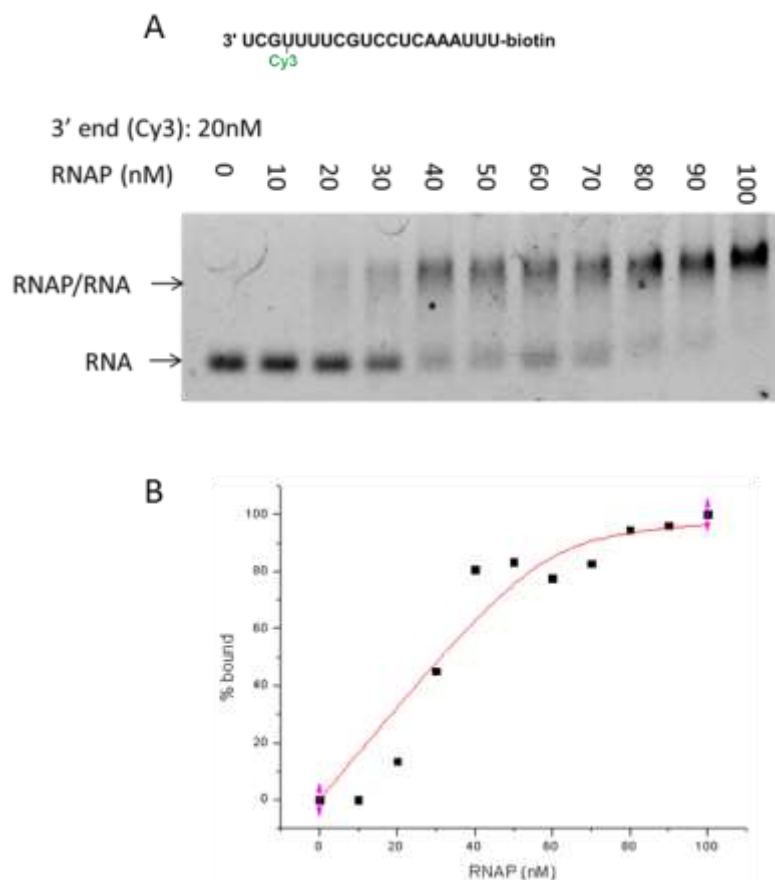


Fig. 5 – A) EMSA of an increasing concentration of polymerase (10-100nM) titrated to 20nM 3' end strand labelled at position 1 with Cy3. B) Plot of the bound RNA fraction (in % to the total RNA detected) vs. protein concentration, as determined from analysis of the image in 5A (fitted as previously described).

The interaction between the polymerase and the 3' end of the viral genome was further investigated using fluorescence anisotropy. The two naturally occurring variations of the 3' end were employed, with the U4 3' end strand labelled at position 4 (fig. 6A) and the C4 construct labelled at position 18 (fig. 6B). In these experiments, the polymerase exhibited a lower affinity for these terminal sequences than it did for the 5' end; in fact, the interaction did not reach saturation within the range of protein concentrations used,

meaning that while an exact K_D value could not be determined, it is certain that this value is higher than $1\mu\text{M}$ (fig. 6A, 6B). A similar behaviour was observed when the mutated 3' end construct (C2->U; U7->A; G9->C), also a polymerase binding mutant, was investigated (fig. 6C).

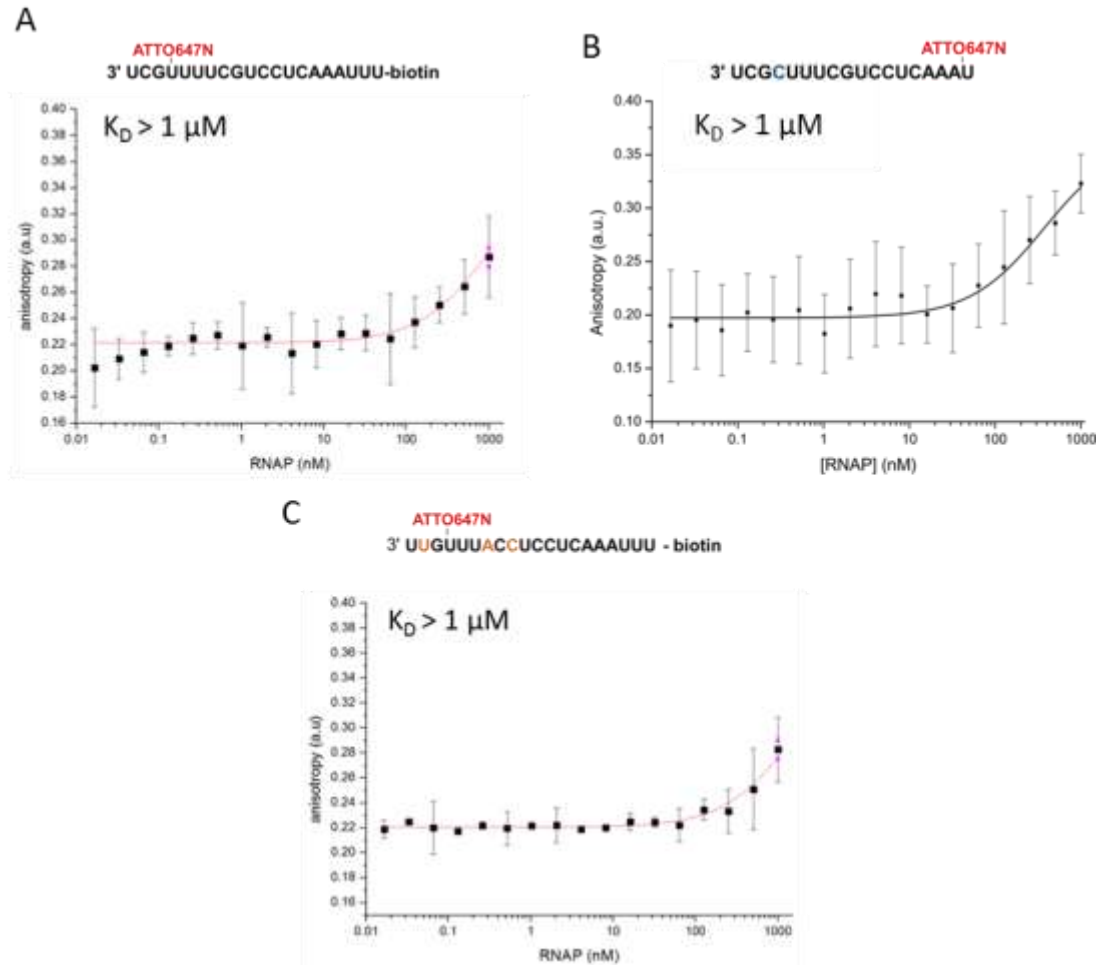


Fig. 6 – A) Fluorescence anisotropy curve for the polymerase binding to 1nM of ATTO647N-labelled 3' end RNA (U4). B) Fluorescence anisotropy curve for the polymerase binding to 1nM 3' end RNA (C4), labelled with ATTO647N at position 18. C) Fluorescence anisotropy curve for the polymerase binding to 1nM ATTO647N-labelled mutated 3' end RNA.

The results of this experiment reinforced the idea that polymerase binding is sequence specific.

2.3.4 Investigating polymerase binding to the 3' end of the vRNA on the surface

As in the case of the 5' end binding, the polymerase – RNA interaction was investigated by initially immobilizing Cy3-labelled RNA on the surface after which 500nM RNA polymerase was added.

When the RNA was imaged on its own, 320 molecules were detected of which 80% were characterised by an average intensity of 1250a.u (ranging between 500 and 2000a.u; fig. 7A,C); when the polymerase was added, out of the 596 particles detected, 10% showed fluctuations in intensity going up to ~3000a.u (fig. 7C), a 2.4x increase in intensity, on average. This increase in intensity was also visible as data was recorded, as shown in the field of view (fig. 7B).

Based on these experiments, 100 time traces showing multiple binding events between the RNAP and the 3' end were further analysed using a Hidden Markov Modelling algorithm (Eddy 2004; McKinney et al. 2006). A Hidden Markov Model (HMM) is a powerful statistical tool allowing for differentiation between specific states in a dynamic process, states that would otherwise be obscured by noise and hence, indistinguishable. In the experiments shown here, HMM allows for determination of the time spent by each molecule in the bound/unbound state (where the 'bound' is defined as the molecule exhibiting the highest fluorescence intensity recorded, while 'unbound' is a molecule showing anything that is not the highest intensity). From the analysis, we could conclude

that the mean time spent in the unbound/off state is 13.5 frames, while the mean time spent in the bound/on state is 6.5 frames; at 100ms temporal resolution, the average time a molecule spends in the unbound state is 1.35 seconds (τ_{free}) and 0.65 seconds, when in the bound state (τ_{bound}) (fig. 7D).

The obtained dwell times for the bound and unbound molecules can be further used to determine K_D (essentially the ratio between τ_{free} and τ_{bound} multiplied by the protein concentration); this calculation suggests that the K_D for this interaction is $\sim 1\mu\text{M}$, which is consistent with the earlier recorded value (from the anisotropy experiments).

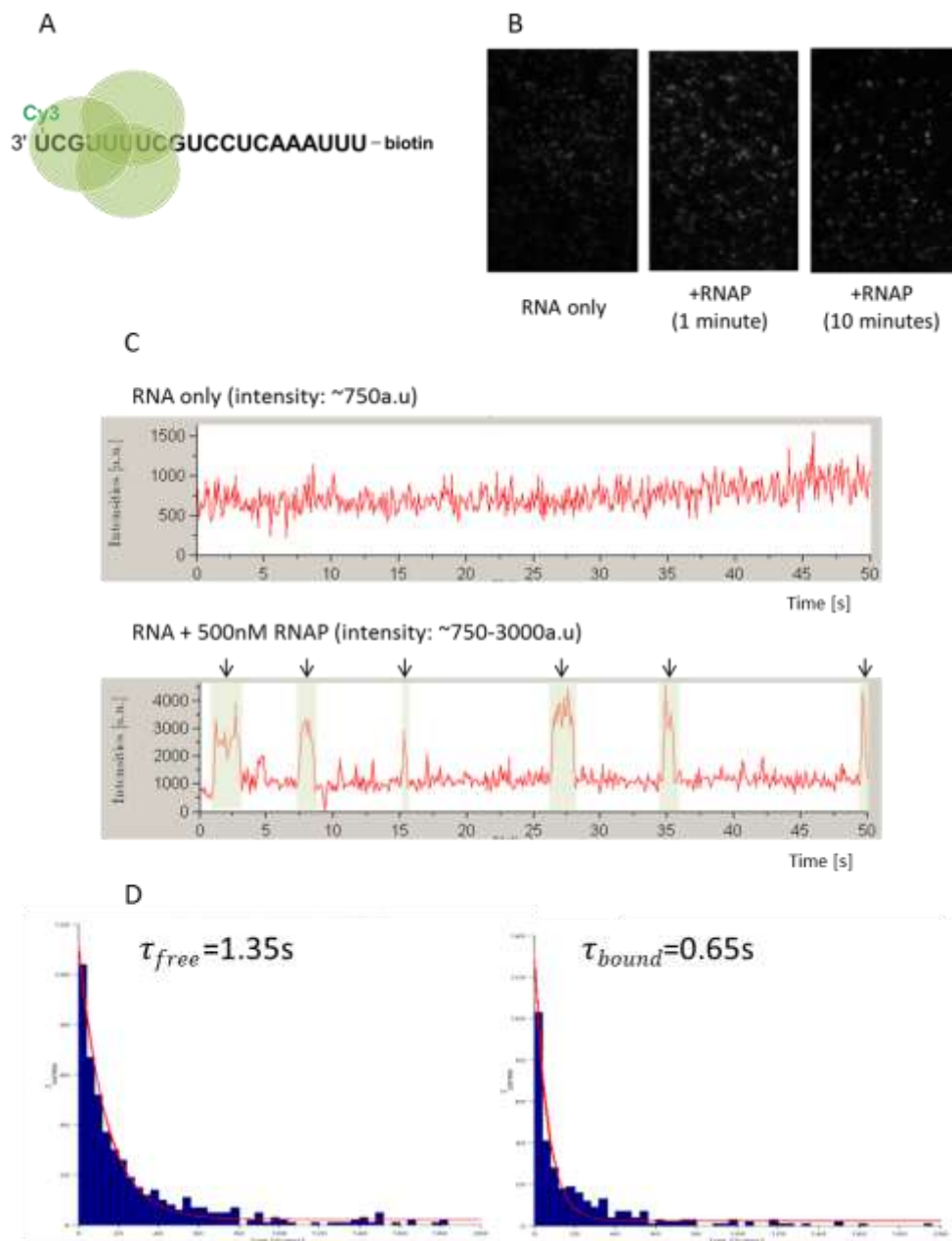


Fig. 7 – A) 3' end construct used for the experiments on the surface, with a Cy3-label placed at position 1 and a biotin group attached at position 20. B) Representative field of view captured during the experimental observation. Panel 'RNA only' shows a still frame from a movie of immobilised Cy3-labelled 3' end RNA (50pM concentration); panel '+RNAP (1 minute)' shows a still frame from a movie taken 1 minute after addition of RNA polymerase; finally, panel '+RNAP (10 minutes)' shows a still frame from a recording taken 10 minutes after the polymerase was added to the well. C) A representative single-molecule time trace for the unbound RNA (~750a.u) and

RNA in the presence of 500nM RNA polymerase (~750-3000a.u). D) Dwell time analysis of unbound RNA (left) and bound RNA (right). The values given are averages obtained by investigating the behaviour of 100 molecules using Hidden Markov Modelling and are characteristic for the time spent by the RNA molecule without being associated with a polymerase (τ_{free}) or when it is in complex with the protein (τ_{bound}).

2.3.5 Polymerase binding to a non-influenza ssRNA

Finally, as a control, the interaction between the influenza polymerase and a non-influenza ssRNA strand was investigated: this experiment was needed to clarify whether the polymerase is capable of binding RNA in a non-specific manner. An EMSA assay showed a faint RNAP/RNA complex band, suggesting that the interaction occurs (fig. 8A). The interaction was further investigated using the fluorescence anisotropy binding assay (fig. 8B); no anisotropy change associated with the protein-ssRNA complex being formed was detected, which contradicts the finding from the previous assay. The contradiction between these results – by which the polymerase binds the non-influenza specific RNA strand (as suggested by the EMSA), on the one hand, but it doesn't bind the same nucleic acid (as suggested by the anisotropy experiment), on the other – can be explained by the fact that the experimental conditions are not identical between the two assays: in the EMSA, the concentration of available RNA is 20 times higher than in the anisotropy-based assay; the concentration of RNA is the same as the concentration of protein (20nM); the conditions during the electrophoretic run could help stabilise the bound complex, even when the interaction is non-specific.

These data taken together support the idea that the polymerase binds specifically to the single-stranded RNA promoter.

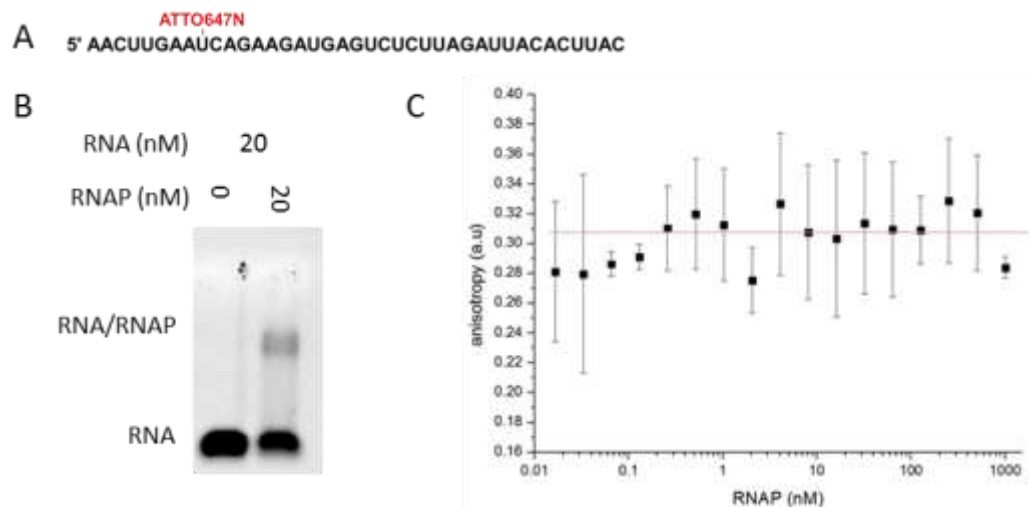


Fig. 8 – A) Sequence of the non-influenza RNA strand, labelled at position 9 with ATTO647N. B) EMSA assay showing the changed mobility of the 20nM ATTO647N-labelled RNA in the presence (20nM) and absence of polymerase. A product that is retarded on the gel can be observed: this is attributed to the RNAP/RNA complex. C) When titrated with increasing concentrations of RNAP (in a serial dilution, as described before), the non-influenza RNA does not show any change in anisotropy.

2.3.6 Investigating the polymerase binding sequentially to the 5' and 3' ends

As several reports have shown that the polymerase binds the two terminal sequences of the viral genome sequentially, investigating whether the affinity for the 3' end is influenced by the presence of the 5' end in the polymerase binding site seemed like the next step.

To investigate this, an excess of 5' end was incubated with the polymerase to allow formation of the binary complex, after which 3' end was added, as to allow for formation of a protein-RNA complex consisting of both viral ends bound by the polymerase (fig. 9A).

Initially, sequential binding was investigated using the EMSA approach: a 10x excess of Cy3-labelled 5' end ($K_D \sim 2.2\text{nM}$; the excess was calculated with regards to this value) was incubated with increasing concentration of polymerase for 15 minutes at room temperature, after which 10nM of ATTO647N-labelled 3' end was added to the reactions, which were incubated for a further 15 minutes and then electrophoresed (fig. 9B). This analysis yielded inconclusive results: while a band corresponding to an RNA polymerase – RNA complex could be detected, co-localisation of the dyes (shown as the orange band in fig. 9B) does not imply that the complex observed is a ternary one, as long as the polymerase – Cy3-5'end and polymerase – ATTO647N-3'end complexes have very similar mobilities.

A better approach for investigating this interaction was fluorescence anisotropy – when the same experiment was performed with 25nM unlabelled 5' end incubated with a serial dilution of RNA polymerase, after which 1nM ATTO647N-labelled 3' end was added: a clear increase was detected as the protein's concentration increased (fig 9C); although only one measurement was taken, a K_D value of $\sim 17\text{nM}$ could be determined.

This K_D value is higher than the one recorded for the 5' end, but much lower than the one determined for the 3' end on its own, suggesting that the polymerase's affinity for the 3' end is indeed dependent on the presence of the 5' terminal sequence.

Sequential binding was also investigated using TIRF-FRET: a biotinylated, Cy3-labelled 5' end construct was immobilised on the glass surface, after which the protein was added. Since the available ATTO4647N-labelled 3' end was also biotinylated, before its addition to the slide, a short incubation with biotin-BSA was performed in order to block any of the

still available neutravidin binding sites (fig. 10A,B). Once both ends were added to the well, oxygen scavengers were also added, the well was sealed and data was taken. Out of all the molecules detected, only those located at the same position in both the green and the red channel were investigated further; these 3 molecules were considered co-localised and a mark of sequential binding occurring (fig. 10C); these molecules were also investigated further. An E/S histogram corresponding to each of them was generated: here, the E/S histogram for the molecule circled in yellow is shown (fig. 10C): the determined E is ~ 0.6 (the similar construct in sequence and having the dyes at the exact same positions as the one used here, studied in solution using confocal microscopy showed a FRET value of 0.5).

Unfortunately, due to the low number of molecules detected, a more quantitative analysis could not be employed.

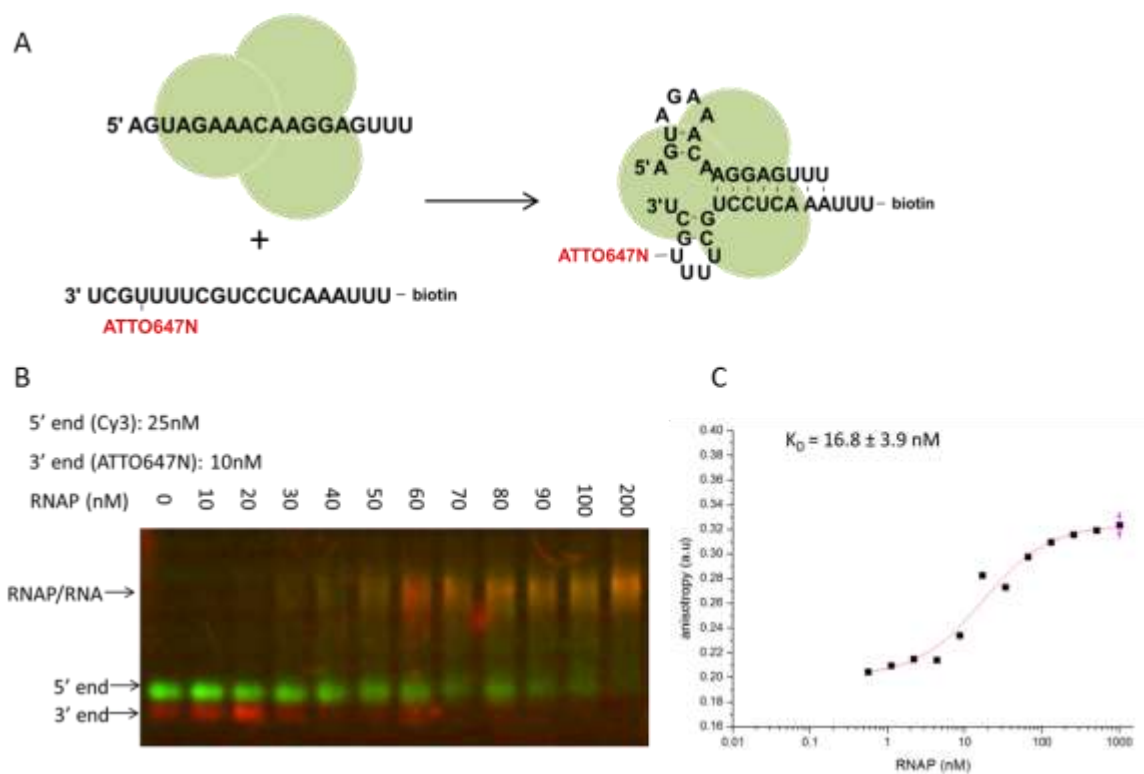


Fig. 9 – A) Diagram of the experimental approach employed in the fluorescence anisotropy experiment (shown in C); the EMSA assay is set up in a similar fashion, the difference being that the 5' end is Cy3-labelled. B) EMSA assay result for testing sequential binding. C) Binding curve resulting from a fluorescence anisotropy experiment (one repeat). The curve was fitted as described in the *Materials and methods* section and a binding constant was determined ($K_D=16.8\pm3.9\text{nM}$).

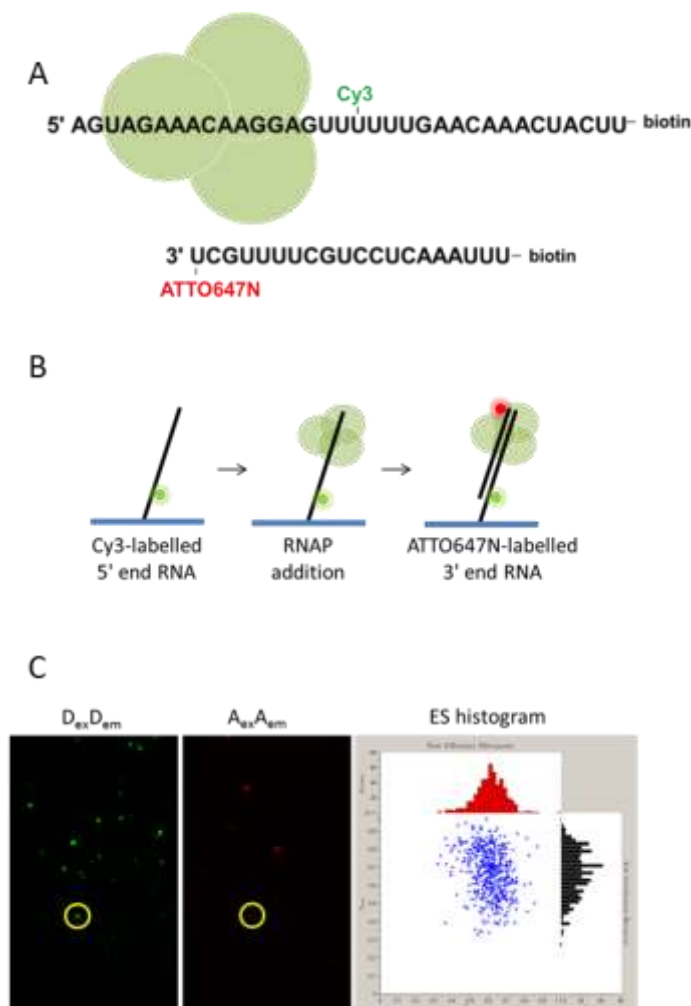


Fig. 10 – A) The two strands used to investigate sequential binding on the surface. B) Sequential binding on the surface experimental scheme: first, the Cy3-labelled 5' end construct was immobilised, after which polymerase was added; then the ATTO647N-labelled 3' end RNA strand was added. C) The same field of view captured in the green ($D_{ex}D_{em}$) channel and in the red ($A_{ex}A_{em}$) channel; the two channels were linked using TwoTone and species that appeared at the same location in both channels were further investigated. The E/S histogram shown is of the circled molecule.

2.3.7 Investigating polymerase binding to the cRNA promoter ends

A full, quantitative binding analysis of the cRNA 5' and 3' ends is beyond to scope of this study; yet, it was tempting to investigate whether the polymerase shows differential

binding affinity for the terminal sequences of the cRNA promoter as it does for the vRNA one. A previous study suggests that the protein interacts with the two cRNA ends equally well (González & Ortín 1999b), which was confirmed by the EMSA assay employed here (fig. 11A and B).

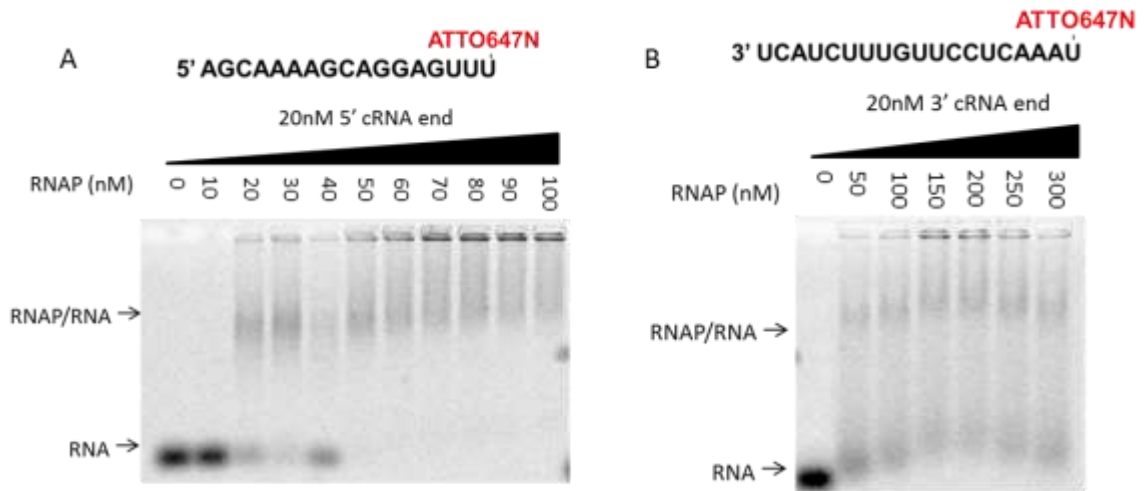


Fig. 11 – A) EMSA of 20nM 5' cRNA end in the presence of increasing concentrations of protein (10-100nM; the range of the protein concentration was chosen in accordance with the observed behaviour of the 5' vRNA end. B) EMSA of 20nM 3' cRNA end in the presence of increasing concentration of protein (50-300nM).

2.4 DISCUSSION

No previous studies have looked at the interaction between the individual ends of the vRNA promoter and the RNA polymerase in a quantitative fashion; several studies have investigated whether the polymerase is capable of binding the ends on their own or whether there is a requirement for the presence of one of them (for example, the 5' end) for binding of the other (the 3' end), but have yielded conflicting results. Here, I present evidence that the protein shows a significantly higher affinity for the 5' end ($K_D=2.2\pm0.6$

nM) than the 3' end ($K_D > 1 \mu\text{M}$), regardless of whether the latter is a U4 or a C4 terminal sequence. When the polymerase is allowed to bind to the 5' terminal sequence first and then to the 3' one, its affinity for the 3' end is higher than in the absence of the other end. This is not surprising given that sequential binding was proposed as the mechanism for polymerase activation (Li et al. 1998; Cianci et al. 1995; Hagen et al. 1994).

In a recent study, it was shown that the polymerase has a 2.4x higher affinity for the U4 dsRNA promoter than for the C4 one (Jiang et al. 2010); as this trend is not apparent in the single-stranded RNA data presented, it is tempting to speculate that the presence of the 5' end might have a differential effect on the protein's affinity for these two 3' variants (in the present study, I only present data on the sequential binding of the U4 end).

Although these results are determined *in vitro*, in an experimental setup in which only the RNA polymerase and the ends of the viral genes were present and in the absence of any other factors that might contribute to binding (such as the NP protein, which is known to interact with both the polymerase and the viral RNA, or any cellular proteins), it is tempting to interpret these results through the currently accepted models for influenza transcription and replication (Fodor 2013).

The latest model for transcription features as a starting point a *cis*-acting polymerase bound to the dsRNA (in the RNP); however, this structure is disrupted and the ends come apart when the capped mRNA primer is made available by the PB2 and PA subunits, shown to cause a conformational change in the polymerase (Resa-Infante et al. 2011). At this stage, the 3' end of the promoter is bound to the active site of the polymerase, where

it interacts with the mRNA primer (transcription initiation); in elongation, the 3' terminal sequence returns to the initial binding site, re-associating with the 5' end in a double-stranded structure. Having a lower binding affinity for the 3' end, potentially allows the polymerase to harbour multiple binding sites for the 3' end (such as the same site as for the 5' end and the active site); the presence of the 5' end was shown to increase the affinity for the 3' terminal sequence (present study; González & Ortín 1999a), which energetically could explain why the two ends re-associate.

During replication, in which the negative-sense vRNA is copied to positive-sense cRNA, two polymerase molecules seem to be involved: one that is *cis*-acting and which remains stably associated with at least one of the two conserved ends at any given time, and a second, *trans*-acting molecule, which moves on the vRNA template as it replicates it, binding the ends transiently. Although it is not completely understood how vRNA->cRNA replication is initiated, the double-stranded promoter structure breaks, with the 5' end remaining associated with the resident polymerase, while the 3' end is bound to the active site of the *trans*-acting molecule. Given our results, it is possible that this polymerase-3' terminal sequence interaction is stabilised by the presence of the initiating nucleotide (in the same way as the capped mRNA-primer might stabilise it in transcription initiation); once this sequence is replicated, the end is returned to the resident polymerase, where it reconstitutes, together with the 5' end, the double-stranded panhandle. The question of whether this double-stranded structure assumes any of the known promoter conformations can also be asked. As the *trans*-acting polymerase

produces full-length cRNA segments, the 5' end needs to be released from the double-stranded promoter structure at the end of replication.

Alternatively, considering the results shown here describing the polymerase binding poorly to the 3' end on its own, doubts can be raised on whether the polymerase ever binds to the 3' end of the vRNA when this strand is on its own, which would suggest that some of the stages of the current accepted transcription and replication models are incorrect. The data shown here is compatible with the idea that a polymerase already associated with a short viral RNA (svRNA) corresponding to a 22-27 nucleotide long 5' vRNA sequence (Perez et al. 2010) is able to bind 3' vRNA and perform replication in *trans*. Copying of vRNA from cRNA could (in theory) assume a similar mechanism, however, currently there is no direct evidence for the existence of short 5' cRNAs.

Unlike transcription, where the newly-synthesised mRNA is excluded from the polymerase as it is being synthesised, in replication, the 5' end of the cRNA remains attached to the active, copying polymerase. To this moment, no quantitative studies focusing on the RNA polymerase interaction with the individual cRNA ends have been performed, but qualitative studies show that the PB1 polymerase subunit binds them equally well (and binds the double-stranded cRNA with a $K_D=70\text{nM}$; González & Ortín 1999b).

2.5 CONTRIBUTIONS

The RNA polymerase used in these experiments was expressed and purified as described in section 1.2.5 by Narin Hengrung (University of Oxford). The Hidden Markov modelling

analysis on data collected to investigate polymerase binding to the 3' viral terminal sequence was performed by Dr. Kristofer Gryte (University of Oxford). The non-influenza RNA strand was provided by Dr. Javier Periz (University of Oxford).

2.6 BIBLIOGRAPHY

- Chi, E.Y. et al., 2003. Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharmaceutical Research*, 20(9), pp.1325–1336.
- Cianci, C., Tiley, L. & Krystal, M., 1995. Differential activation of the influenza virus polymerase via template RNA binding. *Journal of virology*, 69(7), pp.3995–9.
- Desselberger, U. et al., 1980. The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene*, 8(3), pp.315–328.
- Eddy, S.R., 2004. What is a hidden Markov model? *Nature Biotechnology*, 22(10), pp.1315–1316.
- Fodor, E., 2013. The RNA polymerase of influenza A virus : mechanisms of viral transcription and replication. , (2012), pp.113–122.
- Fodor, E., Pritlove, D.C. & Brownlee, G.G., 1994. The influenza virus panhandle is involved in the initiation of transcription. *Journal of virology*, 68(6), pp.4092–6.
- Fodor, E., Seong, B.L. & Brownlee, G.G., 1993. Photochemical cross-linking of influenza A polymerase to its virion RNA promoter defines a polymerase binding site at residues 9 to 12 of the promoter. *The Journal of general virology*, 74 (Pt 7), pp.1327–33.
- González, S. & Ortín, J., 1999a. Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. *Journal of virology*, 73(1), pp.631–637.
- González, S. & Ortín, J., 1999b. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *The EMBO journal*, 18(13), pp.3767–75.
- Hagen, M. et al., 1994. Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity. *J. Virol.*, 68(3), pp.1509–1515.
- Hellman, L.M. & Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols*, 2(8), pp.1849–61.

- Holden, S.J. et al., 2010. Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophysical Journal*, 99(9), pp.3102–3111.
- Jiang, H. et al., 2010. Influenza virus genome C4 promoter/origin attenuates its transcription and replication activity by the low polymerase recognition activity. *Virology*, 408(2), pp.190–6.
- Jung, T.E. & Brownlee, G.G., 2006. A new promoter-binding site in the PB1 subunit of the influenza A virus polymerase. *The Journal of general virology*, 87(Pt 3), pp.679–88.
- Lee, K.H. & Seong, B.L., 1998. The position 4 nucleotide at the 3' end of the influenza virus neuraminidase vRNA is involved in temporal regulation of transcription and replication of neuraminidase RNAs and affects the repertoire of influenza virus surface antigens. *The Journal of general virology*, 79 (Pt 8), pp.1923–34.
- Li, M.L., Ramirez, B.C. & Krug, R.M., 1998. RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. *The EMBO journal*, 17(19), pp.5844–52.
- Mcgeoch, D., Fellnert, P. & Newtont, C., 1976. Influenza virus genome consists of eight distinct RNA species. , 73(9), pp.3045–3049.
- McKinney, S.A., Joo, C. & Ha, T., 2006. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophysical journal*, 91(5), pp.1941–51.
- Perez, J.T. et al., 2010. Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), pp.11525–30.
- Resa-Infante, P. et al., 2011. The influenza virus RNA synthesis machine: Advances in its structure and function. *RNA Biology*, 8(2), pp.207–215.
- Robertson, J.S., 1979. Volume 6 Number 12 1979 Nucleic Acids Research 5'. , 6(12), pp.3745–3757.
- Roy, R., Hohng, S. & Ha, T., 2008. A practical guide to single-molecule FRET. *Nature methods*, 5(6), pp.507–16.
- Sun, W. et al., 2014. U4 at the 3' UTR of PB1 segment of H5N1 influenza virus promotes RNA polymerase activity and contributes to viral pathogenicity. *PloS one*, 9(3), p.e93366.
- Tiley, L.S. et al., 1994. Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *Journal of virology*, 68(8), pp.5108–16.
- De Wit, E. et al., 2004. Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus research*, 103(1-2), pp.155–161.

Chapter 3: Studying the RNA polymerase – promoter interaction

3.1 INTRODUCTION

Each of the eight influenza genomic RNA segments carry conserved and partially complementary sequences at their 3' and 5' ends giving rise to a dsRNA region that plays the role of the viral promoter (referred to as a panhandle; Baudin, Petit, & Ruigrok, 2001). The panhandle is thought to be essential for: (1) binding and stabilising the polymerase, (2) regulation of transcription and replication and (3) vRNA packaging into the RNP. The question of what the possible structure of this double stranded region is has been extensively addressed.

A recent study showed that the two conserved ends of the viral promoter associate to form the double stranded structure even in the absence of the heterotrimer (Noble et al. 2011), but did not shed light on what the precise architecture of this structure might be. A more structural enquiry, employing NMR and investigating both a U4 and a C4 promoter, proposed that non-Watson-Crick base pairing between residues A10 and A11 on the 5' end and residue U10 on the 3' end could give rise to an extended major groove, which would play a role in promoter recognition and binding by the polymerase (Bae et al. 2001; Lee et al. 2003).

Further considering the two partially complementary RNA ends that form the promoter, three regions of interest can be distinguished: the proximal promoter element (comprising of residues 1-9 on both the 5' and 3' end), the distal promoter (residues 11-

16 on the 5' end and 10-15 on the 3' end) and the residue A10 on the 5' end (with no correspondent on the 3' strand) (Flick & Hobom 1999). Depending on the conformation assigned to the proximal promoter, several models explaining the promoter structure when in complex with the polymerase arise (fig. 1): if residues 1-7 are base paired (intermolecular interaction) and residues 8-10 (on the 5' end) and 8&9 (on the 3' end) form bulges, the promoter is thought to be in the duplex panhandle state (fig. 1A; Hsu et al. 1987). If residues 2 and 3 base pair with residues 9 and 8, respectively (intramolecular interaction on each end), forming a tetraloop, the panhandle is in the corkscrew conformation (fig. 1B; Flick et al. 1996). A third possible conformation was proposed, when both ends are single-stranded in the proximal region, but double-stranded in the distal one: this structure is known as a fork (fig. 1C; Fodor et al. 1993; Fodor et al. 1994) and was largely refined by studies showing that the corkscrew 5' & 3' end loops are essential for the polymerase's endonuclease activity (Leahy et al. 2001a; Leahy et al. 2001b), the 5' end loop alone is required for polyadenylation (Pritlove et al. 1999) and that the polymerase is stabilised by interaction with the promoter in the corkscrew conformation (Brownlee & Sharps 2002).

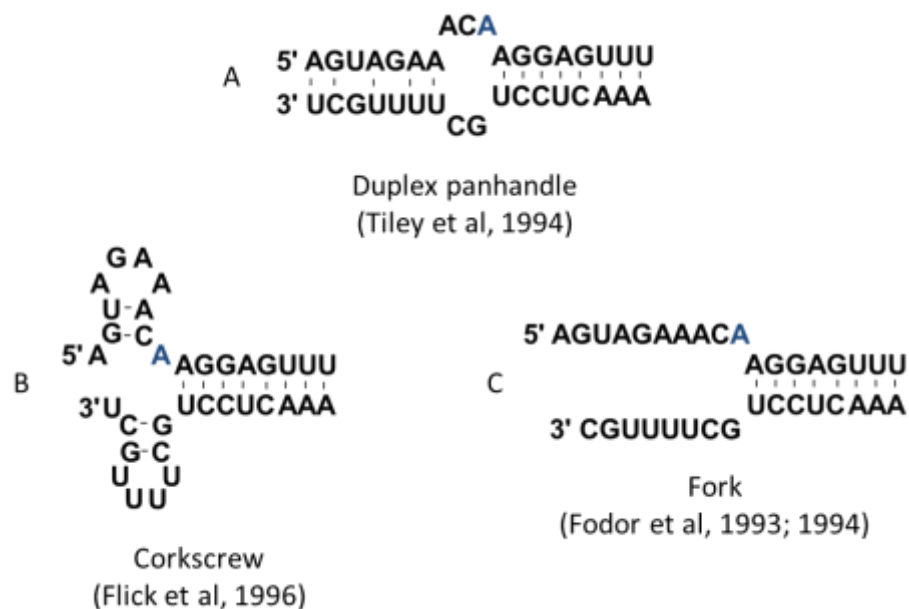


Fig. 1 – An overview of the different models proposed to describe the structure of the viral promoter when the polymerase binds to the viral dsRNA: A) duplex model; B) corkscrew model and C) fork model.

The polymerase is thought to contact the double stranded promoter at residues 8-11 on the 5' end and 1-4 & 9-12 on the 3' end (Bae et al. 2001; Fodor et al. 1993; Tiley et al. 1994). Several studies have addressed the issue of how and where in the PB1 subunit the double-stranded RNA promoter (or its individual terminal sequences) binds (Li et al. 1998; González & Ortín 1999b; González & Ortín 1999a; Jung & Brownlee 2006), but only one of them considered this interaction from a quantitative point of view, suggesting an apparent binding constant for the ds vRNA of 20nM (González & Ortín 1999a).

In this chapter, I present experimental results that lead to a quantitative and structural understanding of the polymerase-promoter interaction, allowing me to reconcile the proposed promoter models with the changes actually occurring in the promoter when it is bound by the polymerase, observed in real-time.

3.2 MATERIALS AND METHODS

3.2.1 RNA polymerase

The recombinant A/Northern Territories/60/68 RNA polymerase (H3N2) used in these experiments was expressed and purified from Sf9 cells as described in Chapter 1.

3.2.2 RNA constructs

The single-stranded, complementary ends of the viral promoter were synthesised and labelled as described in Chapter 2, after which they were annealed in annealing buffer (20mM Tris-HCl pH 8.0, 500mM NaCl, 1mM EDTA) in a thermocycler at 94°C for 1 minute, 90°C for 1 minute, 85°C for 1 minute, 75°C for 2 minutes, 65°C for 5 minutes, 45°C for 10 minutes, 25°C for 5 minutes and 4°C for 3 minutes.

The wild type double-stranded RNA constructs used in this chapter are overviewed in table 1.

Construct	Sequence and labelling position
5'U18(Cy3)/3'U1(ATTO647N)	<pre> Cy3 5' AGUAGAA ACA AGGAGUUU 3' UCGUUUU CG UCCUCAAAUUU – biotin ATTO647N </pre>
5'U18(Cy3)/3'U4(ATTO647N)	<pre> Cy3 5' AGUAGAA ACA AGGAGUUU 3' UCGUUUU CG UCCUCAAAUUU – biotin ATTO647N </pre>
5'U3(Cy3)/3'U4(ATTO647N)	<pre> Cy3 5' AGUAGAA ACA AGGAGUUUUUUGAACAAACUACUU – biotin 3' UCGUUUU CG UCCUCAAAUUU – biotin ATTO647N </pre>

Table 1: Wild type, double-stranded RNA constructs used in this chapter. The notation used to denote the construct indicates 1) the end of the viral promoter (5' or 3'); 2) the position on the strand of the uracil residue carrying the label; 3) the fluorescent dye used.

Experiments were also carried out using mutant dsRNA promoters (table 2).

Construct	Sequence and labelling position
m5'U18(Cy3)/m3'U4(ATTO647N)	<pre> Cy3 5' AAUAGAU AAA AGGAGUUU 3' UUGUUUA CC UCCUCAAAUUU – biotin ATTO647N </pre>
m5'U3(Cy3)/m3'U4(ATTO647N)	<pre> Cy3 5' AAUAGAU AAA AGGAGUUU 3' UUGUUUA CC UCCUCAAAUUU – biotin ATTO647N </pre>

Table 2: Mutated, double-stranded RNA constructs used in this chapter (the notation is the same as described above). Mutated residues are shown in orange.

3.2.3 Electrophoretic Mobility Shift Assay

20nM of dsRNA (Cy3/ATTO647N-labelled) was incubated on its own or in the presence of increasing concentrations of protein in T8 buffer (50mM Tris-HCl pH 8.0, 100mM KCl,

10mM MgCl₂, 100μg/μl BSA, 1mM DTT and 5% glycerol) supplemented with 500mM NaCl and 2U RNasin (40U/ μl, Promega) and incubated at 28°C for 15 minutes. Reactions were stopped by addition of 1x reaction volume of 50% glycerol, after which samples were loaded on a 1% TBE-agarose gel which was run for 2 minutes at 150V and then for 1 hour at 70V. The gel was analysed on a BioRad Molecular Imager PharosFX system as described previously in section 2.2.2.

3.2.4 Polymerase activity assay (*in vitro* transcription)

In order to determine whether the position of the dye interferes with polymerase binding, the capacity of the protein to transcribe RNA was tested: 1.5ul RNAP was incubated with 10nM unlabelled and labelled, double-stranded vRNA promoter (final concentration) in 10mM MgCl₂ and 2mM DTT supplemented with 1.87U RNasin (40U/μl, Promega), 1mM ApG and 50μM:25μM:25μM ATP:CTP:UTP mix, in the presence of 0.3μCi/μl α-³²P GTP (10μCi/μl, Perkin Elmer) and incubated for 2h at 30°C. The transcription reaction was stopped by addition of 10μl loading dye (80% formamide, 10mM EDTA, 0.04% bromophenol blue and 0.04% xylene cyanol FF) and denatured by heating at 95°C for 5 minutes, then electrophoresed on a 6M urea/25% polyacrylamide (19:1 acrylamide:bis-acrylamide) sequencing gel for 3h at 1200V. The products of the reaction were visualised by autoradiography.

3.2.5 Binding assay (single-molecule fluorescence spectroscopy)

Measurements were performed on a custom-built confocal microscope allowing alternating-laser excitation of donor and acceptor fluorophores (Kapanidis et al. 2005), at

average excitation intensities of 250 μ W at 532nm and 60 μ W at 635nm. A dsRNA construct, labelled with ATTO647N and Cy3 was incubated with or without varying concentrations of RNA polymerase (2nM, 5nM, 10nM, 50nM final concentration) in T8 buffer (50mM Tris-HCl pH 8.0, 100mM KCl, 10mM MgCl₂, 100 μ g/ μ l BSA, 1mM DTT and 5% glycerol) supplemented with 500mM NaCl, for 15 minutes at 28°C.

The sample was diluted in T8 buffer (final concentration of the labelled dsRNA: 100pM) and deposited on a glass coverslip (previously incubated with 1.5mg/ml BSA in PBS), after which data was collected. Alternatively, in the experiments described in section 3.3.7 ApG to a final concentration of 1mM was added to the T8 buffer used for imaging.

A custom software package was used to assign fluorescence photons to either donor or acceptor-based excitation depending on their arrival time. Fluorophore stoichiometry and the FRET efficiency (E^*) were calculated for each burst above the minDA=30 threshold, yielding a two-dimensional E^*/S histogram; one-dimensional FRET distributions characteristic for donor-acceptor species were obtained by applying a $0.4 < S < 0.8$ threshold, while the mean E^* value was calculate by fitting the FRET distributions with a Gaussian function. Mean E^* efficiencies were used to determine accurate FRET efficiencies as described in the Introduction to single-molecule fluorescence section of this thesis.

3.2.6 RNA secondary structure prediction

Secondary structures for each of the constructs were determined using the DuplexFold, bifold and/or Fold modules available on the University of Rochester RNAstructure webserver (<http://rna.urmc.rochester.edu/RNAstructureWeb/>; Reuter & Mathews 2010).

Given the proposed models for the influenza dsRNA promoter, certain residues were forced base-paired; the two independent terminal sequences of the promoter were joined in one molecule, allowing them to be modelled as a hairpin, rather than two independently interacting molecules as the software used cannot represent two independent, interacting RNAs). Files were downloaded in the ‘.ct’ or ‘.dot’ format (converted to the appropriate format using the ct2dot or dot2ct modules on the RNAstructure webserver).

3.2.7 2D to 3D conversion

To construct 3D models of the RNAs of interest, the secondary structure information determined previously was used in the Chimera plugin Assemble2 (Jossinet et al. 2010). Based on the secondary structure of the construct of interest, each individual residue or intramolecular structure (loop, helix) was represented and once the whole molecule was translated in 3D space, the different building blocks were linked into one single molecule, which was exported from Assemble2/Chimera as a ‘.pdb’ file.

3.2.8 Dye positioning

Using the ‘.pdb’ file generated for each construct, we identified the number of the atoms in the structure to which the dyes are linked (as a general rule, the dye is attached to a uridine on the C5’ of the sugar residue); the dyes were previously modelled in ChemDraw and their features are given in table 3.

Fluorophore	<u>Cy3</u>	<u>ATTO647N</u>
Linker length	14.2nm	17.8nm

Linker width	4.5nm	4.5nm
Radius 1 (x)	8.2nm	7.4nm
Radius 2 (y)	3.3nm	4.8nm
Radius 3 (z)	2.2nm	2.6nm

Table 3: Properties of the two dyes as determined in ChemDraw (Timothy Craggs, University of Oxford). The different radii describe the length, width and depth of each dye, and are necessary for determining the spatial dimensions of the fluorophores, which are not well approximated by a sphere (Kalinin et al. 2012).

To determine the different positions the dye can assume relative to the previously modelled 3D RNA structure, a FRET positioning and screening software (FPS; Kalinin et al. 2012) was used. Once the coordinates of the centre of the dye and the dye cloud (signifying all the possible positions occupied by the dye) were determined, they were exported as a '.xyz' file, which can be loaded together with the RNA structure in PyMol, for visualisation.

3.2.9 Determining distances in the 3D RNA structures

The '.pdb' file of each of the conformations investigated was opened in PyMol alongside the corresponding '.xyz' files for the dyes attached to it; the centre of the dye was represented as a sphere. Distances between the centres of the dye pairs involved in FRET were determined using the standard 'measure distance' function available in PyMol.

3.3 RESULTS

3.3.1 Is binding influenced by the fluorophore position on the dsRNA construct?

As this study relies on fluorophore reporters to characterise the RNA polymerase – RNA promoter interaction, it's advisable to determine whether the position of these fluorophores on the RNA construct might interfere with binding. To perform such a check, an activity assay for the polymerase transcribing from the RNA promoters used was employed.

The level of transcription from each of the constructs studied was assessed against that of an unlabelled influenza dsRNA promoter of the same length as those used further in the study (fig. 2 – lane 1). All of the samples show transcription products, but judging from the intensity of the bands, the reaction is not as productive when the promoter is labelled. A singly-labelled promoter, with the dye at position 3 on the 5' end seems to be the least perturbed (lane 3), followed by a doubly-labelled construct with the dyes at positions 18 (5' end) and 1 (3' end; lane 2). Compared to these results, a fluorophore placed at position 4 on the 3' end seems to have the highest negative impact on transcription (lane 4). Finally, a short transcription product was observed when the assay was carried out in the presence of ApA dinucleotide and using the unlabelled mutant promoter (G2->A, A7->U, C9->A on the 5' end strand and C2->U; U7->A; G9->C on the 3' end strand, these mutations are believed to introduce little disruption when the promoter is in the duplex conformation, but prevent formation of a stable corkscrew structure when bound by the polymerase) as a substrate (fig. 2 – lane 5).

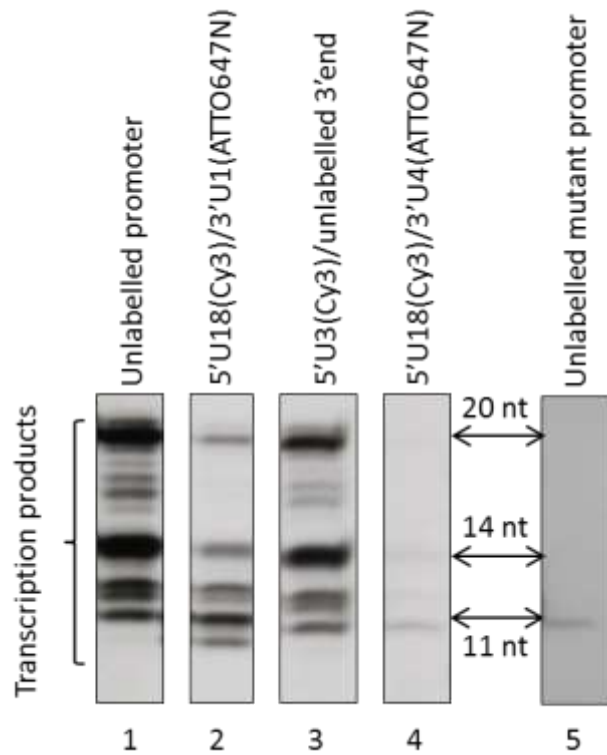


Fig. 2 – *In vitro* transcription assay assessing the activity of the polymerase when it copies the differently labelled RNA constructs used in this study. The reactions using wild type promoter (lanes 1-4, from the left) were set up in the presence of ApG initiating dinucleotide; for the mutant promoter sample ApA was used.

To conclude, based on the transcription profiles observed, all the labelled constructs are all bound by the polymerase.

3.3.2 Qualitative investigation of the polymerase interacting with the dsRNA promoter

Although not useful for a quantitative characterisation of this interaction, an EMSA assay allows for a qualitative study of the polymerase binding to the RNA. In the presence of increasing protein concentrations the doubly-labelled dsRNA used in the assay shows a decreased mobility compared to its mobility in the absence of the protein (fig. 3). The detected intensity of the bands in the gel decreases with increasing protein concentration, which is due to the fact that some of the protein-RNA aggregates,

remaining trapped in the well. Alongside the protein/dsRNA analysis, two other strands are used in the experiment, both with and without polymerase, for comparison purposes: these are the individual termini of the viral genes and I've shown that they are capable of being bound by the polymerase stably (5' end) and unstably (3' end, labelled at position 1), exhibiting different mobilities. These ssRNAs can be used as standards: in the case when the dsRNA is not annealed, the expectation is to detect two separate RNA/RNAP species that would exhibit the same mobility as the ssRNA/RNAP complexes (the dsRNA is an annealed construct of the two ssRNAs used here).

Comparing the mobility of the RNAs in the absence of the protein in fig. 3, two observations stand out: 1) that the ATTO647N-labelled RNA is slightly more retarded on the gel than the Cy3-labelled one, which is expected since the ATTO647-labelled strand is 20nt long and carries a biotin (compared with the Cy3-labelled 5' end, which is 18nt long and non-biotinylated) and 2) that a high proportion of the RNA in the 'dsRNA' lane is actually ssRNA; this does not exclude the fact that some of the RNA detected is double-stranded. In the lanes showing the dsRNA/RNAP complex, the band identified is detected as yellow, suggesting equal amounts of the two strands are present; also, the band is present when the gel is visualised using 532nm excitation and detecting 647nm emission (the equivalent of donor excitation/acceptor emission associated with FRET).

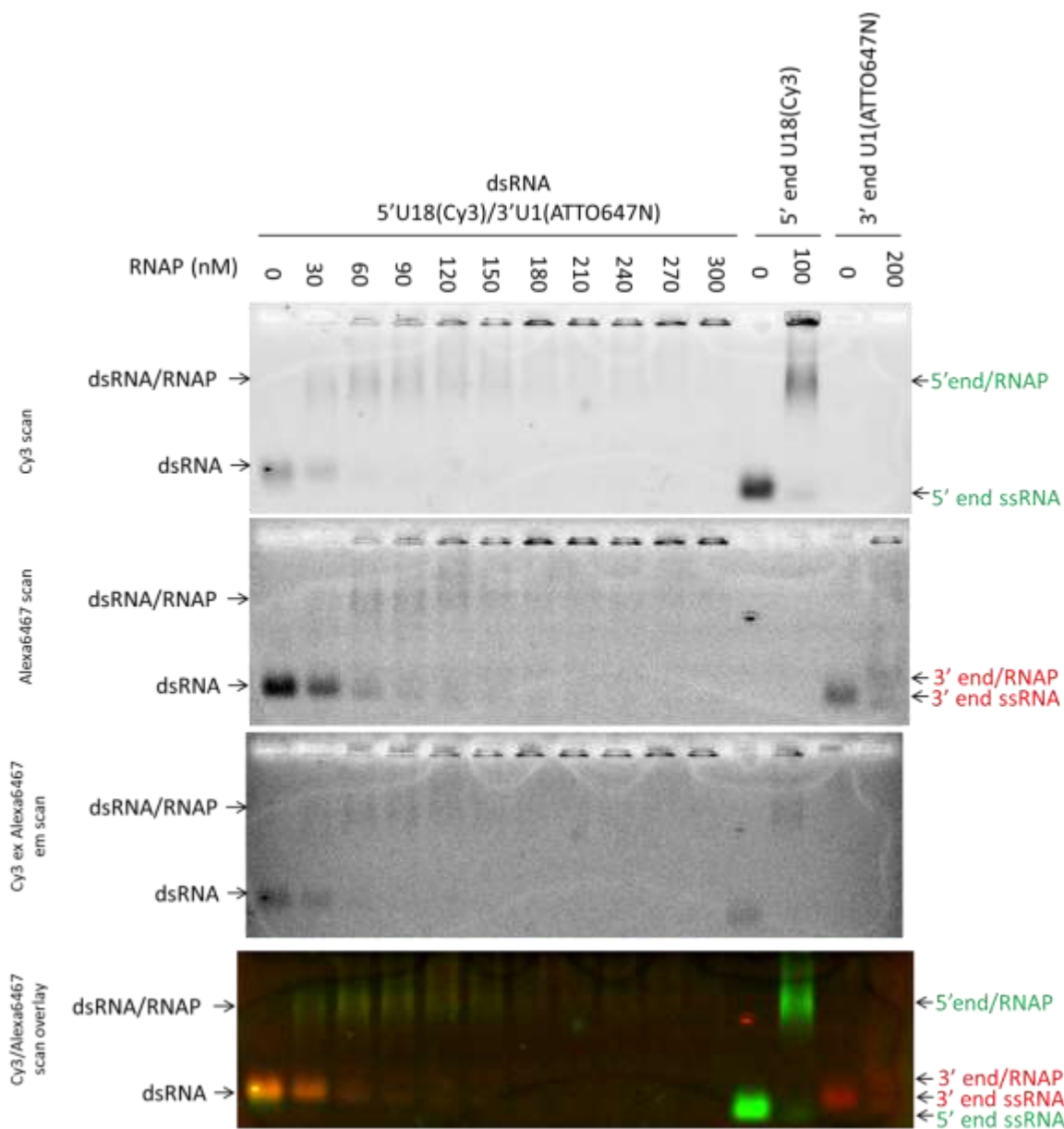


Fig. 3 – Electrophoretic mobility shift assay showing that the dsRNA is retarded when incubated with increasing concentrations of protein; for the purpose of comparison, the individual single stranded RNAs at are incubated with 100nM (5' end) or 200nM (3' end) RNA polymerase. Three individual scans are shown here: one detecting the green dye (Cy3 scan; top panel), one detecting the red dye (Alexa647 scan; second panel) and one looking for the FRET between the two (Cy3 excitation/Alexa647 emission; third panel); the fourth image is a composite of the scans taken to probe the presence of the green and the red dyes. All the RNAs used in the assay are at a 20nM concentration; the 3' end used in this assay is the U4 promoter with the ATTO647N fluorophore

attached at position 1 (U1).

This experiment suggests that the polymerase could be binding to the dsRNA and opens the way for further investigation.

3.3.3 Structural investigation of the polymerase-dsRNA interaction

To further characterise this interaction, I opted for a more sensitive single-molecule Förster resonance energy transfer (smFRET) assay looking at doubly-labelled molecules diffusing in solution (with the donor fluorophore (Cy3) on the 5' strand and the acceptor (ATTO647N) on the 3' strand of the vRNA panhandle). In this assay, FRET is used as a ruler and its distance-dependence is exploited to understand whether the panhandle undergoes structural changes when it is in complex with the polymerase. A set of RNA constructs that are identical in sequence, but have different labelling schemes were employed, as detailed in section 3.2.2.

The wild type 5'U18(Cy3)/3'U4(ATTO647N) labelled promoter (fig. 4A) was detected as a single species with a mean apparent FRET of 0.57 (fig. 4B). Titration of increasing amounts of polymerase (2-50nM) gives rise to a bimodal FRET distribution (fig. 4C-G), indicating the occurrence of two species: one which can be attributed to the unbound RNA and another which is assigned to the dsRNA in complex with the polymerase. As the FRET value corresponding to the free RNA was already determined, it can be fixed, allowing the second peak to be freely fitted with a Gaussian (in green; fig. 4C-G). Treating the data in this manner, allows for determination of the FRET value for the RNA-protein complex ($E^* \sim 0.79$). The affinity of the polymerase for the dsRNA construct is high, with the

complex being clearly visible even at protein concentrations as low as 2nM; at higher concentrations of protein most of the RNA is in the bound form.

Investigating the mutant m5'U18(Cy3)/m3'U4(ATTO647N) labelled promoter (fig. 4H) in the same manner, reveals the unbound RNA peak to be characterised by a mean FRET value $E^* \sim 0.59$ (fig. 4I). Adding 2nM and 5nM RNAP to the mutant RNA (fig. 4J&K, respectively), did not give rise to any change in the FRET distribution, a sign that the RNA-RNAP complex is not formed; only upon addition of 10nM and more RNAP, a bimodal FRET distribution can be detected where the $E^* \sim 0.8$ species is indicative of the RNA-RNAP complex (fig. 4L-N). While addition of higher concentration of polymerase leads to increased complex formation, under these experimental conditions, the bound mutant RNA does not become the dominant species in the sample, suggesting that, the affinity for this construct is lower than for the wild type panhandle.

The wild type (fig. 4A) and mutant (fig. 4H) RNA constructs used in this experiment have the same length and carry the fluorophores at the same positions: since the E^* value for the wild type RNA ($E^* \sim 0.57$) in the unbound state is the same as the one for the mutant RNA ($E^* \sim 0.59$), it can be concluded that the two promoters share the same initial panhandle structure.

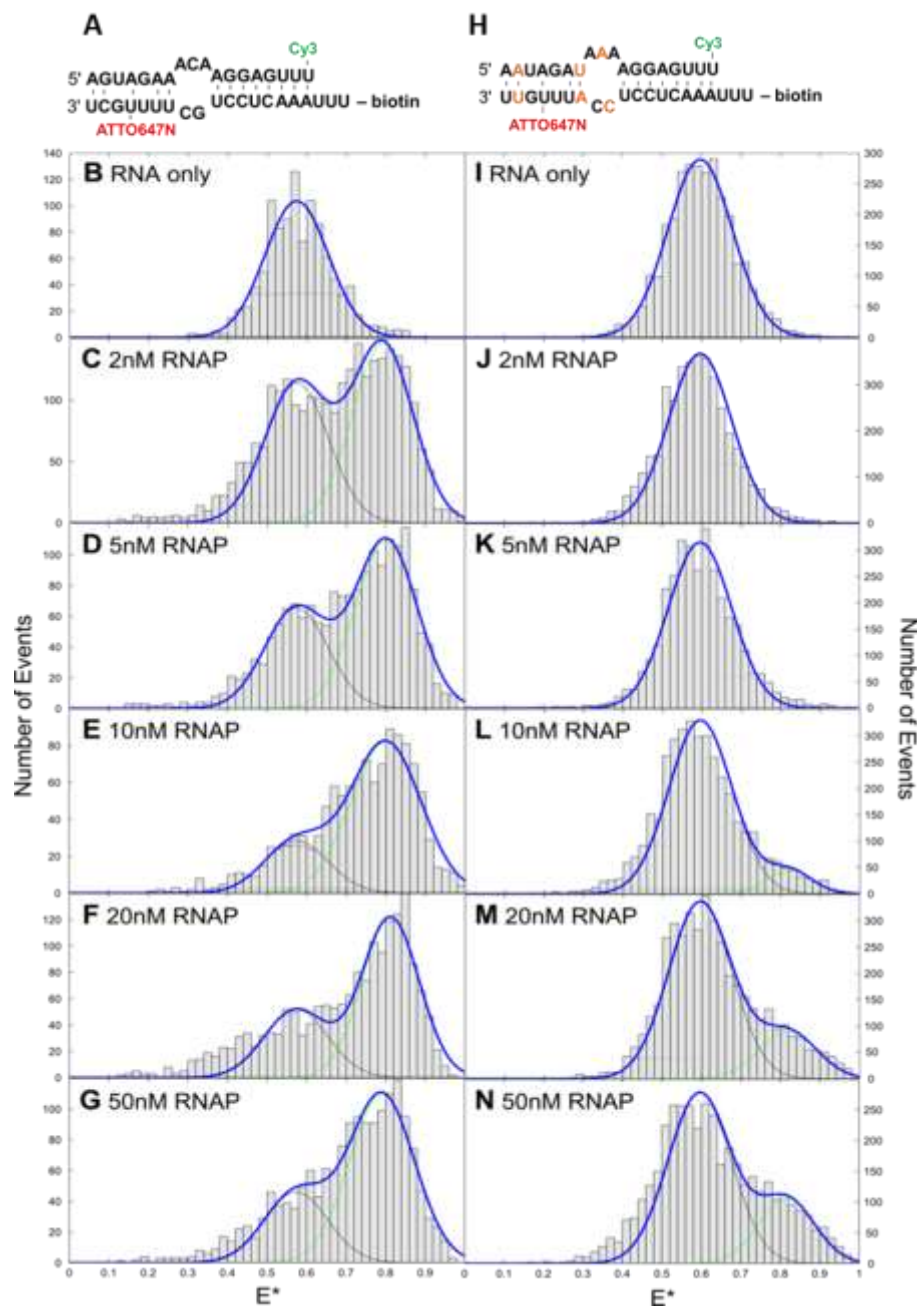


Fig. 4 – Detecting polymerase binding to dsRNA promoter: A) 5'U18(Cy3)/3'U4(ATTO647N) wild type construct was incubated with increasing polymerase concentrations after which the doubly-labelled diffusing molecules were investigated using single-molecule FRET spectroscopy combined with alternating-laser excitation (ALEX). Ratio E^* represents the uncorrected FRET efficiency, and curves were fitted with Gaussian functions to determine the centre of the distributions; the peak assigned to the RNA only species was fixed at mean $E^*=0.57$, while the peak corresponding to the

RNAP/RNA complex was fitted freely (panels B-G). H) A mutant RNA construct with the same labelling scheme was used as a control: the FRET peak for the RNA only species was fixed at $E^*=0.59$ (mutant), while the complex peaks were fitted freely (I-N).

An identical construct in sequence, but carrying the ATTO647N-label at position 1 on the 3' end (rather than position 4; fig. 5A) shows similar behaviour both in the absence (fig. 5B) and in the presence of the protein (fig. 5C-G), suggesting that the position of the dye is not interfering with the binding.

Another way of assessing whether the high FRET species is indeed the dsRNA in complex with the polymerase, is by looking at the fluorophore stoichiometries (S), a measure that is directly related to the D:A ratio ($S \sim 0.5$ for a species with D:A=1:1) and implicitly, to the intensity of the dyes.

It was recently shown that dyes undergoing *cis-trans* isomerisation can have their fluorescence enhanced if their local environment changes, for example when they are in the proximity of a protein; this is due to reducing the rate of isomerisation to the photoinactive *cis* isomer and stabilising the *trans* state. This is referred to as protein induced fluorescence enhancement (PIFE, Hwang et al. 2011) and can be used to monitor changes in S resulting from the dye's increased quantum yield.

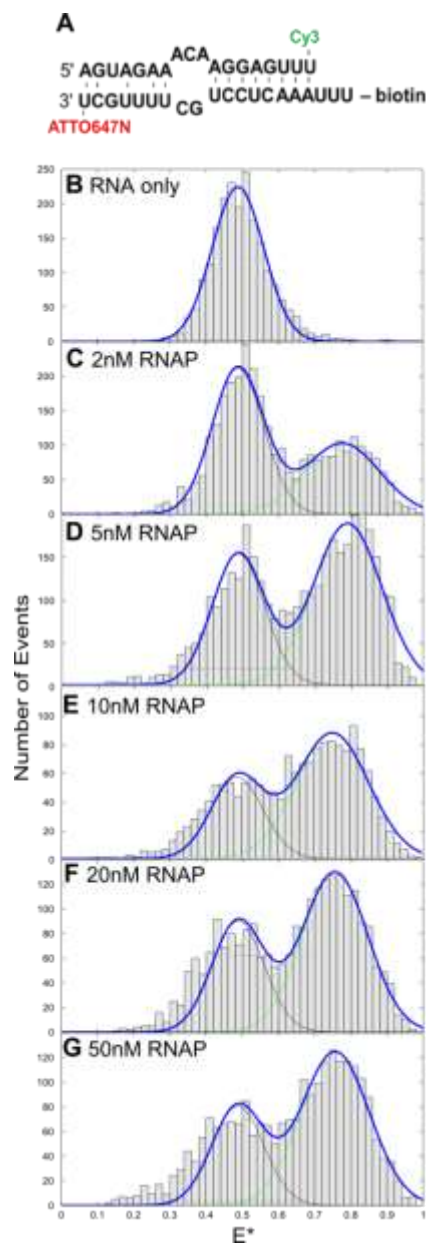


Fig. 5 – A) 5'U18(Cy3)/3'U1(ATTO647N) construct used to investigate polymerase binding to the dsRNA promoter; when measured on its own, the RNA construct is characterised by a mean FRET value of 0.49 (B). Adding increasing concentrations of RNAP gives rise to an RNAP/RNA species characterised by higher FRET (peaks for the complex in panels C-G are freely fitted). As before, the RNA only peak is fixed and the FRET peak attributed to the binary complex is freely fitted using a Gaussian function.

Both wild type and mutant dsRNA promoter constructs used in this study carry a Cy3 on one of the strands, which is an organic fluorophore known to be amenable to PIFE. Analysing the two-dimensional E^*/S histogram of the free dsRNA [wt, 5'U18 (Cy3)/3'U4 (Atto647N), in the absence of the protein], a characteristic stoichiometry value for the construct can be extracted, and is $S \sim 0.45$ (fig. 6A). Upon addition of protein (50nM), the stoichiometry of the detected species increases to $S \sim 0.6$ (fig. 6B), an increase that can be

attributed to PIFE of the Cy3 dye, and hence is another indication that what we detect is a protein-RNA interaction. The unbound mutant RNA [5'U18 (Cy3)/3'U4 (Atto647N)] shows a $S \sim 0.6$ (fig. 6C) that remains unchanged upon addition of 50nM polymerase (fig. 6D), suggesting that, as expected, the protein does not bind the mutant dsRNA construct.

While the FRET efficiency distribution suggests there are two populations present in the 50nM RNAP sample (for both the wild type and the mutant promoters), only one S peak can be detected. This is wider than the one seen for the unbound species and is centred so that it reflects the dominant species within the sample: bound RNA ($S \sim 0.6$, fig. 6B), in the case of the wild type construct and free RNA ($S \sim 0.6$, fig. 6D), for the mutant.

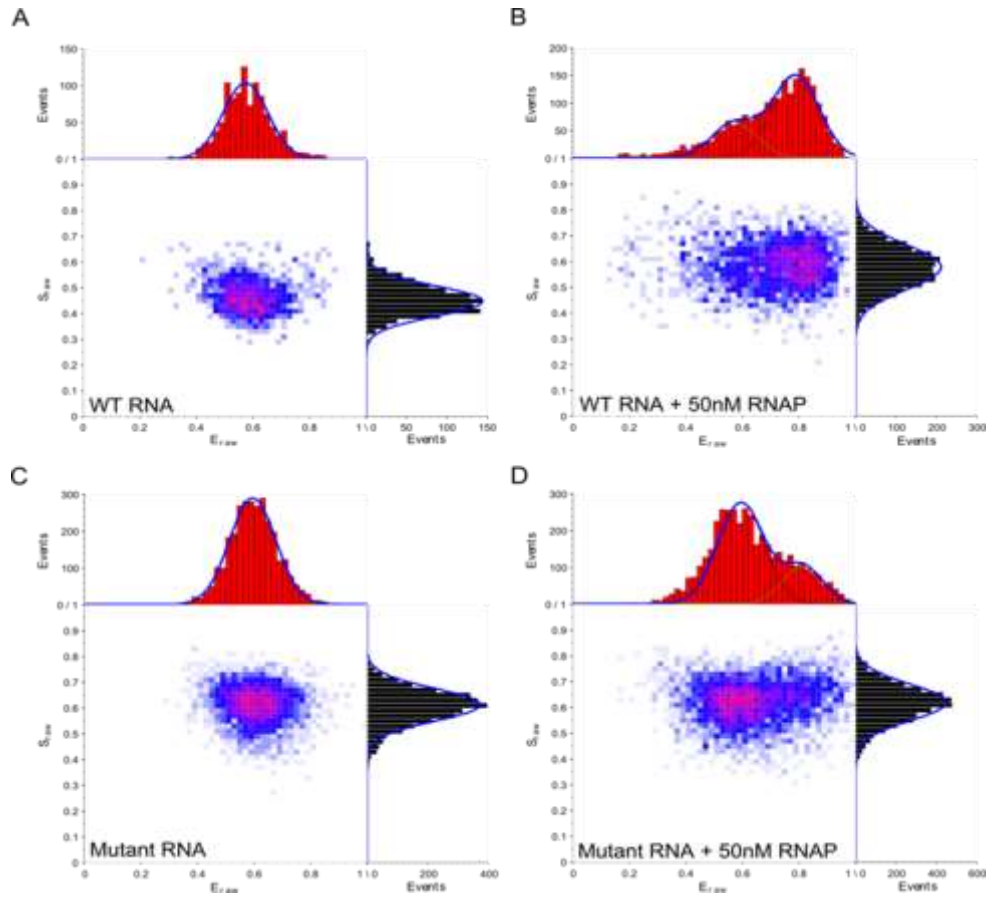


Fig. 6 – E^*/S histograms for the wild type (A) and mutant (C) 5'U18(Cy3)/3'U4(ATTO647N) RNAs in the absence (A, respectively C) and presence of 50nM RNA polymerase (B, respectively D); fluorescence photons were assigned to either donor- or acceptor-based excitation with respect to their photon arrival time and two ratios, fluorophore stoichiometries (S) and apparent FRET efficiencies E^* were calculated for each fluorescent burst yielding a two-dimensional histogram (dual channel burst search, minDA=30).

The data presented so far supports a structural change in the promoter upon polymerase binding, yet it is not possible to distinguish between a structural change in the proximal promoter region (bp 1-10) or in the distal proximal region (bp 11-18).

3.3.4 K_D determination

Using this binding data, the percentage of bound RNA relative to the total RNA available can be quantified, at each protein concentration; these values can be plotted in order to determine the binding constant K_D for each of the wild type promoters used. Since the two constructs have the same sequence and form the same panhandle structure when free, no major differences in binding are expected to occur, and any such differences could be attributed to acceptor fluorophore on the 3' strand interfering with RNA-RNAP complex formation. Plotting the data collected for the 5'U18(Cy3)/3'U4(ATTO647N), promoter a binding constant value $K_D=0.35\text{nM}$ can be determined (fig. 7A); doing the same for the 5'U18(Cy3)/3'U1(ATTO647N) construct, gives a value $K_D=0.7\text{nM}$ (fig. 7B). The values are very similar and the slight difference could be explained by the fact that the presence of the fluorophore at position 1 has a slight detrimental effect on binding.

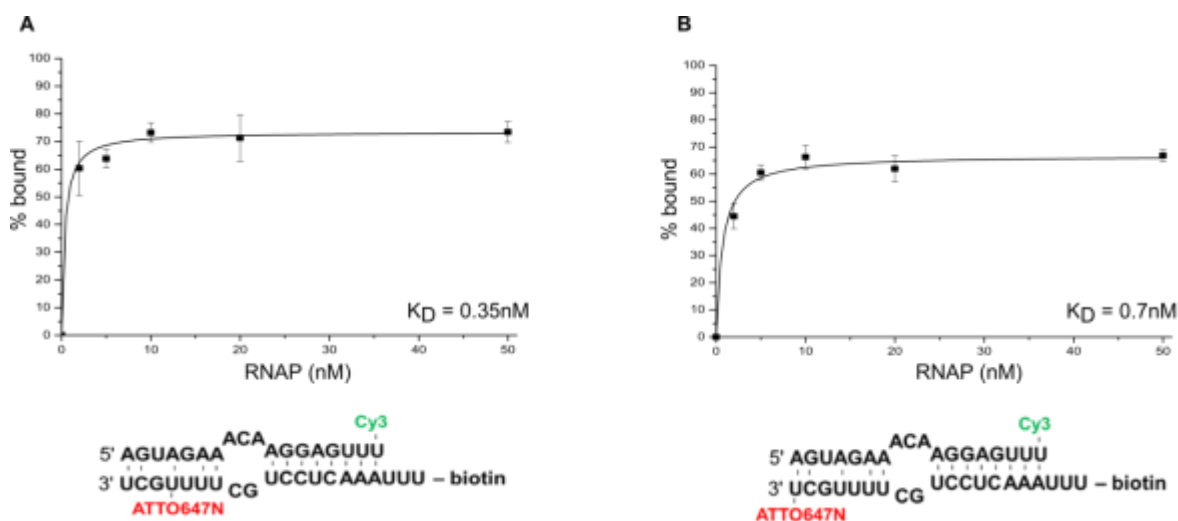


Fig. 7 – A) Binding curve for the 5'U18(Cy3)/3'U4(ATTO647N) dsRNA construct suggests the binding constant for the interaction is 0.35nM. B) The polymerase shows a lower affinity for the

5'U18(Cy3)/3'U1(ATTO647N) construct, with the binding constant determined to be 0.7nM.

Unfortunately, due to the fact that saturation isn't reached when the mutant construct is used, a value for this interaction's binding constant could not be determined.

3.3.5 Analysis of binding to a double-stranded vRNA promoter using single molecule quenchable FRET (quFRET)

Quenchable FRET has been previously described as a promoter opening assay (Cordes et al. 2010). In brief, when the two reporting fluorophores are attached to different strands of the nucleic acid investigated, but in close proximity to each other (<3nm), they are likely to undergo contact-induced quenching, which suppresses fluorescence emission. When the two strands move apart relative to each other (such as in promoter opening), leading to an increase in distance between the donor and acceptor, the fluorescence is un-quenched and energy transfer between the two dyes is re-established; as a result, a FRET signal can be detected. In this work, quFRET is employed to investigate whether the observed structural change brought about by the polymerase binding to the viral RNA promoter occurs in its proximal or distal region. If the structure changes in the distal region, the FRET pair would remain quenched over the course of the titration and no signal would be detected; however, if the change occurs in the proximal region (as suggested by both the corkscrew and fork models), a FRET signal should be present.

For this assay, both a wild type and a mutant promoter are employed: the Cy3 donor on the 5' end is attached to the residue at position 3, while the ATTO647N acceptor remains attached on 3'U4). The wild type construct [5'U3(Cy3)/3'U4(ATTO647N) labelled; fig. 8A]

was investigated first, and no fluorescence could be detected initially (fig. 8B); when adding the protein, already in the presence of as little as 2nM RNAP, a population characterised by a high FRET value ($E^* \sim 0.8$; fig. 8C) could be identified. Addition of more protein leads to this population increasing, which suggests that more complex is formed (fig. 8D-G). This result is proof that the conformational change induced by the polymerase binding the dsRNA promoter occurs in the proximal promoter region.

Compared to the quFRET wild type strand discussed above, the mutant RNA promoter [m5'U3 (Cy3)/m3'U4 (ATTO647N) labelled; fig. 8H] shows very little complex formation and when this occurs, it is in the presence of at least 20nM RNAP (fig. 8M-N).

3.3.6 Distance determination and 3D modelling

It has been previously shown that FRET data acquired under alternating laser excitation can be used for generating structural information about the molecule investigated (Lee et al. 2005). As several possible conformations have been proposed for the viral promoter interacting with the RNAP, an investigation on whether the FRET data collected here can yield enough structural information as to permit distinguishing between the fork and the corkscrew models proposed (the results presented above support a structural change in the bound dsRNA which contradicts the idea that the promoter remain in the duplex structure proposed in Tiley et al. 1994).

To achieve this, the experimentally-determined apparent FRET values have to be converted to accurate FRET values by correcting for background, cross talk and γ -factor (a combined measure of differences in quantum yield and detection efficiency; a more detailed explanation of how accurate FRET values are obtained can be found in the

Introduction to single-molecule fluorescence section at the beginning of this thesis); once accurate FRET is determined for each of the D-A pairs, we can use it to find the distance between these fluorophores (table 4).

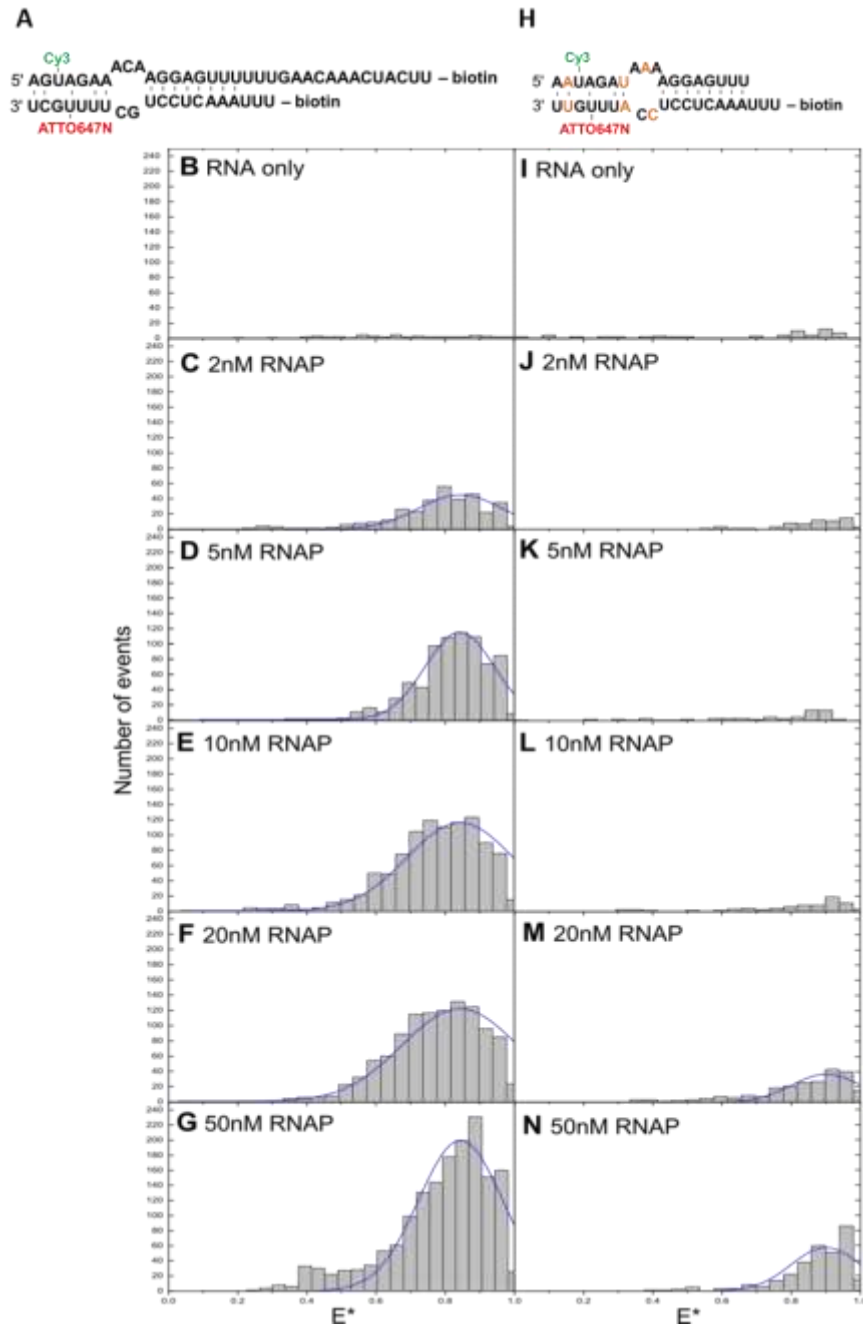


Fig. 8 – Detecting polymerase binding with qFRET. The initial experiment was done using a wild type 5'U3(Cy3)/3'U4(ATTO647N) promoter construct on which the 5' end is longer than the 3' end and both carry a biotin; the reason why this construct is different from the other constructs used is that the 5' end was initially meant for experiments looking at polymerase binding to single-stranded RNA on the surface (A). When no polymerase is added, no FRET is detected (B); a high FRET peak ($E^*=0.84$) is detected when polymerase is added (C-G), suggesting binding occurs. The

same experiment was carried out with a mutant labelled in the same manner (H), which also showed no FRET when imaged on its own (I), but showed a high FRET peak when incubated with the polymerase, albeit at higher concentrations of protein than the wt RNA (J-N). The detected FRET peaks shown here were freely fitted using a single Gaussian function.

	Labelling scheme	Corrected FRET efficiency (a.u)	Determined distance (nm)
RNA only	5'U18(Cy3)/3'U1(ATTO647N)	0.42±0.03	6.25±0.14
	5'U18(Cy3)/3'U4(ATTO647N)	0.55±0.03	5.69±0.10
	5'U3(Cy3)/3'U4(ATTO647N)	n/a	3.00±0.00*
RNA+RNAP	5'U18(Cy3)/3'U1(ATTO647N)	0.68±0.01	5.19±0.05
	5'U18(Cy3)/3'U4(ATTO647N)	0.82±0.03	4.55±0.15
	5'U3(Cy3)/3'U4(ATTO647N)	0.82±0.03	4.57±0.17

Table 4 – Distances determined by taking into account the corrected FRET efficiency, for each of the RNA strands studied, both with and without the protein. (*this value could not be determined experimentally since the fluorophores are quenched; the value of 3nm given here is an approximation based on the 1) distance across a base pair and 2) the calculated linker lengths for the dyes, given in section 3.2.8.)

But do these experimentally determined distances correspond to any of the proposed promoter structural models? Since no 3D-representation of the panhandle or the corkscrew exists, a natural next step was to model these structures *ab initio*.

First, starting from the predicted secondary structure for the unbound RNA panhandle (Tiley et al. 1994; fig. 9A, shown as a hairpin), a 3D reconstitution of this construct was produced (fig. 9B). To check the validity of the generated structure, it was superimposed with the only available 3D structure of the influenza promoter (generated using NMR, available in the RCSB Protein Data Bank as 1JO7): for the alignment of the two structures,

the RMS error was 0.25nm; as this error is so small, it can be concluded that the two compared structures were in agreement (fig. 9C).

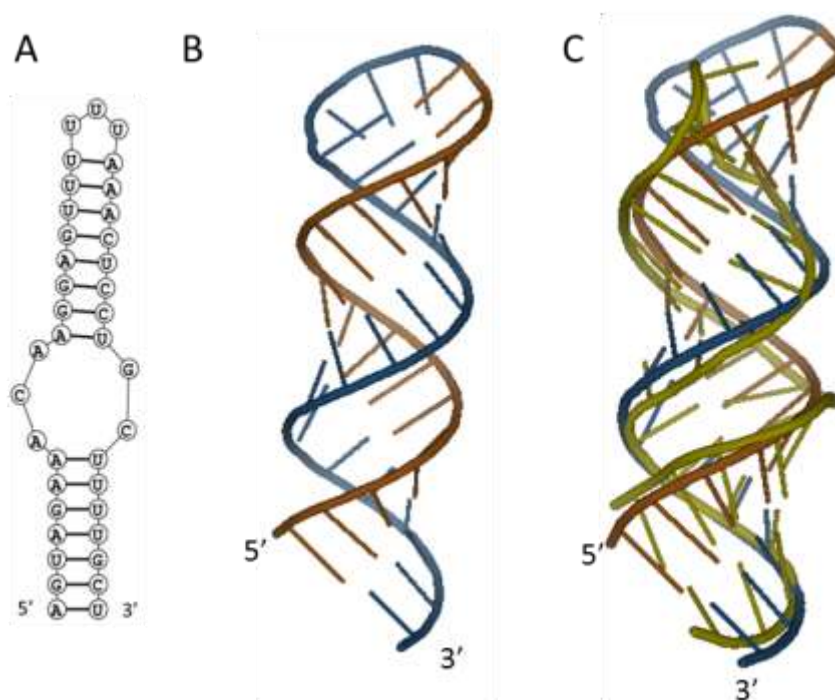


Fig. 9 – A) The 5' end and 3' end RNA strands used in this study were joined to form a hairpin structure whose secondary structure was determined using the BiFold module of the RNAstructure web-based software. B) The Assemble2 3D generated model for the dsRNA construct shown in A: in blue, the 3' end which corresponds to the first 20nt of the neuraminidase gene; in brown, 5' end (18nt). C) Superposition of the modelled structure shown B with the structure of the U4 panhandle determined by NMR (in olive, PDB: 1JO7; Bae et al. 2001). The hairpin modelled here consists of 38nt (from B), while the structure used for comparison is only 31nt long; however, the end regions of both structures, which are in fact the conserved sequences of the viral panhandle, agree well, with an RMS of 0.2nm (as determined in PyMol).

The generated unbound RNA promoter structure was then used as a scaffold for the dyes; once the mean position of the dye was determined for each residue of interest, the distances between the two partners in the FRET pair could be calculated (fig. 10A):

5'U18(Cy3)/3'U1(ATTO647N) are located 6.23nm apart, 5'U18(Cy3)/3'U4(ATTO647N) 5.02nm apart and 5'U3(Cy3)/3'U4(ATTO647N), 3.05nm; these values are in good agreement with the ones determined experimentally, using corrected FRET efficiencies (fig. 10B).

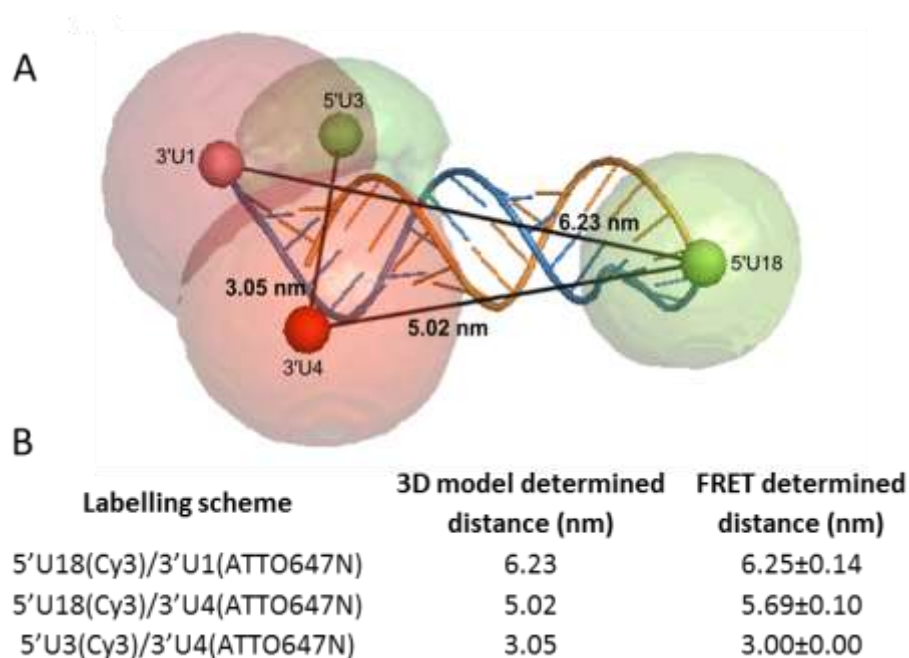


Fig. 10 – A) The RNA in the panhandle conformation, modelled together with the dyes attached to the 5' end (Cy3 at positions 3 or 18) and 3' end (ATTO647N at positions 1 and 4). B) Table showing a comparison between the determined distances between the dyes attached to the RNA in the panhandle conformation modelled in A and the distance determined from the corrected FRET values, for the same labelling scheme of the construct.

A secondary structure of the RNA in the corkscrew conformation was generated by forcing some of the residues to remain single-stranded (fig. 11A); without this constraint, the structure could not be formed and hence no 3D structure could have been derived. The first attempt to generate a 3D representation of the corkscrew failed, as the 5' and 3' extreme loops were represented collapsed onto each other (fig. 11B); this was because

no input on the tertiary structure was available. The second attempt was more successful – this structure was generated by representing each segment (3' and 5' loops and the distal promoter) individually and then assembling them in PyMol (fig. 11C).

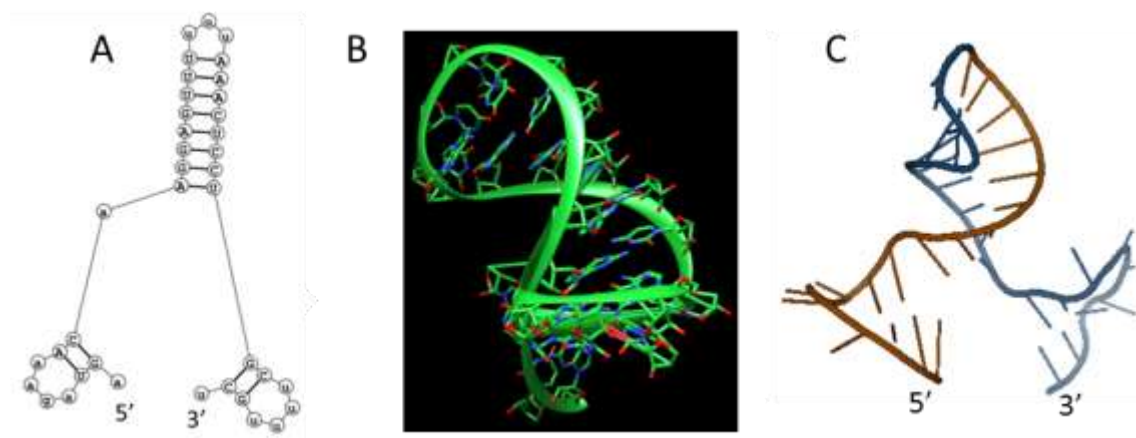


Fig. 11 - A) The 5' end and 3' end RNA strands used in this study in a hairpin structure were constrained to the secondary structure known to be characteristic to the viral RNA promoter in the corkscrew (using the BiFold module of the RNAstructure web-based software). B) The first Assemble2 3D generated model for the dsRNA construct shown in A, in which the 5' and 3' end loops of the corkscrew are collapsed onto each other. C) The second try to generate a 3D model of the corkscrew structure: after translating each individual end loop and the distal promoter in 3D using Assemble2 , the three building blocks were assembled in one structure using PyMol. In blue, the 3' end which corresponds to the first 20nt of the neuraminidase gene; in brown, the 5' end (18nt).

This structure was then used for to find out what the distances between the fluorophores in each FRET pair used is: 5'U18(Cy3)/3'U1(ATTO647N) are located 5.68nm apart, while 5'U18(Cy3)/3'U4(ATTO647N) and 5'U3(Cy3)/3'U4(ATTO647N) are located 4.42nm, respectively 4.40nm apart (fig. 12A), values which are also in good agreement with the distances determined through FRET (fig. 12B).

3.3.7 RNA polymerase binding to the dsRNA construct in the presence of ApG dinucleotide

Having studied the interaction between the polymerase and the dsRNA promoter from a qualitative, quantitative and structural point of view, it would be interesting to investigate the changes that occur in transcription. Hence, employing the same single-molecule assay as described previously, I have looked at the interaction in the presence of the initiating dinucleotide ApG, which mimics the system in transcription initiation.

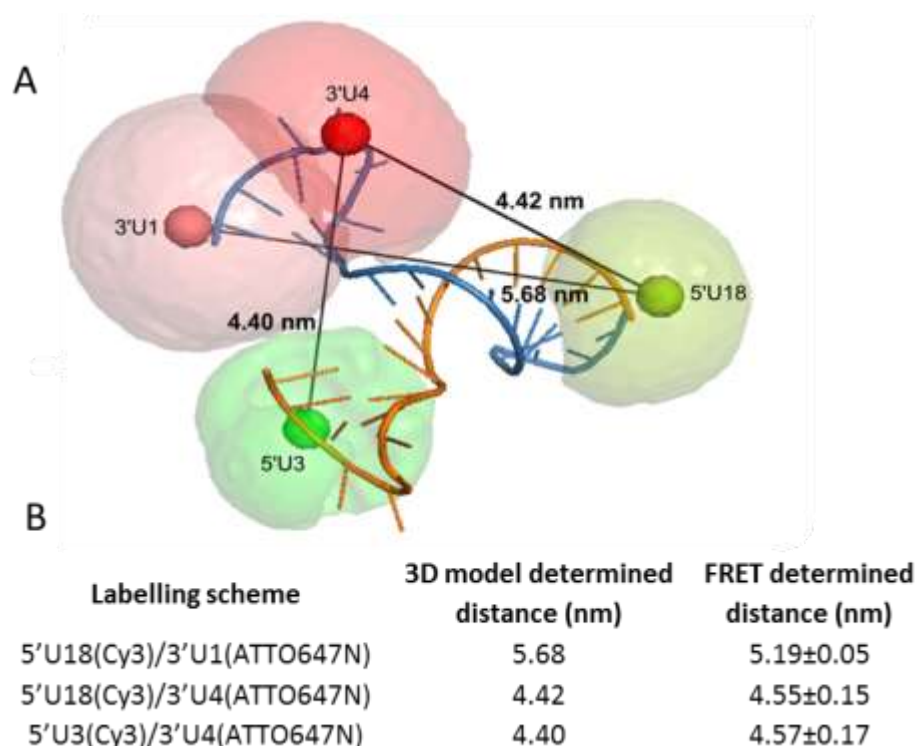


Fig. 12 - The RNA in the corkscrew conformation, modelled together with the dyes attached to the 5' end (Cy3 at positions 3 or 18) and 3' end (ATTO647N at positions 1 and 4). B) Table showing a comparison between the determined distances between the dyes attached to the RNA modelled in the corkscrew conformation and the distances determined from the corrected FRET values determined using smFRET in solution, for the same labelling scheme of the construct.

The results shown here are preliminary and have only been taken using one of the RNA constructs investigated previously (fig. 13A). When the RNA is investigated in the absence of protein, it is characterised by an apparent FRET value $E^*=0.57$ (as determined in section 3.3.3 and seen in fig. 13B); if ApG is added to the RNA, but no protein, the construct will be characterised by a mean FRET value $E^*=0.7$ (fig. 13C). Upon addition of 2nM RNA polymerase, two populations can be detected, as expected (fig. 13D); however, when 1mM ApG is added to the sample, a donor only population can clearly be identified in the E^*/S histogram, both in the FRET projection ($E^*<0.2$) and in the stoichiometry one ($S>0.8$; fig. 13E). The same trend is detected when the dsRNA is incubated with 20nM RNAP (fig. 13F) and the reaction is supplemented with 1mM ApG (fig. 13G).

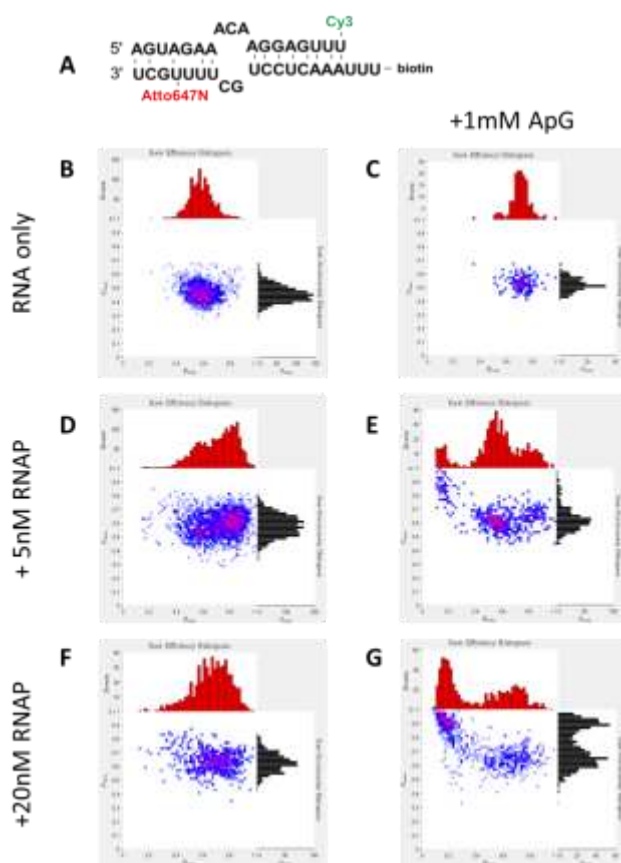


Fig. 13 –

5'U18(Cy3)/3'U4(ATTO647N)

construct (A) used in the experiments aiming to determine whether addition of ApG to the RNAP/RNA complex induces any additional structural changes. RNA in the absence of protein (B) is characterised by an apparent FRET value $E^*=0.57$. The samples were prepared by incubating RNA with the indicated RNA polymerase concentration at 28°C for 15 minutes and were imaged either in the absence (D, F) or presence of 1mM ApG (E, G). The data presented here is analysed using a dual channel burst search and a minDA=30 threshold.

3.4 DISCUSSION

Having determined that the polymerase binding to the promoter is not deterred by the positions of the dye on the RNA construct and that the interaction occurs, I use a single-molecule FRET under alternating laser excitation to investigate the interaction between these partners in a quantitative manner, and determined that the affinity of the polymerase for binding the double-stranded RNA promoter is higher ($K_D=0.6-0.7$ nM) than binding any of the ends individually (lowest binding constant determined was for the 5' end, 2.2nM). All this is not surprising given that the polymerase is attached to the RNA

panhandle in the RNP and the RNA structure remains double-stranded in transcription, when due to steric constraints, a poly(U) tract close to the 5' terminus is copied repeatedly, giving rise to the poly(A) tail of the viral mRNA molecule.

The binding constant reported here is ~30x smaller than the previously reported binding constant for the PB1 subunit binding to the ds vRNA ($K_D=20$ nM; González & Ortín 1999a); the difference can be attributed partly to the fact that in this study the heterotrimer is used, rather than the PB1 subunit, suggesting that the presence of all of the subunits might influence the binding affinity for RNA. Another way of explaining the difference is by way assessing how the protein used was prepared, with the pure fraction, containing polymerases not associated with a residual RNA, being selected. Finally, the techniques used here are more sensitive than those previously used, allowing for investigation of single-molecules and heterogeneity within the sample, rather than characterise it in bulk (as many solution-based biochemical techniques do).

A recent biophysical study of the vRNA promoter has shown that the 5' and 3' conserved ends of influenza genes can associate and give rise to a duplex panhandle structure even in the absence of the polymerase, while also confirming that the corkscrew structure does not occur in the absence of the protein (Noble et al. 2011). The single-molecule FRET results presented here confirm both these findings and unambiguously show that the conformational change in the promoter is dependent on its interaction with the polymerase. Moreover, using quFRET, we are able to show that the conformational change does indeed take place in the proximal promoter region, suggesting the promoter could be in the corkscrew conformation.

Investigating the changes in FRET efficiency correlated with the detected structural change of the same wild-type promoter sequence under three distinct labelling schemes, allowed determination of the relative distances between the reporter dyes used in our assays. The distances obtained characterise an RNA molecule in a three-dimensional space and considering them in a two-dimensional representation of the promoter could lead to misleading interpretation. Unfortunately, there is no 3D structure available for the influenza promoter in the corkscrew conformation; the unbound promoter was previously investigated using NMR (Bae et al. 2001), but the structure derived from this early study (PDB: 1J07) consists of only the conserved promoter sequence (shorter than the constructs used here) and is a hairpin (versus two interacting molecules, which would describe the constructs used here better). To overcome these limitations, I decided to model the possible RNA conformations *ab initio*.

While modelling the panhandle structure was straightforward, reconstituting the structure of the RNA in complex with the protein proved much more difficult; secondary structure prediction algorithms (Reuter & Mathews 2010) did not return the corkscrew conformation, suggesting that formation of this structure is dependent on protein interaction or, possibly, on intra-molecular tertiary structure. Posing constraints on the secondary structure (by forcing certain residues of the proximal promoter to be base paired), a secondary structure that represented a corkscrew could be generated and further used as a template for 3D generation. Unfortunately, when trying to generate the 3D model of the corkscrew, it became clear that having no tertiary structure information was going to be a significant limitation: when translating the 2D structure to 3D, the

corkscrew loops were represented collapsed onto each other, which is not what we would expect to occur in the presence of the protein. To overcome this, I decided to model the distal promoter region and the two corkscrew loops independently and assemble them into one structure, assuming that the only way the loops can move relative to the distal region is through rotation along the bond between residues A10-A11 (on the 5' end strand) and G9-U10 (on the 3' end strand), where the first residue is involved in the loop and the second is part of the distal promoter helix.

The experimentally determined distances agree with the ones from the derived models and support the panhandle structure for the free RNA and the corkscrew structure for the bound RNA. To my knowledge, this study is the first to be compatible with the corkscrew conformation, while also being the first to observe its formation directly (rather than extrapolating from loss of function experiments employing mutagenesis) and in real-time. One question that arises, though, is whether the double-stranded structure that exists in transcription initiation, elongation and termination assumes a panhandle or a corkscrew conformation, or possibly, a different structure altogether. To start investigating this question, the binding assay was modified so that the RNAP/RNA complex is in the presence of the correct initiating dinucleotide (ApG) for the promoter used; setting the experiment in this way mimics transcription initiation *in vitro*. While the results presented here are preliminary and a more thorough refinement of the assay is needed, for now there is some suggestion that the structure of the promoter changes even further when ApG is present, as indicated by the detection of a donor-labelled species in the experimental E/S histograms. It is tempting to speculate that the promoter falls apart and

the two individual ends are bound to different binding sites on the protein (it was previously suggested that there are two different binding sites for the 5' end in the PB1 subunit, while the 3' end is primarily bound by base pairing with the 5' end; Jung & Brownlee, 2006), however more experiments would be needed to establish what exactly happens.

3.5 CONTRIBUTIONS

The H3N2 recombinant RNA polymerase was expressed and purified as described in section 1.2.5 by Narin Hengrung (University of Oxford). For dye positioning I used the values determined by Dr. Timothy Craggs (University of Oxford), who modelled the dyes attached to DNA using ChemDraw.

3.6 BIBLIOGRAPHY

- Bae, S.H. et al., 2001. Structural features of an influenza virus promoter and their implications for viral RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), pp.10602–7.
- Baudin, F., Petit, I. & Ruigrok, R.W., 2001. Structure of the RNA inside influenza virus RNPs. *International Congress Series*, 1219, pp.451–456.
- Brownlee, G.G. & Sharps, J.L., 2002. The RNA Polymerase of Influenza A Virus Is Stabilized by Interaction with Its Viral RNA Promoter. *Journal of Virology*, 76(14), pp.7103–7113.
- Cordes, T. et al., 2010. Sensing DNA opening in transcription using quenchable Förster resonance energy transfer. *Biochemistry*, 49(43), pp.9171–80.
- Flick, R. et al., 1996. Promoter elements in the influenza vRNA terminal structure. *RNA (New York, N.Y.)*, 2(10), pp.1046–57.
- Flick, R. & Hobom, G., 1999. Interaction of influenza virus polymerase with viral RNA in the “corkscrew” conformation. *The Journal of general virology*, 80 (Pt 10, pp.2565–72.

- Fodor, E., Pritlove, D.C. & Brownlee, G.G., 1994. The influenza virus panhandle is involved in the initiation of transcription. *Journal of virology*, 68(6), pp.4092–6.
- Fodor, E., Seong, B.L. & Brownlee, G.G., 1993. Photochemical cross-linking of influenza A polymerase to its virion RNA promoter defines a polymerase binding site at residues 9 to 12 of the promoter. *The Journal of general virology*, 74 (Pt 7), pp.1327–33.
- González, S. & Ortín, J., 1999a. Characterization of influenza virus PB1 protein binding to viral RNA: two separate regions of the protein contribute to the interaction domain. *Journal of virology*, 73(1), pp.631–637.
- González, S. & Ortín, J., 1999b. Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *The EMBO journal*, 18(13), pp.3767–75.
- Hsu, M. et al., 1987. Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proceedings of the National Academy of Sciences*, 84(November), pp.8140–8144.
- Hwang, H., Kim, H. & Myong, S., 2011. Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), pp.7414–8.
- Jossinet, F., Construct RNA 3D Architectures with Assemble2.
- Jossinet, F., Ludwig, T.E. & Westhof, E., 2010. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics (Oxford, England)*, 26(16), pp.2057–9.
- Jung, T.E. & Brownlee, G.G., 2006. A new promoter-binding site in the PB1 subunit of the influenza A virus polymerase. *The Journal of general virology*, 87(Pt 3), pp.679–88.
- Kalinin, S. et al., 2012. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nature methods*, 9(12), pp.1218–25.
- Kapanidis, A.N. et al., 2005. Alternating-laser excitation of single molecules. *Accounts of chemical research*, 38(7), pp.523–33.
- Leahy, M.B. et al., 2001. Mutagenic analysis of the 5' arm of the influenza A virus virion RNA promoter defines the sequence requirements for endonuclease activity. *Journal of virology*, 75(1), pp.134–42.
- Leahy, M.B., Dobbyn, H.C. & Brownlee, G.G., 2001. Hairpin Loop Structure in the 3' Arm of the Influenza A Virus Virion RNA Promoter Is Required for Endonuclease Activity Hairpin Loop Structure in the 3' Arm of the Influenza A Virus Virion RNA Promoter Is Required for Endonuclease Activity. , 75(15).

- Lee, M.-K. et al., 2003. A single-nucleotide natural variation (U4 to C4) in an influenza A virus promoter exhibits a large structural change: implications for differential viral RNA synthesis by RNA-dependent RNA polymerase. *Nucleic Acids Research*, 31(4), pp.1216–1223.
- Lee, N.K. et al., 2005. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophysical journal*, 88(4), pp.2939–53.
- Li, M.L., Ramirez, B.C. & Krug, R.M., 1998. RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. *The EMBO journal*, 17(19), pp.5844–52.
- Noble, E. et al., 2011. Biophysical analysis of influenza A virus RNA promoter at physiological temperatures. *The Journal of biological chemistry*, 286(26), pp.22965–70.
- Pritlove, D.C. et al., 1999. A Hairpin Loop at the 5 ' End of Influenza A Virus Virion RNA Is Required for Synthesis of Poly (A) + mRNA In Vitro A Hairpin Loop at the 5 J End of Influenza A Virus Virion RNA Is Required for Synthesis of Poly (A) z mRNA In Vitro. , pp.5–11.
- Reuter, J.S. & Mathews, D.H., 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC bioinformatics*, 11, p.129.
- Tiley, L.S. et al., 1994. Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5 ' ends of the viral RNAs . Sequence-Specific Binding of the Influenza Virus RNA Polymerase to Sequences Located at the 5 ' Ends of the Viral RNAs. , 68(8).

Part B: Fluorescence-based assays for RNA detection

Chapter 4: A biosensor for influenza RNA detection

4.1 INTRODUCTION

Several global pathogenic threats have emerged in the last decade which required a rapid, structured response that would focus both on the identification of the causative agent and on containment. Overall, close to 1500 species of infectious agents that pose disease risks in humans have been determined – with 50% of these being viruses (including prions) and bacteria, and 61% originating in animals (Taylor, Latham, & Woolhouse, 2001). Of these identified pathogens, 175 are associated with diseases that have not occurred previously in humans or that have had either a very low incidence or were found in restricted geographical areas. Recent emerging diseases such swine flu, avian flu and Severe Acute Respiratory Syndrome (SARS) have strengthened the idea that it is impossible to predict outbreaks, but paramount to identify them in a timely fashion. Detecting and correctly diagnosing the conditions such pathogens trigger upon infection cannot simply rely on clinical symptoms, which often occur after some time post-infection, are generic and can at best indicate whether the causative agent is a virus or a bacteria, or a different kind of agent altogether.

Classically, identifying infectious agents was done by employing cell cultures or through serology, both of which are time consuming, labour intensive and can easily become cross-contaminated. Nowadays, it is possible to identify molecular markers that are characteristic for prevalent pathogens by employing polymerase chain reaction (PCR)

based techniques or by using an enzyme linked immunosorbent assay (ELISA) (Lazcka et al. 2007; Nayak et al. 2009; Skottrup et al. 2008; Ivnitski et al. 1999; Nayak et al. 2013). Both of these techniques rely on amplifying the signature signal of the pathogen, with detection occurring in an indirect fashion. While such techniques are faster than the classical detection techniques, they still take up to 10 hours before returning any results, with much of this time committed to sample preparation.

When considering influenza detection specifically, most procedures are aimed at detecting both influenza A and B viruses, the only two viruses in the Orthomyxoviridae known to infect humans. Out of these procedures, serology is by far the slowest, with results becoming available after 2 weeks or more, followed by viral culture, which takes 3-10 days (Gavin & Thomson 2004). A more rapid detection, taking between 2-5 hours is possible using immunofluorescence based techniques, RT-PCR ones or various enzyme immuno-assays; ultra-rapid tests are also available, being able to detect influenza presence in a sample in 15 minutes or less (Storch 2003; Hurt et al. 2007). Rapid tests show a sensitivity for identifying influenza of roughly 50%, with false negative results being a particular problem (Blyth 2009; Faix 2009); moreover, not all these rapid detection tests can distinguish between influenza A and influenza B (Centre for Disease Control and Prevention, www.cdc.gov/flu).

In view of all this, direct, fast detection methods with increased reliability are desired. Owing to recent technical advancements, single-molecule fluorescence spectroscopy techniques are an alternative to the detection methods already mentioned and more assays based on them are starting to become available (Lympieropoulos et al., 2010;

Crawford, Kelly, & Kapanidis, 2012; Yim et al., 2012). This is not surprising given that assays based on single-molecule fluorescence have the potential to be more sensitive, specific and rapid than other detection techniques.

Proposed detection assay

To detect the presence of influenza RNA in a sample, a single-molecule assay based on the use of two different probes was devised, as detailed below (and in fig.1):

a) A capture probe, used to pull down the influenza RNA to be detected. As thoroughly described in Chapter 2 of this thesis, the negative-sense, segmented viral genome carries conserved sequences to its 3' and 5' ends; moreover, these conserved sequences show partial complementarity, allowing the ends to come together in a dsRNA structure. As this feature of the viral RNA is conserved across all influenza A viruses (and to a certain extent, also in influenza B viruses; Desselberger, Racaniello, Zazra, & Palese, 1980), it is advantageous to exploit it and hence design a capture probe that is universal and is based on the conserved sequence of the 3' end.

b.1) A detection probe: initially, for the proof of principle experiments, only one gene would be detected – the neuraminidase gene of the influenza A virus (NA/A). As the pathogenic RNA to be detected is cell-extracted and given the life cycle of the virus and the relative abundance of the different viral RNA species in such a sample, sensing the positive-sense viral nucleic acids (mRNA and the cRNA) seemed like the best approach; the detection probe is raised against a 30nt region of the NA/A gene found (roughly) in the centre of the gene. The probe is labelled with Alexa647 and the first four residues on both its ends are phosphorothioate variant residues, carrying a

sulphur atom instead of an oxygen one on the phosphate backbone; such a modification ensures that the probe is less susceptible to nuclease degradation (in later experiments, this probe was replaced with one that is identical in sequence and labelling position, but which carries no phosphorothioate linkages).

b.2) Alternatively, instead of using the detection probe described above, two shorter probes targeting the same region of the NA/A gene could be used: the two probes are labelled with Cy3b and ATTO647N (separated by 29nt) and are expected to have a FRET signature of 0.1 (Periz et al. 2013). The advantage of using two probes with a characteristic FRET value is apparent when considering multiplexing, which would be the next step (with detecting one more gene of influenza A – the haemagglutinin gene, and at least one gene from the influenza B virus).

The approach proposed here is based on a step-wise addition of the different reagents (fig. 1, inset):

- 1st step: addition and immobilisation of the biotinylated capture probe to the glass surface;
- 2nd step: addition of extracted RNA from 293T cells that have been infected with influenza A/WSN/1933 virus, which would allow the pathogenic RNA of interest to be retained due to it becoming hybridised to the immobilised capture probe;
- 3rd step: addition of the detection probe;
- 4th step: data collection.

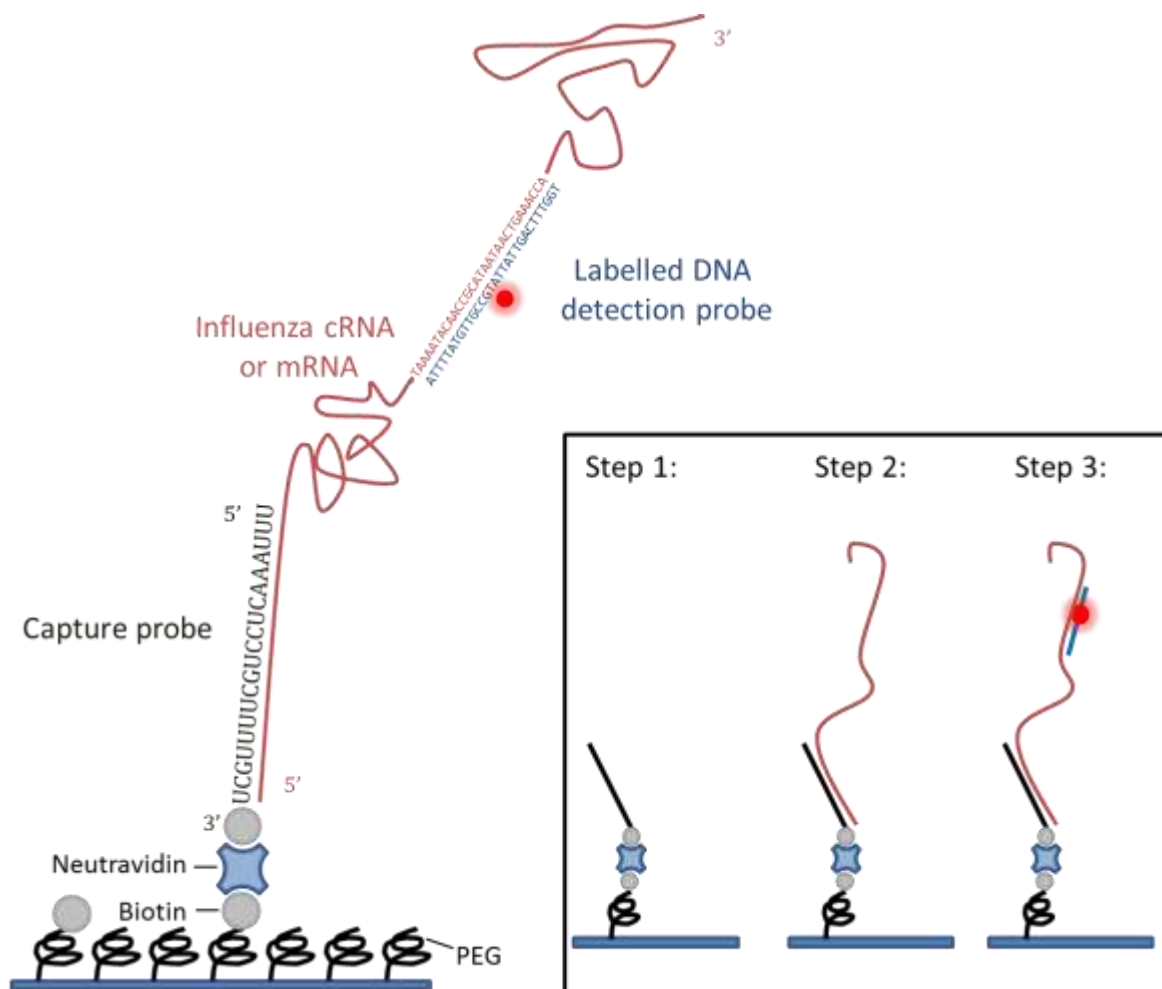


Fig. 1 – Schematic of the biosensor, showing the immobilisation procedure, the different probes used and the expected interaction between them and the RNA to be detected. The inset depicts the step-wise manner of the assay (for clarity, the different nucleic acids used in the assay have been represented using the same colours).

The biosensing assay development was informed by knowledge of the viral life cycle (overview in fig. 1A of Chapter 1 of this thesis; Palese & Shaw in Fields, 2007), on the one hand, and the relative stability of nucleic acid duplexes, on the other. An understanding of the polarity of the viral RNAs (where vRNA is of negative polarity and the mRNA and cRNA species are of positive polarity) and their ends' morphology, as well as the temporal regulation of their expression further informed the design of the biosensing assay proposed here.

Finally, the assay also took into account the properties of hybrids consisting of different nucleic acids: several studies have found that RNA:RNA hybrids are the most stable type of hybrid duplexes, followed by either DNA:DNA or DNA:RNA hybrids, depending on the sequence of the strands involved and the formamide concentration of the buffer used for hybridisation (Clark et al. 1997; Sugimoto et al. 1995; Casey & Davidson 1977). Moreover, a study investigating whether phosphorothioate substitutions in the DNA backbone of probes used for hybridisation have any influence on the stability of formed hybrids found that the presence of this linkage lowers the melting temperature of the hybrid by $\sim 8^{\circ}\text{C}$, rendering it less stable (Clark et al. 1997), which reflects in the choice of switching between a modified probe to an unmodified one (both Alexa647-labelled).

4.2 MATERIALS AND METHODS

4.2.1 Probes

Single-stranded RNA and DNA probes were synthesised and labelled as described previously in Chapter 2; the different probes referred to in this chapter are reviewed in table 1.

Probe	Nucleic acid	Sequence (3'->5')
Capture probe	DNA	3' TCGTTTTCGTCCTT-biotin 5'
NA/A detection probe (phosphorothioate linkages*; Alexa647-labelled)	DNA	3' A*T*T*T*TATGTTGCCG7ATTATTGACTT*T*G*G*T 5'
NA/A detection probe (Alexa647-labelled)	DNA	3' ATTTTATGTTGCCG7ATTATTGACTTTGGT 5'
NA/A FRET detection probes (ATTO647N- & Cy3b-labelled)	DNA	3' GACATAATTTTA7GTTGCCGT 5' 3' ATTATTGACTTTGGTATTTT 5'
3' end vRNA (capture probe)	RNA	3' UCGUUUUCGUCCUCAAUUU-biotin 5'
5' end vRNA (partially complementary to the capture probe)	RNA	3' UUUGAGGAACAAAGAUGA 5'

Table 1 – Overview of the probes used in the assays described in this chapter. The residue shown in italics is the one to which the label is attached; three different labels have been used (ATTO647N, Alexa647 – red acceptor probes and Cy3b – green donor probe).

The cRNA sequence of the NA gene of the A/WSN/1933 virus (PubMed accession: J02177) is given in table 2; this is the influenza gene that the assay aims to detect.

1	AGCGAAAGCAGGAGT	TTAAATGAATCCAAACCAGAAAATAATAACCATTGGGTCAATCTG
61	TATGGTAGTCGGAATAATTAGCCTAATATTGCAAATAGGAAATATAATCTCAATATGGAT	
121	TAGCCATTCAATTCAAACCGGAAATCAAACCATACTGGAATATGCAACCAAGGCAGCAT	
181	TACCTATAAAGTTGTTGCTGGGCAGGACTCAACTTCAGTGATATTAACCGGCAATTCATC	
241	TCTTTGTCCCATCCGTGGGTGGGCTATACACAGCAAAGACAATGGCATAAGAATTGGTTC	
301	CAAAGGAGACGTTTTTTGTCATAAGAGAGCCTTTTATTTTCATGTTCTCACTTGGAAATGCAG	
361	GACCTTTTTTCTGACTCAAGGCGCCTTACTGAATGACAAGCATTCAAGGGGGACCTTTAA	
421	GGACAGAAGCCCTTATAGGGCCTTAATGAGCTGCCCTGTCGGTGAAGCTCCGTCCCCGTA	
481	CAATTC AAGTTTGAATCGGTTGCTTGGTCAGCAAGTGCATGTCATGATGGAGTGGGCTG	
541	GCTAACAATCGGAATTTCTGGTCCAGATGATGGAGCAGTGG	CTGTAT TAAAAATACAACCG
601	CATAATAACTGAAACCA	TAAAAAGTTGGAGGAAGAATATATTGAGAACACAAGAGTCTGA
661	ATGTACCTGTGTAAATGGTTCATGTTTTACCATAATGACCGATGGCCCAAGTGATGGGCT	
721	GGCTCGTACAAAATTTTCAAGATCGAGAAGGGGAAGGTTACTAAATCAATAGAGTTGAA	
781	TGCACCTAATTCTCACTACGAGGAATGTTCTGTACCCTGATACCGGCAAAGTGATGTG	
841	TGTGTGCAGAGACAATTGGCACGGTTCGAACCGACCATGGGTGTCCTTCGACCAAAACCT	
901	AGATTATAAAATAGGATACATCTGCAGTGGGGTTTTTCGGTGACAACCCGCGTCCCAAAGA	
961	TGGAACAGGCAGCTGTGGCCCAGTGTCTGCTGATGGAGCAAACGGAGTAAAGGGATTTTC	
1021	ATATAAGTATGGCAATGGTGTGTTGGATAGGAAGGACTAAAAGTGACAGTTCCAGACATGG	
1081	GTTTGAGATGATTTGGGATCCTAATGGATGGACAGAGACTGATAGTAGGTTCTCTATGAG	
1141	ACAAGATGTTGTGGCAATAACTAATCGGTCAGGGTACAGCGGAAGTTTCGTTCAACATCC	
1201	TGAGCTAACAGGGCTAGACTGTATGAGGCCTTGCTTCTGGGTGAATTAATCAGGGGGCT	
1261	ACCTGAGGAGGACGCAATCTGGACTAGTGGGAGCATCATTTCTTTTTGTGGTGTGAATAG	
1321	TGATACTGTAGATTGGTCTTGCCAGACGGTGCTGAGTTGCCGTTCCACATTGACAAGTA	
1381	GTTTGTTCAAAAAACTCCTTGTCTTCTACT	

Table 2 – cRNA sequence of the NA gene of A/WSN/1933 influenza virus as retrieved from GeneBank (accession: J02177; Hiti & Nayak, 1982); the regions of the NA gene that are targeted by the probes are highlighted or represented in the same colour as the probes shown in table 1.

4.2.2 Eukaryotic cell culture

Human embryonic 293T or Madin-Darby bovine kidney epithelial (MDBK) cells were grown in 75cm² flasks (Costar) in DMEM medium (for 293T cells; PAA Laboratories) or

MEM medium (for MDBK cells; Fluka) supplemented with 10% heat-inactivated FCS (Sigma Aldrich) and 2mM L-glutamine (Sigma Aldrich), at 37°C with 5% CO₂; when confluent, the medium was discarded and the monolayer was washed with PBS, followed by incubation with 2ml trypsin (PAA Laboratories), at 37°C, until all the cells detached from the flask's surface and separated. Fresh MEM was added to the flask and the cell suspension was further separated by repeatedly pipetting through a 10ml pipette, before being used to seed fresh culture flasks at a pre-determined density, in MEM.

4.2.3 Virus purification

MDBK cells were seeded for infection in four 175cm² flasks, in MEM medium and grown to 80% confluency, after which they were infected with A/WSN/1933 at a multiplicity of infection (MOI) of 0.01, by adding 10µl virus (titre = 7.2×10^8 pfu/ml) in 10ml MEM, 0.5% foetal calf serum (FCS) and 1% L-glutamine. The cells were incubated for 1 hour at room temperature, with side rocking every 15 minutes; 1 hour post-infection, the culture medium was replaced with 30ml MEM supplemented with 0.5% heat-inactivated FCS and 1% L-glutamine, after which the cells were incubated for 48 hours at 37°C. The culture medium was collected in 50ml tubes and clarified by centrifugation for 30 minutes at 3000rpm and 4°C; the supernatant was transferred to pre-chilled (at 4°C) SW28 tubes and subjected to ultracentrifugation for 30 minutes at 10000rpm and 4°C (in a Beckman SW28 rotor) and then transferred once more onto pre-cooled 30% sucrose cushions, also in SW28 tubes. The sample was pelleted by centrifugation for 1.5 hours at 25000rpm and 4°C; the supernatant was discarded, each pellet was resuspended in 50µl NTC buffer (20mM Tris-HCl pH 7.4, 100mM NaCl, 5mM CaCl₂) and all the fractions were pulled

together and stored on ice overnight. The sample was applied on a discontinuous sucrose gradient (2.5ml each of 30%, 40%, 50%, 60% (w/v) sucrose) in NTC buffer and further fractionated by centrifugation for 2.5 hours at 35000rpm and 4°C in a Beckman SW41Ti rotor. The virus-containing fraction sediments at the 40% sucrose mark, appearing as a white, milky band; the fraction of interest is extracted and collected in a syringe (1-2ml sample), transferred to a SW41 tube with 8ml NTC buffer and centrifuged for 1 hour at 30000rpm and 4°C to pellet the virions. Finally, the pellet is resuspended in 1ml NTC, aliquoted and stored at -80°C.

4.2.4 Plaque assay

A ten-fold serial dilution of A/WSN/1933 virus purified as above, prepared in MEM supplemented with 0.5% FCS and 2mM L-glutamine, was used to infect MDBK cells grown in monolayer in 35mm dishes, for 1 hour at room temperature. Then, the virus solution was replaced with 2ml agarose solution containing 2% agarose in PBS and an equal volume of MEM supplemented with 0.5% FCS, which was allowed to set. The dishes were incubated for 48 hours at 37°C with 5% CO₂, after which the agarose pad was removed and the plaques were visualised by staining with 2ml 0.2% Coomassie Brilliant Blue dye in 7.5% glacial acetic acid and 50% ethanol (fig. 2).

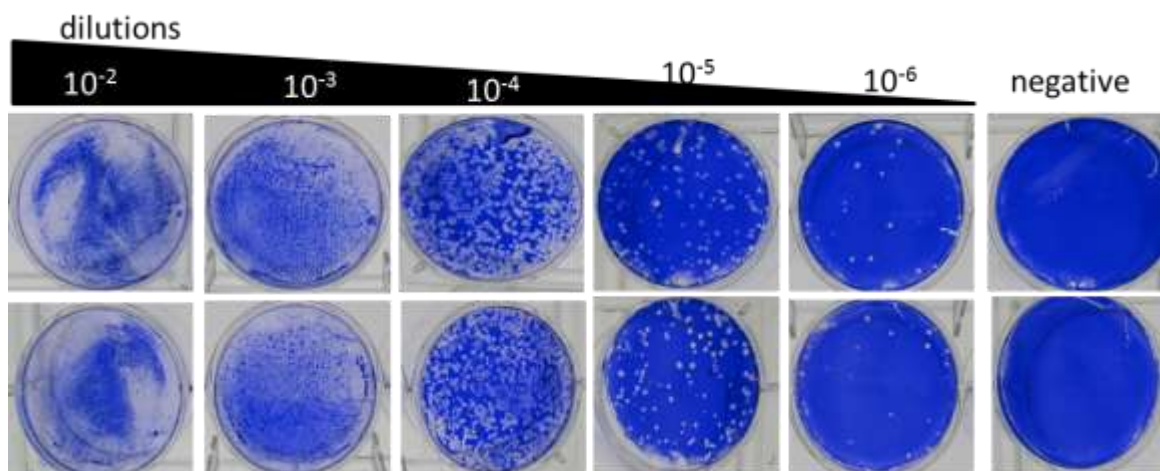


Fig. 2 – Plaque assay performed in duplicate, photographed after staining. The well with the a medium density of plaques was used for quantitation of the virus titre – specifically, the number of plaques was counted and multiplied by the dilution factor, after which it was divided by the volume of virus added to yield the virus titre in ‘plaque forming units/ml’.

4.2.5 Viral RNA preparation

293T or MDBK confluent cells in a monolayer (in 10cm dishes) were infected with purified influenza A/WSN/1933 virus of known titre (3.3×10^8 pfu/ml). To determine how much virus needs to be added (in number of μ l to be used), given the virus titre and the desired multiplicity of infection, the following relation was used (Fields 2007):

$$[(N \times \text{MOI}) / \text{virus titre}] \times 10^3$$

N = number of cells/dish (8.8×10^6 cells/10cm dish)

MOI = multiplicity of infection, a measure of the % of cells infected with at least one viral particle (MOI = 1 -> 63.2% cells infected; MOI = 3 -> 95% cells infected)

Virus titre = the concentration of the stock virus, expressed in ‘plaque forming units’ (pfu) and assessed via the plaque assay

Having determined how much virus needs to be used for the desired multiplicity of infection (usually MOI = 3), the virus stock diluted in DMEM supplemented with 0.5% FCS (or MEM supplemented with 0.5% FCS, in the case of MDBK cells) and added to the dish, followed by incubation for 1 hour at room temperature, with rocking every 15 minutes. The medium was then replaced with fresh DMEM + 0.5% FCS (or MEM + 0.5% FCS) and the cells were allowed to recover and grow for 4 hours at 37°C with 5% CO₂. At the end of the incubation, if the cells had dislodged from the bottom of the dish, they were harvested by scraping in the medium, which was decanted in a 15ml tube and clarified by centrifugation for 1 minute at 1000rpm and 4°C; the supernatant was discarded and the pellet was resuspended in 1ml TRIzol (Life Technologies); if the cells were still attached to the dish, the medium was aspirated and the cells were harvested by scraping in 1ml TRIzol; in both cases, the cell lysates were then stored at -20°C overnight or for longer. Total RNA was extracted from the samples by incubation with 200µl chloroform for 5 minutes at room temperature, followed by centrifugation for 20 minutes in a benchtop centrifuge at 13000rpm at 4°C, after which the top layer of liquid was collected in a fresh tube to which 500µl isopropanol was added; the sample was treated as before, but this time the supernatant was discarded and the pellet was washed with 1ml 100% ethanol for 10 minutes, 13000rpm at room temperature. The ethanol was discarded and the pellets were air dried, after which they were resuspended in 50-100µl RNase-free H₂O (Ambion) or TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA) and stored at -20°C. Total RNA was also extracted from uninfected cells and used in the assay as a negative control.

4.2.6 Primer extension assay

Total RNA was extracted as described in the previous section; 1/25th of each total RNA sample was added together with an excess of two DNA primers, labelled with ³²P and targeted to the NA gene, to a final volume of 5µl and denatured by heating for 5 minutes at 95°C, followed by incubation on ice for 5 minutes. The sample was re-heated to 50°C for 2 minutes, after which it was mixed with 5µl 2x transcription mix (2x First Strand Buffer (Invitrogen), 20mM DTT, 4mM dNTP mix and 250U SuperScript II Reverse Transcriptase (200U/µl; Invitrogen)), pre-heated at 50°C. The reaction was incubated for 1 hour at 50°C, after which it was stopped by enzyme inactivation through heating for 10 minutes at 70°C, prior to addition of 8µl of loading buffer (90% formamide, 10mM EDTA pH 8.0, 0.04% bromophenol blue and 0.04% xylene cyanol FF) and heat denaturation at 95°C for 3 minutes. The samples were electrophoresed on a 7M urea/6% polyacrylamide (19:1 acylamide: bis-acrylamide) gel in 1x TBE (89mM Tris-borate, 2mM EDTA pH 8.3) for 3 hours at 1200V. The products were visualised by autoradiography and quantitated using Fujifilm imaging plates and a Fujifilm FLA-5000 fluorescent image analyser.

The primers used in this assay had the following sequence:

(Forward NA primer; detects vRNA): 5' TGGACTAGTGGGAGCATCAT 3'

(Reverse NA primer; detects mRNA and cRNA): 5' TCCAGTATGGTTTTGATTCCG 3'

4.2.7 Biosensing assay

50pM of Cy3b-labelled, biotinylated capture probe (DNA capture probe or AT2 vRNA probe; labelled with Cy3b) in PBS was immobilised on a glass slide functionalised using PEG:biotin-PEG and neutravidin (described in chapter 2, section 2.2.6) or on a glass slide

that was treated with 20 μ l of 1mg/ml biotin-BSA (Sigma Aldrich) for 5 minutes at room temperature, followed by 20 μ l 0.5mg/ml neutravidin for a further of 5 minutes. Total purified RNA was deposited onto the slide and incubated for 5 minutes, after which the sample was replaced with 100pM of Alexa647-labelled detection probe in PBS or annealing buffer (20mM Tris-HCl pH 8.0, 500mM NaCl, 1mM EDTA).

Data was collected on a custom-built total internal reflection microscope (TIRF) under alternating laser excitation, using a 543nm laser at 500 μ W intensity and a 640nm laser at 800 μ W intensity and 10ms time resolution; 300 frames/movie were acquired for each field of view and 5-10 movies for each condition investigated.

4.2.8 Binding troubleshooting: PAGE electrophoresis

Target DNA or target vRNA (in the case of the detection probe troubleshooting assay) were incubated with the relevant probes at a ratio of 5:1 (50nM:10nM) in PBS or annealing buffer and incubated either at room temperature or at 37°C for 30 minutes, after which they were electrophoresed on a 20% or a 15% non-denaturing PAGE (mini), in 1xTBE for 70 minutes at 180V at 4°C and visualised as previously described. Where staining with SybrSafe was necessary, the manufacturer's protocol was followed.

4.2.9 Binding troubleshooting: digestion assay

3 μ l of total RNA extracted from 293T cells, infected with A/WSN/1933 at an MOI = 3 or non-infected, was incubated with 1 μ l 200nM Alexa647-labelled NA/A DNA detection probe in hybridisation buffer (20mM Tris-HCl pH 7.4, 500mM NaCl, 1mM EDTA, 40% formamide) to a final volume of 30 μ l and hybridised at 42°C for 2.5 hours after being

denatured at 95°C for 4 minutes. Once the hybridisation is completed, the sample was incubated with 2500U Ribonuclease T1 (1700U/μl; Life Technologies) in digestion buffer (10mM Tris-HCl pH 7.4, 300mM NaCl, 5mM EDTA), for 30 minutes at 37°C; the digestion reaction was stopped by addition of 5μl proteinase K (20mg/ml; NEB) and further incubation at 37°C for 30 minutes. The sample was ethanol precipitated and the resulting pellets were resuspended in loading buffer (50% formamide, 5mM EDTA) and electrophoresed on a 2% agarose gel in the presence of SybrSafe nucleic acid stain. The bands of interest were visualised as before.

4.3 RESULTS

4.3.1 Initial attempts

The initial attempt to sense the presence of influenza RNA in a sample was based on a one-colour assay, in which the reporter label was on the detection probe, while the capture probe remained unlabelled (fig. 3A); once all the experimental components were added, images of the coverslip were recorded and a red signal was detected, suggesting that the labelled detection probe is bound (fig. 3B). This kind of signal could have also arisen from the Alexa647 probe binding non-specifically to the glass surface. To test whether this is the case, the assay was modified (fig. 3C) so that both the capture and the detection probe are labelled with a red-emitting fluorophore. Addition of the capture probe leads to an easy identification of the slide's surface and is also a way of determining how efficient the coverage is (fig. 3D); bleaching all the fluorophores corresponding to the capture probe, while not moving to a new field of view, followed by

addition of the labelled detection probe, ensures that any signal detected is from the newly added probe which binds either to the capture probe (based on the sequence of the two nucleic acids, this is not expected to happen) or non-specifically to the surface. As no signal could be detected, the conclusion of the experiment is that the detection probe does not bind to the surface-immobilised probe or the surface, and more importantly, that the signal detected (fig. 3B) is from the detection probe sensing the pathogenic RNA.

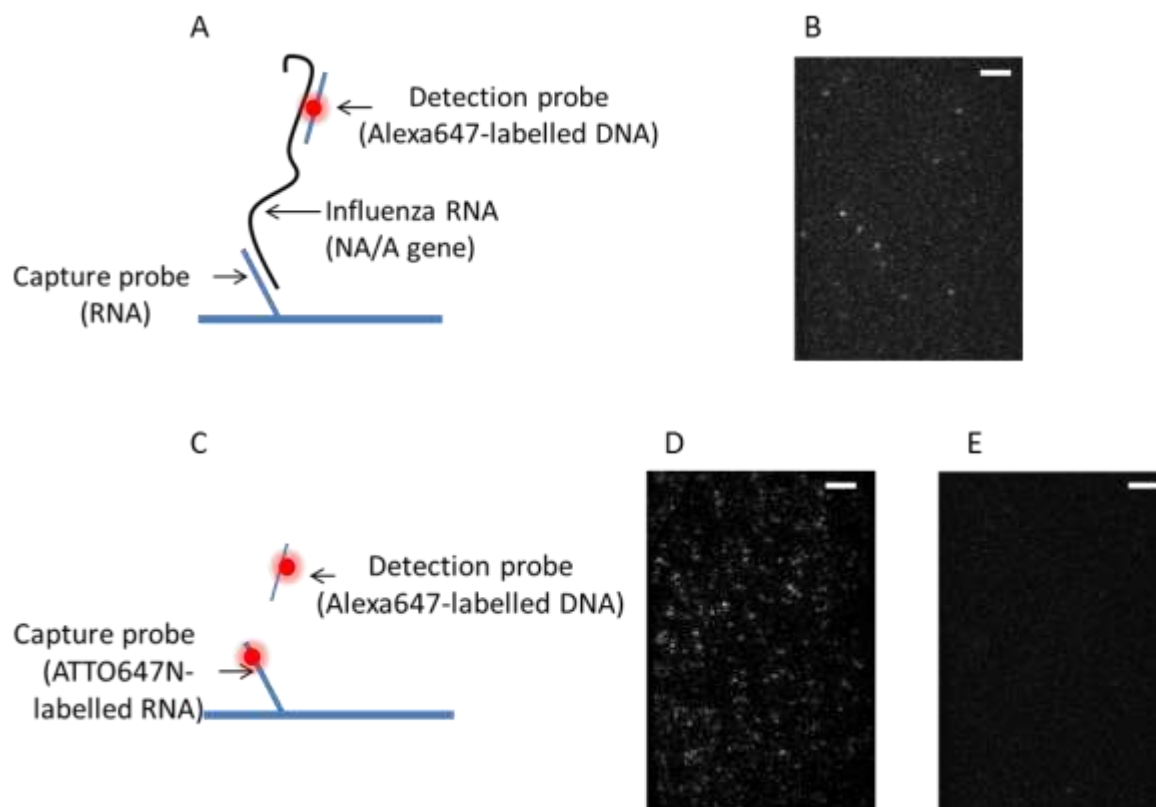


Fig. 3 – A) Scheme detailing the initial setup of the biosensing assay: an unlabelled RNA probe (3' end, described in section 4.2.1) is immobilised on the glass surface and used as a capture probe (50pM); 20 μ l total RNA extracted from influenza infected 293T cells is added to the slide and incubated for 3 minutes, after which it is replaced with 100pM Alexa647-labelled DNA detection probe, which is incubated for 5 minutes before imaging. B) Field of view after 5 minutes post-addition of the Alexa647-labelled detection probe (640nm laser illumination, 800 μ W continuous wave). C) Control assay: setup is as before, with the difference that the capture probe is ATTO647N-labelled and there is no extracted RNA added. D) Field of view captured after the addition of the ATTO647N-labelled capture probe (illumination as previously described). E) Field of view as imaged after 5 minute incubation with the Alexa647-labelled detection probe (this is the same field of view as in D, but prior to addition of the detection probe, all the fluorophores tagging the capture probe were bleached). (For images B-D, the width of the scale bar is

equivalent to 3µm.)

Encouraged by this result, the next step was to attempt the assay using a Cy3b-labelled capture probe (an RNA probe identical to the one used previously) and the same Alexa647-labelled detection probe (fig. 4A). Immobilisation of the Cy3b-capture probe results in good coverage of the surface with the probe (identified in the green 'G' channel, fig. 4B); after addition of total RNA extracted from influenza infected 293T cells, the detection probe is also added – 5 minutes after incubation with it, a faint signal can be detected (in the red 'R' channel, fig. 4B). When data is collected after 30 minutes of incubation with the detection probe, a much higher signal is identified, suggesting that binding of the probe to the RNA of interest occurs, but requires a longer time (fig. 4C). Finally, as a control, total RNA extracted from uninfected cells was used in the second step of the assay (fig. 4D): after addition of the Alexa647 probe and incubation for 30 minutes, no signal could be detected in the R channel, suggesting that the detection probe has nothing to bind to (as expected) and also doesn't bind to the surface (confirms that what was observed in fig. 4B&C is genuine detection).

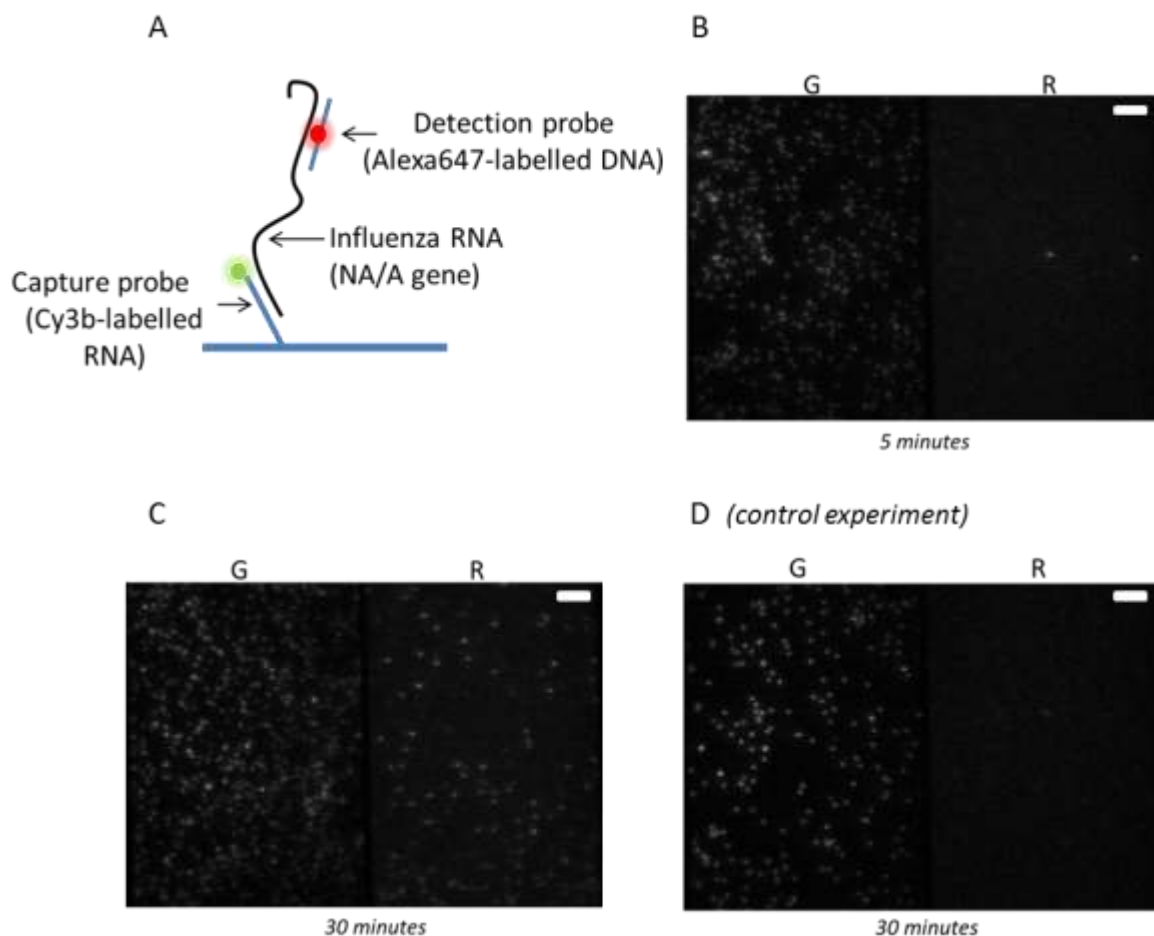


Fig. 4 – A) Schematic depiction of the modified sensing assay using two labelled probe: a detection RNA probe labelled with Cy3b and a DNA detection probe labelled with Alexa647. B) Field of view showing two channels (G=green channel, in which Cy3b emission is detected and R=red channel, in which Alexa647 emission is seen) taken after addition of 100pM Alexa647-labelled detection probe and incubation for 5 minutes (assay performed as described in section 4.2.7). C) Field of view taken after addition of the detection probe and incubation for 30 minutes. D) Control experiment, performed under the same conditions as the experiment detailed in A-C, but in which the sample RNA used was extracted from 293T cells that were not infected with the influenza virus. (Scale bar: 3 μ m)

The results of the initial attempts shown here suggest that the approach employed for influenza RNA detection is valid, however, the assay proved not to be reproducible. The capture probes used in the two experiments described above (and those that followed) are both RNA and have the exact same sequence; the difference between them lies in the fact that in one experiment, the capture probe is unlabelled (fig. 3A&B), while in the other, it carries a Cy3b fluorophore (fig. 4), but as influenza RNA is detected in both, it is safe to conclude that this difference doesn't have an impact on the assay. Initially, the detection probe used for these experiments was a DNA strand carrying phosphorothioate linkages (described in section 4.2.1), which makes it more resistant to nuclease activity; detection was also tested using a DNA probe that has the same sequence, carries the same fluorophore (and at the same position), but had no backbone modification. When this latter probe was employed, the assay proved to be unsuccessful; however, returning to the use of the phosphorothioate-modified DNA detection probe, also didn't yield the expected results, with no influenza RNA being identified. In view of this, it is unlikely that the nature of linkages in the detection probe's backbone plays a significant role in the irreproducibility of the assay. Finally, the extracted RNA used in the experiments detailed in fig. 3&4 was from the same sample; using this same total RNA preparation in a later experiment didn't lead to influenza RNA being detected. At the time, this was believed to be due to RNA degradation (the sample being a month old). The experiment was repeated with a fresh sample of extracted total RNA from both infected and non-infected cells, but also didn't yield the desired result. Several different RNA extraction strategies

were then employed, but none of the resulting samples expected to contain influenza RNA led to a positive detection result when used in the biosensing assay.

4.3.2 Troubleshooting

Since the assay's results obtained initially could not be replicated, recovering and improving reproducibility were seen as a priority, followed by optimisation of the assay. Several different problems that might have arisen when repeating the assay were identified, amongst which:

- a) there is not enough viral RNA to be detected in the sample;
- b) the purified viral RNA is rich in secondary structure which would prevent binding of the DNA detection probe;
- c) the binding conditions of the RNA to the capture probe or of the DNA detection probe to the immobilised RNA are not optimal and there is no detectable RNA:DNA hybrid formed.

4.3.2.1 RNA availability in the sample

One cause for which the biosensing assay might not be reproducible could be linked to the availability of the pathogenic RNA in the total RNA extracted. The vast majority of the extracted RNA will be of eukaryotic origin, and particularly ribosomal RNA – rRNA – known to make up about 80% of the total RNA in eukaryotes. While a precise, quantitative determination of the exact fraction made up by the pathogenic RNA within the total RNA is beyond the scope of these experiments, determining the relative levels of the three viral RNA species (vRNA, cRNA and mRNA) found in total extracted RNA might

offer insights, especially since the influenza RNA in the sample will be directly correlated with the multiplicity of infection, on the one hand, and with the metabolic state of the infected cells, on the other.

To determine the optimal levels at which the cells should be infected, a primer extension assay (fig. 5A), in which the total extracted RNA from 293T cells infected with influenza A/WSN/1933 virus at varying MOIs was used as a template (where MOI stands for 'multiplicity of infection', a measure of the % of cells infected with at least one viral particle). The results of this experiment suggest that cells infected at a MOI lower than 1 (and including 1) have little viral RNA, while infecting cells at an MOI = 3 seems ideal (results are comparable with those obtained from RNA extracted from cells infected at an MOI = 12; fig. 5B).

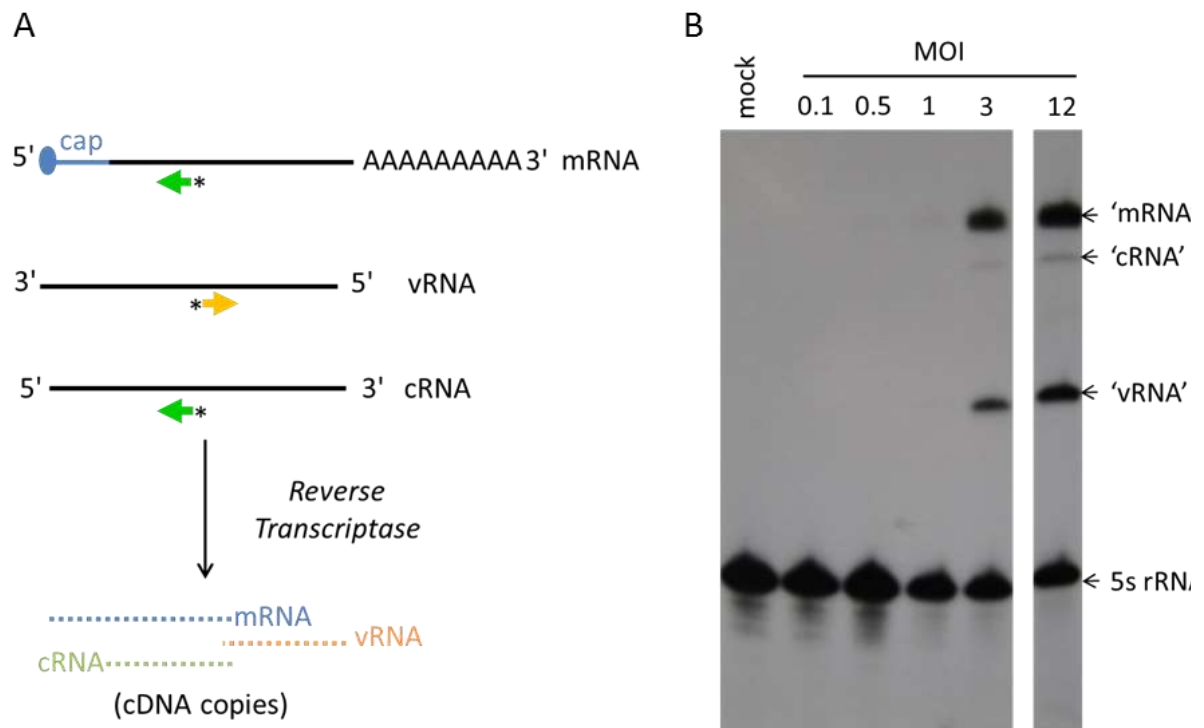


Fig. 5 – A) Primer extension assay: two ^{32} P-labelled DNA primers (depicted here as a green arrow, recognising the NA/A mRNA and cRNA, and a yellow arrow, for the NA/A vRNA) are incubated with total extracted RNA in the presence of a reverse transcriptase, which will yield cDNA copies of the different viral RNAs. B) Primer extension assay performed with total extracted RNA from 293T cells that were infected at different multiplicity of infection (0.1, 0.5, 1, 3, 12); for comparison, total extracted RNA from uninfected cells was also used as a template ('mock').

Having determined what the optimal MOI is for infection, the next step was to assess whether there are any differences between the 293T and MDBK cells in infection; both of these cell lines are routinely used in influenza work. While both these cell lines grow in a monolayer, in practice, it was observed that the 293T cells are less prone to remaining adherent (and so, in a monolayer) when handled extensively, which is the case upon

influenza infection; this was believed to be a problem, as infection of cells in a culture dish relies partly on newly produced virus from neighbouring cells. The same cannot be said about the MDBK cells, which are more adherent and more robust when it comes to influenza infection. Both cell lines were used to generate total RNA samples containing influenza RNA, which is why comparing their profiles in infection seemed necessary.

As previously, the cells were infected ($\text{MOI} = 3$) and the total RNA extracted was used in a primer extension assay, which revealed that 293T cells are more susceptible to influenza infection than the MDBK ones (fig. 6A). A quantitative analysis of the results of this assay suggest that the level of all the three viral RNAs found in MDBK cells-extracted total RNA is about a quarter of that extracted from infected 293Ts (fig. 6B); moreover, the abundance of the three viral RNA types (mRNA, cRNA and vRNA) relative to the total RNA detected is consistent within the samples, with mRNA being the most abundant, followed by vRNA and cRNA.

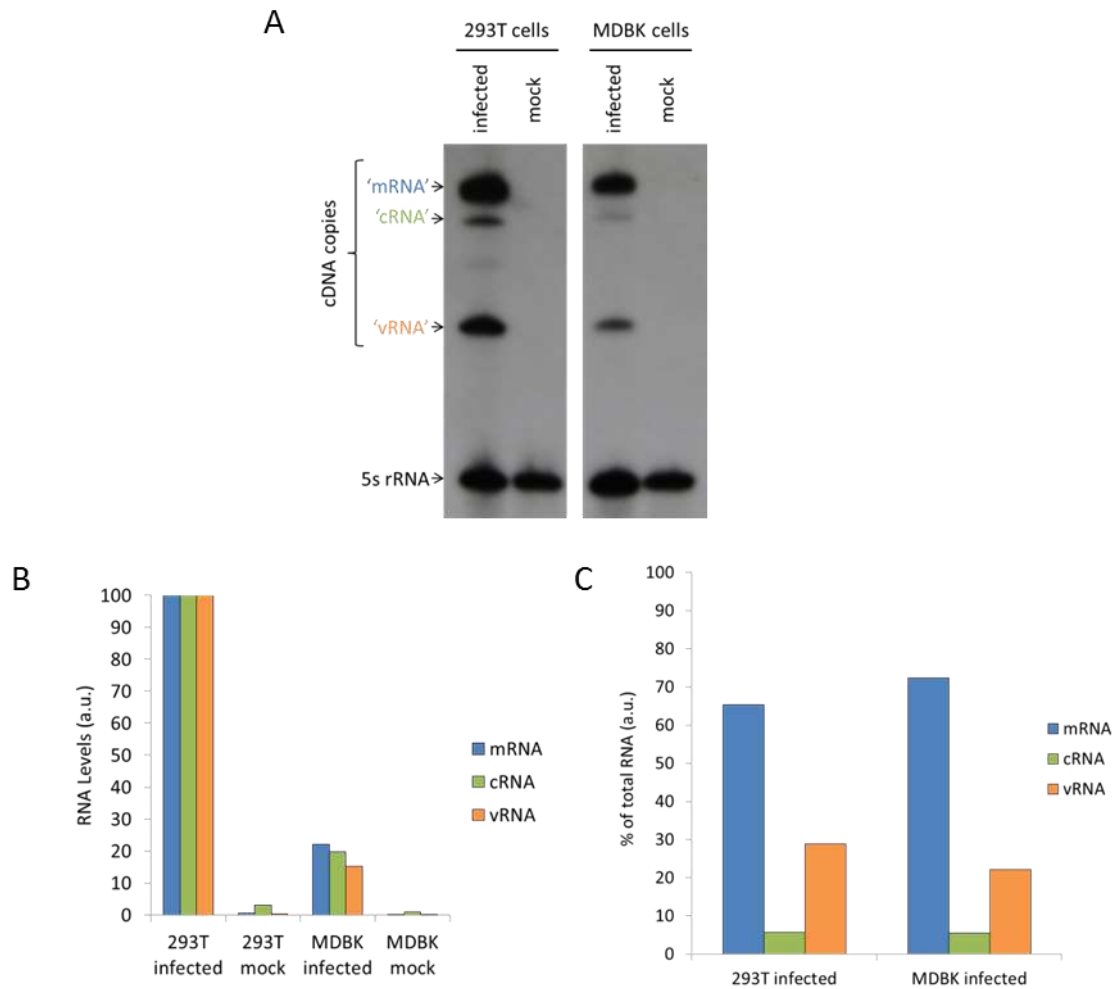


Fig. 6 – A) Autoradiograph of the primer extension assay in which total extracted RNA from non-infected and influenza-infected 293T and MDBK cells was used as a template, with the cDNA products corresponding to the viral RNA species indicated. B) Comparison between RNA levels in 293T and MDBK cells, infected and mock infected. C) Levels of specific RNA species relative to total RNA detected in each cell type.

4.3.2.2 The viral RNA extracted is rich in secondary structure

Another reason why the detection might not be successful is if the targeted RNA to be detected has rich secondary structure around the region to which the detection probe is expected to bind. To test whether this might be the case, I repeated the single-molecule biosensing assay using influenza infected 293T-extracted total RNA that was heated at 95°C for 10 minutes, prior to addition to the glass slide and incubation with the Cy3b-capture probe, followed by Alexa647-labelled detection probe (non-modified backbone DNA probe). Data was taken 5 minutes (fig. 7A) and 30 minutes (fig. 7B) after addition of the DNA Alexa647-labelled detection probe; the data collected revealed that unfortunately, heating the sample did not yield a better result, suggesting that secondary structure might not be an issue. Alternatively, it might be possible that secondary structure is reformed in the short time needed to remove the denatured sample from the heating block and addition onto the slide; hence, denaturing the total RNA in the presence of the probe and annealing might be a better future strategy.

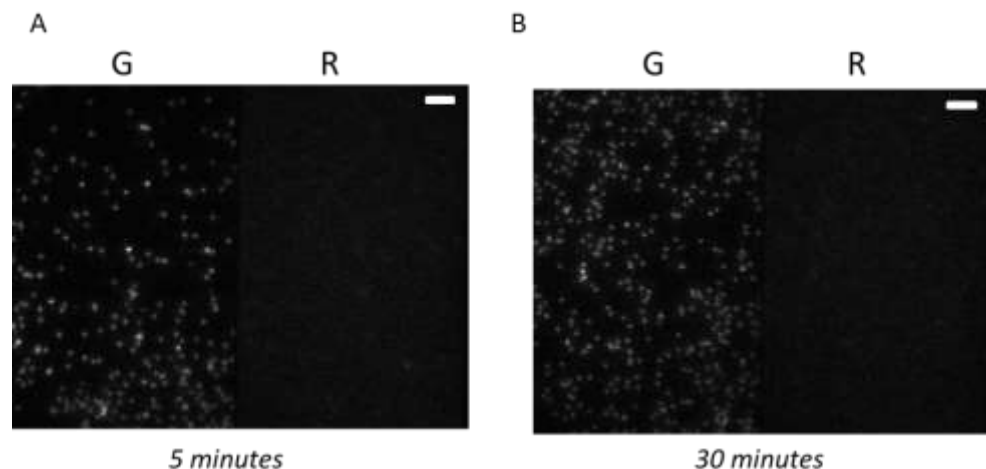


Fig. 7 – Cy3b-labelled RNA capture probe was immobilised on the surface and incubated for 5 minutes (at room temperature) with total RNA extracted from infected 293T cells (denatured at 95°C for 10 minutes and then added straight on the slide), after which it was removed and replaced with 50pM Alexa647-labelled DNA detection probe. A) Field of view captured under green and red illumination, 5 minutes after addition of the detection probe; B) field of view captured under the same conditions, but after 30 minutes since the addition of the detection probe. (Scale bar: 3 μ m)

4.3.2.3 Binding conditions – capture probe

It might be possible that the binding conditions used for the capture probe ‘pull down’ of the viral RNA to be non-optimal. As the capture DNA probe is essentially the same as the 3’ end of the viral RNA, it is complementary to the 5’ end of the viral RNA and hence binding conditions could be investigated by incubating the DNA capture probe with the 5’ end RNA segment (which was used in experiments described in Part A of this thesis).

The biochemical investigation set up to test whether the RNA:DNA hybrid is formed, suggests that the DNA capture probe is able to bind to the RNA when the two strands are incubated together in PBS, at both room temperature and 37°C (conditions which are the same as the ones used when the assay was performed on the surface; fig. 8).

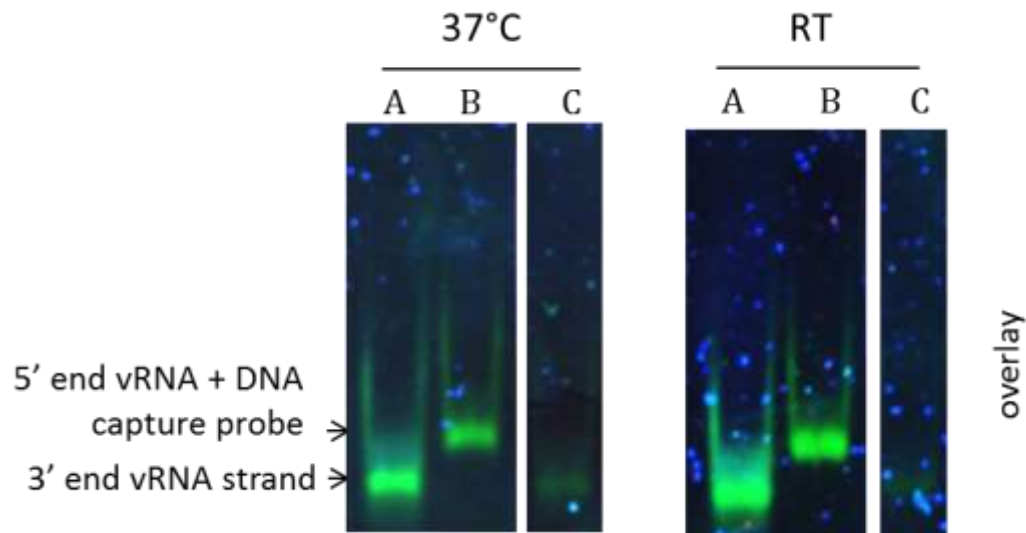


Fig. 8 – A biochemical investigation (15% PAGE) of the DNA capture probe's capacity to bind to a short strand mimicking the end of the viral RNA it is meant to pull down. Lane A: Negative control - 3' end vRNA incubated with capture probe (no binding expected). Lane B: DNA capture probe incubated with the 5' end vRNA short strand (has the same sequence as the mRNA and cRNA that is meant to be recognised by the capture probe); the DNA is unlabelled and 15nt long, while the target RNA is 18nt long. Lane C: control - 3' end of the viral RNA (20nt), this is the mimic of the DNA capture probe (and this strand has also been used as a capture probe in the experiments that use an RNA capture probe).

4.3.2.4 Binding conditions – detection probe

It might be possible that the binding conditions of the DNA detection probe to the immobilised RNA are not optimal and hence no detectable RNA:DNA hybrid is formed. To test whether this is the case, two approaches were employed: a single-molecule fluorescence one and a biochemical one, and both of them were aimed at checking whether the probes are able to bind to a DNA strand with an identical target sequence as the one targeted on the NA/A gene, but that is only 50nt long as opposed to the NA/A gene which is ~1300nt long.

As the availability of the RNA to be detected in the sample was also identified as a possible limiting factor for the success of the assay, employing a complementary DNA to the detection probe and not the extracted RNA as a target, allows the focus to be exclusively on the binding conditions, rather than the nature of the binding partners.

The single-molecule troubleshooting approach relies on detection of the pre-annealed biotinylated target DNA (containing the sequence to be detect on the influenza RNA) with the DNA detection probe (labelled with Alexa647), immobilised on a glass cover slip (fig. 9A). The target DNA: Alexa647 detection DNA probe duplex was annealed in a thermocycler, in annealing buffer (as described in section 4.2.8), to a final concentration of 100nM. When deposited on the surface, the duplex was diluted to 50pM in PBS; after 10 minutes of incubation, data was collected – a signal from the red probe was detected (fig. 9B), suggesting that the DNA:DNA complex is formed.

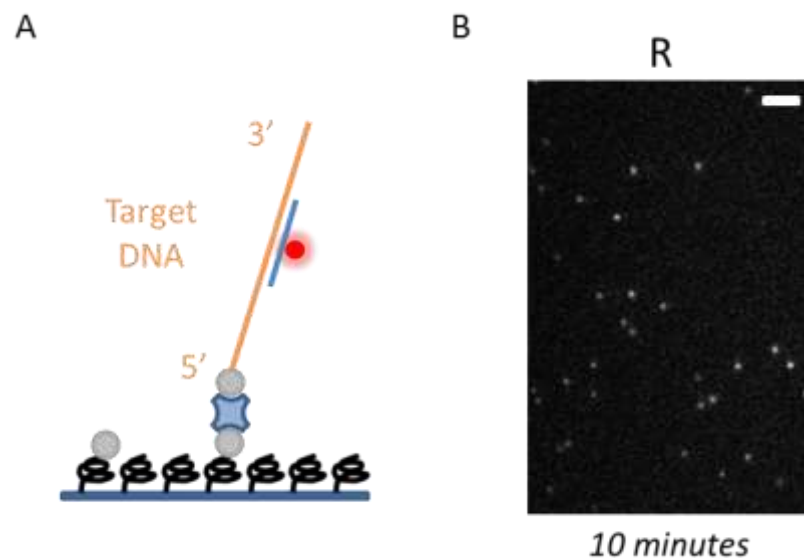


Fig. 9 – A) schematic explaining the experiments setup. B) A field of view taken at 10 minutes after the pre-annealed DNA:DNA duplex was deposited on the surface (scale bar: 3 μ m).

However, the assay relies on the detection probe interacting with the target without the two being annealed at a controlled temperature and for a certain period of time; the next step in the troubleshooting process was to check whether the detection probe can bind to the DNA target in solution: the two nucleic acids were mixed together in annealing buffer or PBS and the samples were incubated for 30 minutes at 37°C, after which they were electrophoresed on a 20% non-denaturing PAGE – all the samples exhibited a band that was interpreted as the DNA:DNA duplex (fig. 10, lanes A&C). Alongside these samples, binding of the target DNA by two shorter, Cy3b- or ATTO647N-labelled probes (targeting the same region of the NA/A gene as the original Alexa647-labelled detection probe), under the same conditions was investigated: a major band of very similar mobility

to the one observed previously could be identified, alongside a minor band which travelled further within the gel (fig. 10, lanes B&D). The overlay of the two images taken to detect the Cy3b and ATTO647N signals, suggests that the detected major band is the DNA-detection probes complex, while the minor band is an excess of the ATTO647N-labelled short probe.

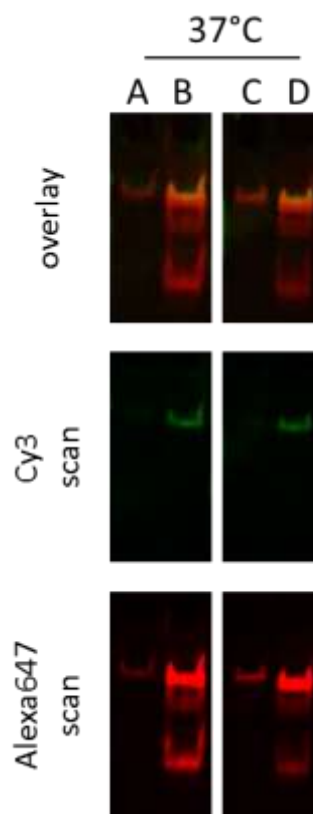


Fig. 10 – A 20% native PAGE for investigation of whether the DNA target: DNA detection probe can form in annealing buffer (lane A) or PBS (lane C), upon incubation at 37°C. Alternatively, the target DNA was incubated with two shorter probes, targetting the same region of the NA/A as the original detection probe and labelled with Cy3b and ATTO647N, signifying that a FRET signal is expected when the two probes are bound to the same segment at the same time; in annealing buffer (lane B) or PBS (lane D), at 37°C. Top: overlay of images taken at 532nm and 640nm laser intensities; 'Cy3 scan' – image resulting after 532nm laser scan; 'Alexa647 scan' – image resulting after 640nm laser scan

Having seen that the probes anneal to the target DNA in both PBS and annealing buffer, the experimental conditions were limited to further investigating the impact of temperature on binding occurring in PBS. The effect of the temperature was further

investigated, with the experiment repeated at both 37°C and room temperature (fig. 11), which confirmed that temperature is not a crucial factor for binding in this case.

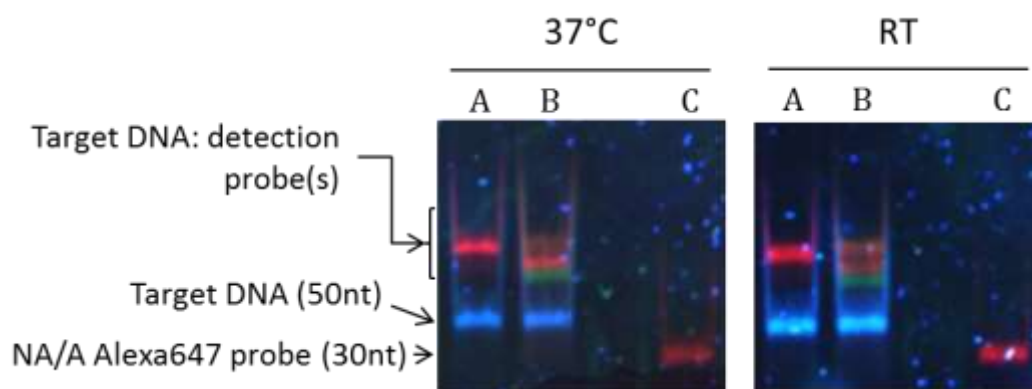


Fig. 11 – 15% PAGE investigating the binding conditions between the DNA target and the different capture probes (an overlay of three different images: in red – Alexa647 scan, in green – Cy3 scan, in blue – SybrGold scan). Lane A: target DNA incubated with a 30nt, Alexa647-labelled DNA probe; the top band (red) corresponds to the complex, the bottom band (blue) corresponds to the unbound target DNA. Lane B: target DNA incubated with two detection probes, one 20nt long (Cy3b labelled) and the other 21nt long (ATTO647N labelled); top band corresponds to the target DNA to which both DNA detection probes are bound; immediately under it, a band corresponding to the target DNA: ATTO647N-labelled probe (red), followed by a band attributed to the target DNA: Cy3b-labelled probe (green); the bottom band (blue) denotes unbound target DNA. Lane C: control lane in which only the 30nt, Alexa647 probe was loaded.

After establishing that the Alexa647-labelled detection probe interacts with a DNA mimic of the targeted region of the NA/A gene, checking whether the DNA probe also interacts with the extracted RNA by taking a classical biochemical approach seemed like the next

step. Specifically, extracted total RNA from influenza infected and non-infected (mock) cells was incubated with detection probes directed to the NA gene of the influenza A virus; in case the probe binds, the double-stranded product would be stable upon digestion with Ribonuclease T1, an enzyme that digests single-stranded RNA after guanine residues. The products of the digestion reaction were electrophoresed on a 2% agarose gel (in the presence of SybrSafe, a nucleic acid stain) and some protected products could be observed in the sample that contained viral RNA and had been incubated with the DNA detection probe (fig. 12); surprisingly, rather than only one protected product, the electrophoresis analysis revealed five such products, which suggests either multiple binding sites for the detection probe or degradation products of the probe that protect slightly different lengths of RNA. No such products were observed in the sample containing extracted RNA from non-infected cells and no fluorescence could be observed from the Alexa647-labelled probe (which is due to the fact that the concentration of the probe is too low to be detected by our scanner).

	x		x	x	total RNA (infected cells)
		x			total RNA (mock)
x	x	x			A647N NA/A detection probe
	x	x		x	RNase T1

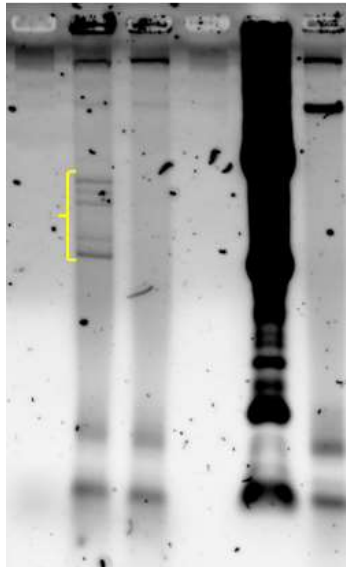


Fig. 12 – T1 ribonuclease digestion assay (image is a SybrGold scan and the identified protected products are marked by the yellow paranthesis).

4.4 DISCUSSION

While the initial experiments presented here suggest that the biosensing assay is functional, they were not reproducible (as explained at the end of section 4.3.1). Hence, I embarked on trying to find the cause which makes this experiment so hard to repeat with successful results.

Identifying three different stages at which the assay could be suboptimal, I investigated each of them individually. I first looked at the availability of the viral mRNA and cRNA in the sample used, comparing the levels of all the different influenza RNAs in a total RNA

extract from infected cells: here I found that MDBK cells are less susceptible to Influenza A infection than 293T cells, so in the future 293T cells are preferable. Also, the dominant viral RNA species is the viral mRNA, followed by vRNA and then cRNA; the probes used in the assay detect positive-strand cRNA and mRNA, and these two species combined represent 70-80% of the viral RNA present in the total RNA extraction, which one would hope is enough to be detected.

I then investigated whether denaturing the total extracted RNA prior to addition of the detection probe could lead to more DNA:RNA hybrids being formed: unfortunately, it didn't seem to make any difference. It is unlikely that the NA/A gene region to which the DNA detection probe is targeted presents very stable secondary structure as it is not a GC-rich regions for which such occurrences have been described previously.

Finally, I looked at whether the binding conditions of the detection probe to the RNA that is to be detected could be suboptimal: investigating binding at both room temperature and at 37°C, in both PBS and annealing buffer (a buffer with a higher salt content), no differences were observed. To simplify the troubleshooting investigation, this study was initially performed on the DNA probe aimed at being used in the assay binding to a complementary DNA target – once it was established that under the described conditions, no differences in DNA:DNA hybrid formation could be detected, the investigation was extended to DNA:RNA hybrids that are true to the ones expected under biosensing conditions. A biochemical/RNase T1 digestion assay yielded several protected products that could not be observed when the DNA probe was incubated with total RNA extracted from non-infected cells, suggesting that the protected products are a result of the

DNA:RNA hybrid occurring; surprising, there were several such products that might point to either multiple binding sites of the probe on the NA/A gene or to a partial degradation of the DNA detection probe (as experiments described in this step were performed with the probe that had no phosphorothioate linkages and so was susceptible to nuclease degradation).

When performing these checks, it has become apparent that the attachment point of the capture probe might also be suboptimal: the capture probe has a biotin residue attached to its 5' end, which suggests that the captured RNA gene will bind with its 3' end towards the glass surface; this means that the bulk viral genomic RNA has its free end facing the glass surface, rather than float freely in solution (as depicted in fig. 1). This being the case, it is likely that the viral RNA segment, if captured, interacts with the surface and is not available for binding by the labelled detection probes.

4.5 PERSPECTIVES

Given all of the above, it is likely that the biosensing assay would benefit from a few changes such as a redesign of the capture probe (by which the sequence remains the same, but the biotin label is moved to the 3' end of the probe) or performing the assay in a buffer formulation that contains formamide, which is known to aid DNA:RNA hybridisation (Casey & Davidson 1977).

As it stands, the biosensor is fit to detect positive polarity RNAs (mRNA and cRNA) found in a sample; however, if this assay is to be used for detection in a medical practice, it would only work on extracted RNA from cells (probably obtained from a swab) and would

not be simple to use (as an RNA extraction step would need to be performed); to detect exhaled virus, the probes would need to be targeted to the negative-sense vRNA – i.e. target the viral genome that is packaged in the virions. This would raise the question of disrupting the virions, on the one hand, and freeing the RNA from the RNPs, on the other. This would of course lead to redesigning not only the capture probe, but also the detection one; moreover, as there is only one copy of each of the influenza genes in a virion, it might be necessary to detect several genes at the same time, in which case using FRET codes might be more advantageous.

In the case that vRNA detection becomes the preferred route for the continuation of this project, *in vitro* proof of principle experiments can be performed on purified RNA from pure virus rather than from infected cells. This approach would be convenient as the virus titre could be adjusted to mimic that of the virus exhaled upon human-to-human transmission (for example), offering the biosensing assay a highly realistic scenario.

Alternatively, if the assay is to be further investigated as it stands (detection of positive-strand RNA), a full length vRNA gene segment could be generated via *in vitro* transcription and used as a target, rather than using total extracted RNA. Additionally, as a similar assay was developed in our research group for detection of rotavirus RNA (Periz et al. 2013), total extracted RNA samples containing influenza RNA could be spiked with a rotavirus gene that could be detected alongside the desired RNA, as an internal positive control.

4.6 BIBLIOGRAPHY

- Blyth, 2009. Rapid-Test Sensitivity for Novel Swine-Origin Influenza A (H1N1) Virus in Humans — NEJM. Available at: <http://www.nejm.org/doi/full/10.1056/NEJMc0909049>.
- Casey, J. & Davidson, N., 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic acids research*, 4(5), pp.1539–52.
- Clark, C.L. et al., 1997. CD, absorption and thermodynamic analysis of repeating dinucleotide DNA, RNA and hybrid duplexes [d/r(AC)]₁₂[d/r(GT/U)]₁₂ and the influence of phosphorothioate substitution. *Nucleic acids research*, 25(20), pp.4098–105.
- Crawford, R., Kelly, D.J. & Kapanidis, A.N., 2012. A protein biosensor that relies on bending of single DNA molecules. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 13(4), pp.918–22.
- Desselberger, U. et al., 1980. The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene*, 8(3), pp.315–328.
- Faix, 2009. Rapid-Test Sensitivity for Novel Swine-Origin Influenza A (H1N1) Virus in Humans — NEJM. Available at: <http://www.nejm.org/doi/full/10.1056/NEJMc0904264>.
- Fields, B., 2007. *Fields virology*, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Gavin, P.J. & Thomson, R.B., 2004. Review of Rapid Diagnostic Tests for Influenza. *Clinical and Applied Immunology Reviews*, 4(3), pp.151–172.
- Hiti, A.L. & Nayak, D.P., 1982. Complete nucleotide sequence of the neuraminidase gene of human influenza virus A/WSN/33. *Journal of virology*, 41(2), pp.730–4.
- Hurt, A.C. et al., 2007. Performance of six influenza rapid tests in detecting human influenza in clinical specimens. *Journal of Clinical Virology*, 39(2), pp.132–135.
- Ivnitski, D. et al., 1999. Biosensors for detection of pathogenic bacteria. *Biosensors and Bioelectronics*, 14(7), pp.599–624.
- Lazcka, O., Campo, F.J. Del & Muñoz, F.X., 2007. Pathogen detection: A perspective of traditional methods and biosensors. *Biosensors and Bioelectronics*, 22(7), pp.1205–1217.
- Lymperopoulos, K. et al., 2010a. Single-molecule DNA biosensors for protein and ligand detection. *Angewandte Chemie (International ed. in English)*, 49(7), pp.1316–20.

- Lymperopoulos, K. et al., 2010b. Single-molecule DNA biosensors for protein and ligand detection. *Angewandte Chemie (International ed. in English)*, 49(7), pp.1316–20.
- Nayak, M. et al., 2009. Detection of microorganisms using biosensors—A smarter way towards detection techniques. *Biosensors and Bioelectronics*, 25(4), pp.661–667.
- Nayak, M. et al., 2013. Integrated sorting, concentration and real time PCR based detection system for sensitive detection of microorganisms. *Scientific reports*, 3, p.3266.
- Periz, J. et al., 2013. Rotavirus mRNAs are released by transcript-specific channels in the double-layered viral capsid. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29), pp.12042–7.
- Skottrup, P.D., Nicolaisen, M. & Justesen, A.F., 2008. Towards on-site pathogen detection using antibody-based sensors. *Biosensors and Bioelectronics*, 24(3), pp.339–348.
- Storch, G.A.M., 2003. Rapid diagnostic tests for influenza. *Current Opinion in Pediatrics*.
- Sugimoto, N. et al., 1995. Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. *Biochemistry*, 34(35), pp.11211–11216.
- Taylor, L.H., Latham, S.M. & Woolhouse, M.E., 2001. Risk factors for human disease emergence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356(1411), pp.983–9.
- Yim, S.W. et al., 2012. Four-color alternating-laser excitation single-molecule fluorescence spectroscopy for next-generation biodetection assays. *Clinical chemistry*, 58(4), pp.707–16.

Chapter 5: Labelling short RNAs using click-chemistry

5.1 INTRODUCTION

5.1.1 Motivation

Francis Crick proposed the Central Dogma of Molecular Biology (1958) in which information is transferred sequentially, residue-by-residue between DNA, RNA and proteins in a process generally referred to as “expression of genetic information”. Originally, Crick proposed that information can flow in all directions through all the nodes (DNA, RNA, protein); subsequent experimental evidence revised the dogma emphasising that the transfer of information once at the protein level is irreversible. The various steps of the transfer are achieved by specialised molecular machines in a highly controlled fashion. In all living organisms, the first step of genetic information transfer is *transcription*, a process in which the RNA polymerase enzyme produces messenger RNA (mRNA).

Much of our current understanding of how transcription occurs in all domains of life has come from employing classical biochemical techniques, such as *in vitro* transcription assays, whose products are analysed under denaturing conditions using gel electrophoresis; such analyses rely on the incorporation of a radiolabelled nucleotide in the generated nucleic acid products. While these assays are reliable (having been thoroughly optimised) and relatively easy to set up, they take a long time until the results become available (with autoradiography of some dilute samples having to take days).

Moreover, safety precautions in place for handling radioactive materials render these assays tedious.

A faster, easier method for conducting these same assays is desired, and with the advancement of fluorescence in recent times, it seems that such alternatives should be readily available.

Detecting DNA and RNA after electrophoresis using nucleic acid stains such as ethidium bromide or Sybr Gold (for DNA) and Sybr Green I & II (for DNA and ssRNA, respectively) could be such an alternative: however, these are intercalating dyes which require that the nucleic acid studied is of a certain minimal length (e.g. 50nt for Sybr Green II), meaning they would not be effective for visualisation of short RNAs (such as abortive transcription products). Moreover, their efficiency is directly linked with whether the nucleic acid under investigation is double-stranded or not, with some of these stains having reduced sensitivities for single-stranded species.

For the purpose of studying transcription, a safe and fast detection method for single-stranded RNA is desired. Since this reaction relies on the building blocks of RNA being supplied in the reaction mixture, it follows that it could be possible to add a modified NTP to the reaction, so that it is incorporated in the growing nucleic acid strand as it is being synthesised; such a modified NTP could be carrying either a reactive handle (such as a thiol-, amine-, alkyne- or azide-reactive group) or an unreactive tag (such as a fluorescent dye or a biotin). Indeed this kind of approach has been investigated, with results suggesting that fluorescently-modified NTPs are harder to incorporate than the NTPs carrying reactive groups. Out of the reactive groups incorporated, amines, alkynes and

azides are the most frequent ones used for nucleic acid labelling (thiol-modified NTPs can also be used).

Using click chemistry to fluorescently label RNAs is far less common than for DNA labelling; this might be explained by the reaction's reliance on copper catalysis, an element known to induce RNA strand cleavage (Dojahn et al. 2013). In the few studies investigating click chemistry for labelling modified RNAs generated using the *in vitro* transcription reactions, the assay was run at much higher concentrations of initiating dinucleotide, nucleotides and polymerase than usually employed for biochemical investigations (Rao, A. A. Sawant, et al. 2012; Dojahn et al. 2013). Successful click reactions were also reported for RNAs as short as 9nt (Paredes & Das 2011), in the presence of copper; in acetonitrile, rather than a biologically compatible buffer; and when the modification on the RNA was introduced post-synthetically (using enzymatic labelling). Paredes & Das performed click labelling on *in vitro* generated RNAs carrying a 5' end modification, using copper-catalysed cycloaddition and in acetonitrile, but on a strand that was longer, measuring 66nt.

In this chapter, I explore the possibility of using alkyne-azide cycloaddition, specifically strain-promoted alkyne-azide cycloaddition (SPAAC; Agard et al. 2004) to label short RNA products as small as 3nt. Single-stranded RNA transcripts were generated through an *in vitro* transcription reaction; the reaction was carried out by the *E. coli* RNA polymerase, using a variant of the *lac* promoter (*lacCons+2*, sequence given in fig. 2) that is known to produce abundant abortive transcripts, which are short RNAs (2-11nt long) that accumulate prior to the enzyme entering elongation (Hsu 2009). To make the short RNA

products amenable to click chemistry, transcription was primed with an initiating dinucleotide that was 5' azide-modified. The products were subjected to labelling with fluorescent dyes (via SPAAC), either before, during or after being separated through denaturing PAGE, each strategy having its own advantages and disadvantages (and which will be discussed in the relevant section), but all leading to the end result much faster than in the case of radioactivity-based transcription assays.

Should a copper-free click chemistry strategy be successful at labelling RNAs of varying lengths and equally sensitive to the currently employed radioactivity-based methods, it would be applicable to various transcription systems other than the bacterial one and hence become a viable alternative to the use of radioactivity in such assays.

5.1.2 Bacterial transcription

The process of transcription has been widely studied in bacterial model organisms such as *Escherichia coli*. The key enzyme responsible for transcription in *E. coli* is a DNA-dependent RNA polymerase (fig. 1) which has 5 subunits ($\alpha^2\beta\beta'\omega$) and is referred to as "core" RNAP. Transcription factors (such as σ) can bind to core RNAP to form a 6-subunit protein called "holoenzyme" (Young et al. 2002), which shows an increased affinity and specificity for the template DNA (Callaci & Heyduk 1998).

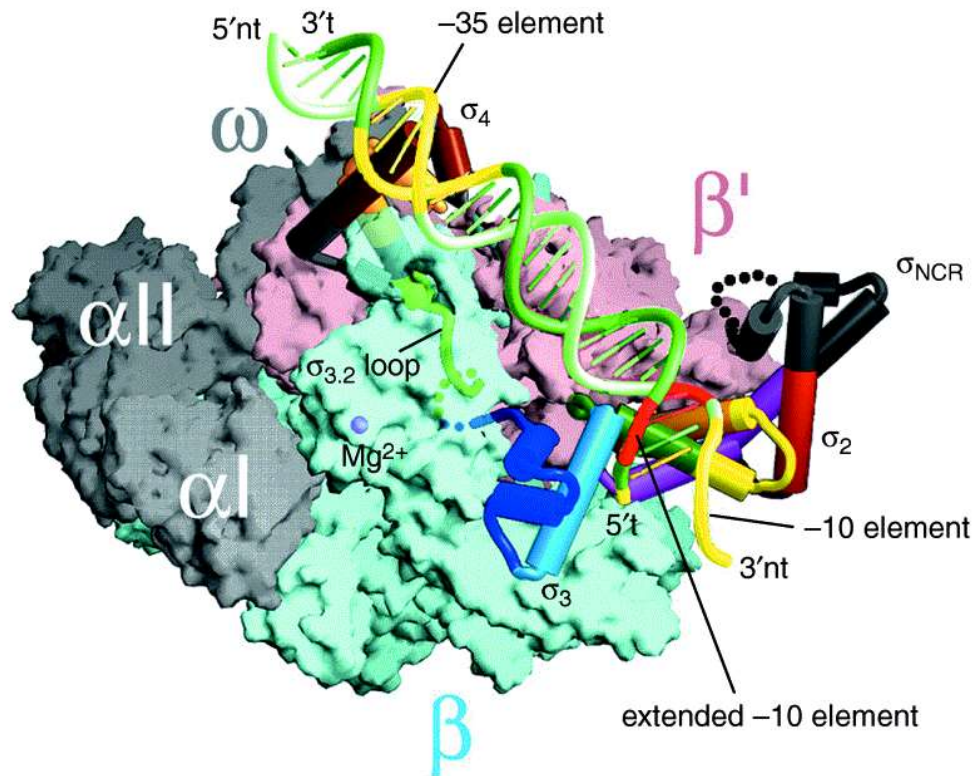


Fig. 1 - The DNA-dependent RNA polymerase from *Escherichia coli* shown in a complex with dsDNA. In the core (shown as a molecular surface), the α and ω subunits serve as scaffold for the assembly of the protein, while the two α subunits are also involved in recognizing a specific DNA sequence known as the promoter (shown as a double helix with the non-template strand in light green, the template strand in dark green, the -35 promoter element in yellow and the -10 promoter element in red). The β subunit is involved in binding the DNA and together with the β' subunit forms the catalytic site of the enzyme. Around the catalytic site, four mobile modules (β_1 , β_2 , the β flap and the clamp) allow for conformational changes when the enzyme is actively involved in transcription. The σ subunit (shown as cylinders) is comprised of four flexibly-linked globular domains: $\sigma_{1.1}$, σ_2 , σ_3 and σ_4 ; between the σ_3 and σ_4 domains a 33-residue linker known as the $\sigma_{3.2}$ loop is found which blocks the exit channel of the RNAP (Young et al., 2002; figure from Murakami & Darst, 2003).

Transcription is a multi-step process which can be described as occurring in three main stages: initiation, elongation and termination.

In initiation, the holoenzyme binds to a specific, conserved DNA sequence (called a promoter) present upstream from a gene (when referring to directionality on the DNA, the transcription start site (+1) is taken as a point of reference¹). Three important features in a promoter necessary for RNAP binding are: the -35 element, the -10 element and the spacer between them. Binding is achieved sequentially, with the σ subunit first contacting the -35 element (via its σ_4 domain), followed by binding the -10 region (through σ_2). To be able to bind these two regions simultaneously, the holoenzyme bends the DNA and the stable DNA-RNAP interaction formed is referred to as the closed complex (RP_c ; Murakami, Masuda, & Darst, 2002).

Transition from RP_c to the open complex (RP_o) is achieved by melting of the promoter dsDNA between positions -12 and +2; this melted region inside the enzyme is known as the transcription bubble. The enzyme undergoes conformational changes to bind the two separated strands. The exposed -10 element on the non-template strand interacts with three conserved amino acids on the σ_2 domain of the σ (tryptophan, tyrosine and phenylalanine), while the other strand kinks over the active-site channel of the enzyme, being accommodated in a positively charged tunnel (Murakami & Darst 2003).

The holoenzyme subsequently undergoes several cycles of abortive initiation, producing short RNAs while maintaining all contacts with the promoter DNA. When synthesising at this stage, instead of translocating along the template, RNAP pulls the downstream DNA inside the bubble, in a mechanism known as DNA scrunching (Kapanidis et al., 2006; Revyakin, Liu, Ebright, & Strick, 2006). While short RNA is being synthesised, increased

¹ Upstream is before this point (starting -1), downstream is after it (starting +2); by convention, there is no position "0"; Pribnow, 1975)

strain accumulates within the transcription bubble which can be relieved either by RNA release via the secondary channel or through displacement of the $\sigma_{3.2}$ loop blocking the RNA exit channel (possible if the RNA is 11-12nt long). Unblocking of the exit channel together with the presence of a 12nt mRNA are thought to destabilise the transcription bubble, forcing the holoenzyme to let go of the promoter contacts, in what is termed promoter escape (Hsu 2002). At this point, the σ subunit could be released from the holoenzyme, but this is not a requirement for progression to elongation.

Elongation is the stage of transcription in which the RNAP translocates along the DNA, copying it into mRNA. During elongation, the enzyme, DNA template and the nascent mRNA are associated in a very stable complex, known as tertiary/transcription elongating complex (TEC). In elongation, the σ factor is stochastically released (Raffaella et al. 2005; Kapanidis et al. 2005).

Termination is the final stage of transcription, in which the polymerase stops synthesising mRNA, releases it and the RNAP dissociates from the DNA. Once transcription is completed, the polymerase is able to rebind σ and reinitiate transcription.

5.1.3 Fluorescently labelling short RNAs

Generally, RNAs can be modified to carry a fluorescent label (or a reactive group that can be used to attach a fluorescent dye) at either their 5' or their 3' ends, or internally.

Two main strategies of introducing modifications in RNA can be distinguished: the first relies on adding the desired modification post-synthesis, by using enzymes that can add a modified nucleotide to the 3' end of the RNA strand (such as the poly(A) polymerase or the poly(U) polymerase. The second strategy relies on using a modified base during

synthesis, regardless of whether during solid-phase synthesis or enzymatic production of RNA. Amino-modifications are the most commonly employed, but also alkyne- and azide-modifications, which lend the RNA to labelling via click chemistry (Winz et al. 2012). This approach can be adopted for producing RNAs that are modified at either ends or internally alike. Internally modified nucleic acids can also be generated by using splint ligation, via a longer DNA template that acts as a scaffold for the ligation reaction, from an end labelled RNA that is capable of reacting with another RNA strand (Kurschat et al. 2005; Stark et al. 2006; El-Sagheer & Brown 2010).

5.1.3.1 Enzymatic labelling

The enzyme most commonly used for 3' end labelling of RNA strands is the poly(A) polymerase (Martin & Keller, 1998; Lingner & Keller, 1993). Of the two poly(A) polymerases routinely used, the yeast poly(A) polymerase is known to be both more productive and faster than the *E.coli* poly(A) polymerase, which is known to incorporate fluorescently-labelled or biotinylated ATP with a low efficiency. Both of these enzymes require the presence of a divalent cation (Mn^{2+} or Mg^{2+}) and are better at attaching the labelled nucleotide if the substrate RNA is longer (for the *E.coli* enzyme, at least 20 nucleotides long, but no lower limit indicated for the yeast enzyme). Another such enzyme that can be employed for this scope is the poly(U) polymerase (Wickens & Kwak, 2008; Kwak & Wickens, 2007).

Other enzymes that can be used for 3' end labelling on RNAs are T4 RNA ligase (for which the literature suggests a trinucleoside diphosphate with a 3' terminal hydroxyl to be a

minimal substrate- England & Uhlenbeck, 1978) and terminal deoxynucleotidyl transferase or TdT (Zhao et al. 2012; Rosemeyer et al. 1995).

For 5' end enzymatic labelling, one can use either a T7 RNA polymerase (Li et al. 2005) or a T4 RNA ligase (Kinoshita et al. 1997).

5.1.3.2 Click chemistry strategies

In the last decade, a class of bio-compatible reactions have been described and thoroughly investigated: they are characterised by high yields, efficiency and specificity (limited side-reactivity), progress in water, are fast and, most importantly, orthogonal, meaning that they do not interfere with any other known biological functional groups (Best 2009). These reactions are classed under 'click chemistry' and rely on the use of an azide functional group that is a small, highly reactive and specific toward its partner, which is most often an alkyne.

The first click chemistry reactions to be described in the context of labelling biological molecules was the modified ***Staudinger ligation*** in which an azide group reacts with a triarylphosphine; the reaction was first described in 2000 and was used to label a mammalian cell with biotins, by reacting an azide-modified sugar presented on the cell surface with biotinylated phosphine (Saxon & Bertozzi 2000). This strategy is most often employed for generation of modified peptides and for *in vivo* labelling (Köhn & Breinbauer 2004), but it has also been used for post-transcriptionally labelling RNAs that have been synthesised in the presence of an azide-UTP nucleotide (Rao, A. a Sawant, et al. 2012).

Shortly after the bio-compatible Staudinger ligation was reported, two independent research groups publish a variation of the Huisgen cycloaddition that relies on a molecule carrying an azide-moiety reacting with another containing an alkyne group in the presence of copper Cu^I catalyst (Rostovtsev et al. 2002; Tornøe et al. 2002). This reaction is known as the **copper-catalysed azide-alkyne cycloaddition (CuAAC)** and has found applicability in *in vitro* biological studies, as its reliance on high concentrations of copper make it unsuitable for use in living organisms (Phelps et al. 2012). The CuAAC strategy is probably the most used click reaction for RNA studies, having been employed to fluorescently label single-stranded RNAs that ranged in length from 9nt (Paredes & Das 2011) to 200nt long (Samanta et al. 2013) at either end or at a well-defined internal position, both *in vitro* (Dojahn et al. 2013; Rao, A. A. Sawant, et al. 2012; Winz et al. 2012; Paredes & Das 2011) and in fixed cells (Jao & Salic 2008).

Another azide-alkyne cycloaddition was described by the Bertozzi group, a reaction in which the alkyne is incorporated within a strained cyclooctyne system which leads to a much higher reactivity towards azide groups and removes the requirement for copper catalysts (Agard et al. 2004). The **strain-promoted (copper-free) azide-alkyne cycloaddition (SPAAC)** has only recently been used for RNA labelling, in a study that relied on enzymatic tailing to introduce a reactive azide-modified base at the 3' end of the nucleic acid which was subsequently reacted with a strained alkyne-modified fluorescent dye (Winz et al. 2012). This latter strategy, SPAAC, is the strategy of choice for the experiments presented in this chapter.

5.2 MATERIALS AND METHODS

5.2.1 Nucleic acid strands

A dsDNA construct, corresponding to the wild-type *lacCons+2* promoter, with the -35 and -10 elements mutated to consensus and separated by a 17bp spacer (fig. 2) was used as a template in the *in vitro* transcription (Kapanidis et al. 2005).



Fig. 2 – The *lacCons+2* promoter, with consensus -35 and -10 elements boxed. The green box indicates the position of the transcription start site (position +1). The full length product expected is 25nt long (green bar); while the most frequent abortive products observed are 5nt, 6nt, 7nt long (blue bars; Robb et al. 2013).

In order to assess the length of the products produced in the *in vitro* transcription assay, three RNA strands of varying lengths, all of which carry a fluorescent dye label (labelled as described in section 2.2.3), were used as markers; alternatively, custom synthesised RNA strands labelled as described in section 5.2.3 were employed (table 1).

Strand	Length	Sequence (5'→3'; modified residue in bold)
RNA17	17nt	AGCGAAAGCAGGAGUU(U-ATTO647N)
RNA18 (no label)	18nt	AGUAGAAACAAGGAGUUU
RNA18	18nt	(ATTO647N-U)AAACUCCUGCUUUCGCU
RNA7	7nt	(azide-A)AUUGUG
RNA25	25nt	(azide-A)AUUGUGAGAGCGGAUAACAAUUUC

Table 1 – Characteristics of the RNA strands of varying lengths, labelled with either Cy3, ATTO647N or Cy5 (the RNA strands carrying the azide-modification), which were used as markers in the electrophoresis runs shown in this chapter

5.2.2 Specific click chemistry reagents

Two different fluorescent dyes, functionalised to be amenable to click chemistry were used in this investigation: dibenzylcyclooctyne(DBCO)-Sulfo-Cy5 (Jena Biosciences; fig. 3A), dissolved in DMSO and stored as 10µl aliquots of 10mM each at -80°C and bicyclo[6.1.0]nonyne(BCN)-Cy3 (kindly provided by Tom Brown, University of Oxford; BCN moiety shown in fig. 3B), also solubilised in DMSO and stored as 3µl aliquots of 50nmol each at -80°C.

In the *in vitro* transcription reactions, a 5' azide-modified ApA (kindly provided by Tom Brown and synthesised by Afaf El-Sagheer, University of Oxford) initiating dinucleotide was used (fig. 3C); the reagent was suspended either in mqH₂O or in TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0) to a final concentration of 10mM.

For post-labelling clean-up, azide-modified agarose (22µmol/ml; Jena Biosciences) was employed (fig. 1D); this reagent is supplied as a suspension in 20% ethanol and was buffer exchanged into TE prior to usage.

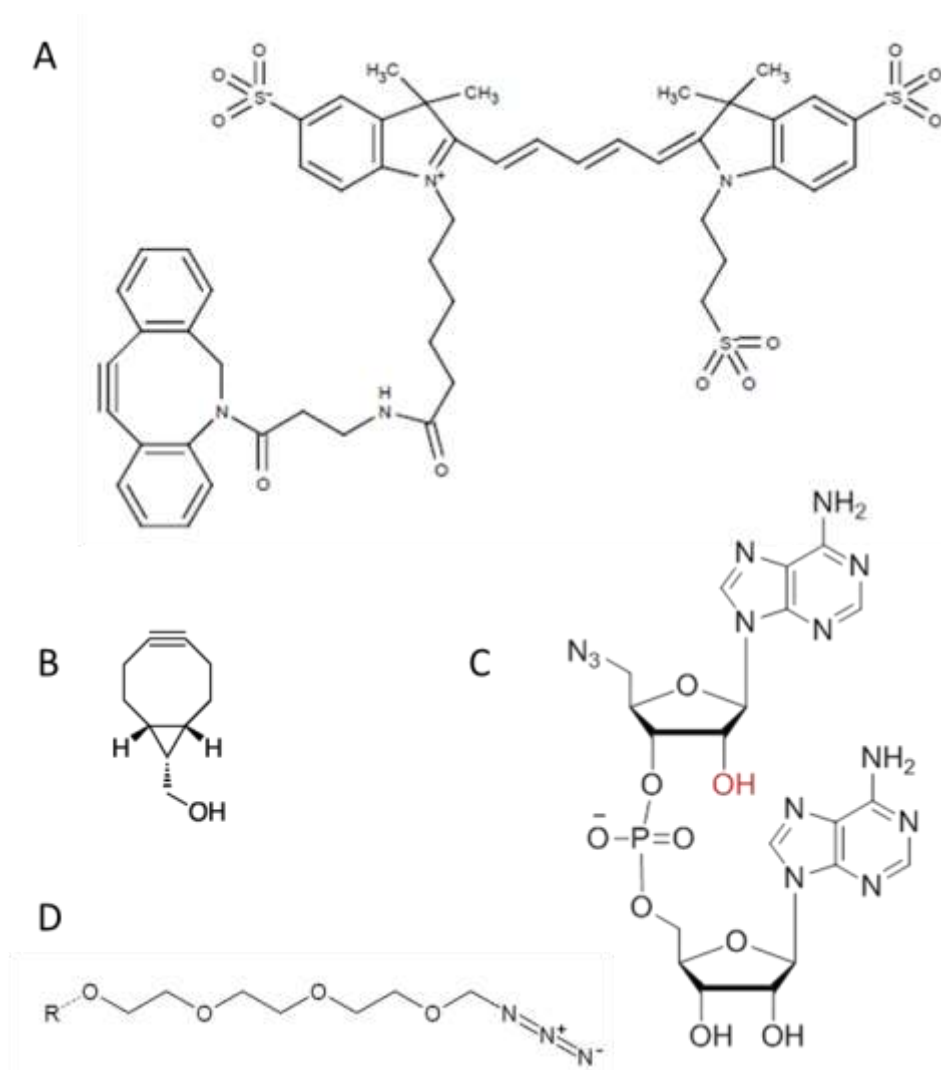


Fig. 3 – A) Chemical structure of the dibenzylcyclooctyne-Sulfo-Cy5 (DBCO-Sulfo-Cy5) reactive dye, as given by the manufacturer. B) Bicyclo[6.1.0]nonyne (BCN) reactive moiety (not conjugated with a fluorescent dye). C) Structure of the initiating dinucleotide ApA, carrying the azide reactive group on the sugar of the 5' most nucleoside; the presence of the hydroxyl group on the sugar also highlights that the ApA used here is RNA (A. El-Sagheer, University of Oxford). D) Azide-modified agarose, as depicted by the manufacturer.

5.2.3 Click chemistry labelling of standard RNAs

The two azide-modified RNA strands described in table 1, supplied lyophilised, were resuspended in TE buffer to a final concentration of 200µM. Each of the strands (final concentration: 50µM) was reacted with DBCO-Sulfo-Cy5 supplied in a 10x excess, in TE buffer supplemented with 20U RNasin (40U/µl; Promega), and incubated for 1 hour at room temperature. Post-incubation, the labelling reaction was split in two equal parts; one part was challenged with 1µl azide-agarose and incubated further at room temperature, 450rpm shaking for 4 hours, after which both samples were stored at 20°C.

5.2.4 Open complex formation

A 5x mix containing transcription-competent RNA polymerase – DNA complexes was set up by incubating 250nM *E.coli* RNA polymerase holoenzyme (250U; Epicentre) with 50nM unlabelled lacCons+2 DNA in 1x KG7 buffer (40 mM Hepes–NaOH pH 7, 100 mM potassium glutamate, 10 mM MgCl₂, 1 mM DTT, 100 µg/ml bovine serum albumin, 1 mM mercaptoethylamine and 5% glycerol) to a final volume of 5µl at 37°C for 20 minutes. To remove any non-specific binding of the RNA polymerase to the DNA and remove unbound protein, the reaction was challenged using 0.5µl of 100mg/ml heparin sepharose (GE Healthcare); upon addition of the heparin sepharose, the reaction was mixed by pipetting up and down (10 times) and then incubated for 30 seconds at 37°C, after which the samples were centrifuged for 10 seconds. At this point, 3µl of the resulting supernatant was transferred to a fresh tube.

5.2.5 *In vitro* transcription (IVT) reaction

The *in vitro* transcription reaction was set up by mixing 1x E.coli RNA polymerase – DNA template open complex (formed as described) with 500μM 5' azide-modified ApA in 1x KG7 buffer supplemented with 50μM each of UTP, CTP, GTP, ATP in the presence of 12U RNasin. The reaction was incubated at 37°C for 30 minutes and was stopped by incubation at 95°C for 10 minutes and then placed on ice, unless it was stopped by addition of 1x reaction volume of 90% formamide loading buffer.

Alternatively, the IVT assay was set up as before, using an unmodified or 5' azide-modified ApA initiating dinucleotide in the presence of 50μM each of UTP, CTP, ATP and 0.2μCi/μl α-³²P GTP (10μCi/μl, Perkin Elmer) and incubated for 10 minutes at room temperature. The reaction was stopped by addition of one reaction volume of loading dye (90% (v/v) formamide, 10mM EDTA, 0.04% (w/v) bromophenol blue and 0.04% (w/v) xylene cyanol FF) and incubation for 5 minutes at 95°C; the sample was electrophoresed on a 7M urea/20% polyacrylamide sequencing gel (3 hours at 1200V) and visualised by autoradiography.

5.2.6 Click chemistry labelling of IVT RNA products and clean up

IVT reactions were incubated with the desired excess of reactive dye and the click chemistry was allowed to proceed for 1 hour at room temperature.

If the reaction was further subjected to a clean-up by reacting the excess dye with azide-modified agarose, the sample was incubated with 44nmol azide-agarose, buffer exchanged in TE (22μmol/ml; Jena Biosciences) at room temperature with 450rpm shaking for 4 hours, after which the supernatant was recovered, added to 90% formamide

loading buffer, heated at 95°C for 5 minutes and analysed on a 7M urea/20% polyacrylamide sequencing gel (3 hours at 1200V) and visualised on a BioRad Molecular Imager PharosFX System (detection: Cy5 – 635nm excitation/695nm emission; Sybr Green II – 488nm excitation/530nm emission; Cy3 – 532nm excitation/ 605nm emission).

5.2.7 Silver staining and Sybr Green II staining

The silver staining procedure was performed using the SilverXpress kit (Life Technologies); the staining protocol was modified from the manufacture's guide as follows: the gel was fixed in a solution containing 50ml 100% methanol, 10ml acetic acid and 45ml mqH₂O, for 20 minutes with shaking, after which it was sensitised in 50ml methanol, 2.5ml sensitizer and 45ml mqH₂O for 1 hour (with shaking). The gel was then washed twice in 100ml mqH₂O, for 10 minutes each time and then stained by incubation for 15 minutes in a solution of 2.5ml Stainer A and 2.5ml Stainer B in 45ml mqH₂O, followed by washing twice, 5 minutes each time, in water. The gel was immersed in developing solution (2.5ml developer, 47.5ml mqH₂); the reaction was allowed to proceed until the bands of interest were clearly seen on the gel then stopped by addition of 2.5ml stopper. Before photographing (in ambient conditions), the gel was washed three times with 100ml mqH₂O, for 10 minutes each time.

Sybr Green II (Life Technologies) staining was performed according to the manufacturer's instructions, with the stain diluted 10,000x in 1x TBE and left at room temperature, covered for 15 minutes with gentle shaking, before imaging.

5.3 RESULTS AND DISCUSSION

5.3.1 Transcription profiles: unmodified ApA vs. 5' azide-modified ApA

As the scope of this project is to generate 5' azide-modified short RNA strands (abortive products of an *in vitro* transcription reaction), it was necessary to check whether transcription proceeds as expected when a 5' azide-modified initiating dinucleotide is used, on the one hand, and whether the transcription profile generated is similar when transcription is carried out in the presence of the unmodified or the azide-modified ApA.

When transcription on the *lacCons+2* promoter is allowed to proceed using an unmodified ApA dinucleotide, in the presence of radiolabelled GTP, the major abortive products expected to arise are a 5mer & a 6mer (which have been seen to run as a double-band) and a 7mer. Previously, the reaction was also performed in the presence of a subset of nucleotides, conditions that will regulate the maximum length products produced (for example, if no ATP is supplied in the reaction mix, the maximum length product generated will be a 7mer) and hence allows for assigning lengths to the abortive products identified. Longer products than the 7mer can also be seen in the transcription profile: specifically, an 11mer, which is considered to be the last abortive product produced before the polymerase enters elongation; a 16mer and a 17mer, which are the major products generated when the polymerase pauses on the DNA; and finally, a 25mer, which is the run-off product (Cordes et al. 2010).

Promoter pausing is an event known to occur in early elongation and believed to be brought about by a DNA sequence that is similar to that of the -10 promoter element, a sequence present on the construct used here (pos. -3, -2 and +1 to +6 form the extended

-10-like promoter sequence; Hsu 2002; Leibman & Hochschild 2007). Besides generating these paused products, the promoter used in these experiments has another particularity: while polymerase initiates transcription at position +1 in most cases, it can also initiate at other positions, usually at purines found upstream from the +1 initiation site, such as position -1 on the promoter used here (Robb et al. 2013). When transcription is carried out in the presence of radiolabelled GTP and the polymerase initiates at position -1, shorter abortive products, previously not detectable as they do not incorporate the label (such as those with lengths between 2-4nt) can also be identified; in this instance, the run-off product will be a 26nt long RNA.

Setting up a transcription reaction, using an unmodified or a 5' azide-modified ApA, in the presence of a radiolabelled GTP, revealed that the two transcription profiles are similar (fig. 4): the two major abortive products measuring 5nt and 6nt in length, run together and can be identified in both profiles, as can the 7nt RNA (which is more prominent in the modified-ApA reaction). A series of bands corresponding to longer products can also be identified when azido-ApA has been used. This intensity difference between the two transcription profiles in the case of longer RNA products could be an indication that when the modified dinucleotide is used, the protein escapes the promoter more efficiently, resulting in more transcription events, but maybe not in full length transcription. Indeed, it is not surprising that there isn't a large amount of full length RNA produced in this reaction, as promoter escape is rare in the case of the *lacCons+2* construct used (this DNA template is routinely employed to study abortive initiation).

Yet, when it comes to considering the abortive products that can be identified (fig. 4), the two experimental conditions seem to yield very similar results up to the 7nt mark (and identifying these very short RNAs is the goal of the assay described here).

A caveat that needs to be kept in mind when interpreting the work presented in this chapter refers directly to the signal detected when radioactivity is used as a reporter – the signal will be additive, with longer products generating stronger bands in the gel, as they have several labels incorporated. This will not be the case for the azide-modified RNAs labelled using click chemistry which will always carry just one label, leading to direct correlation between the intensity of the signal detected and the abundance of the product in the sample.

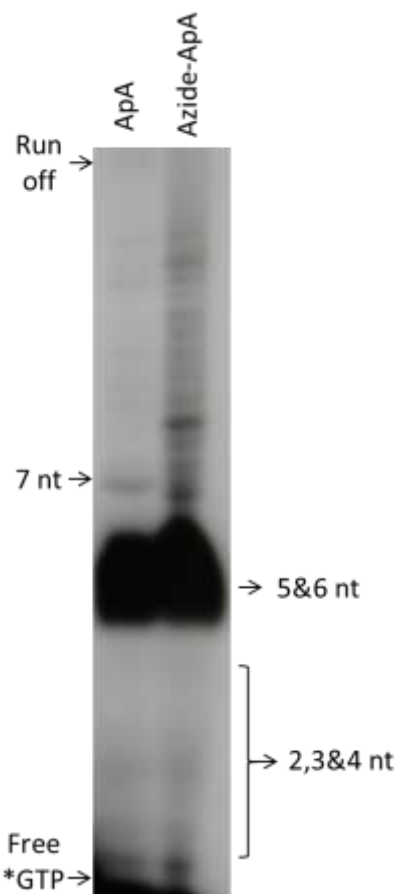


Fig. 4 – *In vitro* transcription products analysed on a 7M urea/20% PAGE and visualised using autoradiography.

5.3.2 Click chemistry labelling of RNA abortive products post electrophoresis

The first strategy investigated in this study deals with the click chemistry being performed post electrophoresis, in the manner of a stain. The azide-modified products of an *in vitro* transcription assay were run on a denaturing PAGE alongside an unmodified, unlabelled RNA strand. Post-electrophoresis, the gel was immersed in a 200nM DBCO-Sulfo-Cy5 solution in 1x TBE and left for 1 hour at room temperature: this resulted in the gel being uniformly stained with the Cy5 dye (fig. 5A). As no individual bands could be detected, the

gel was de-stained in methanol (which is also a fixing agent for nucleic acids post-electrophoresis) and re-imaged for Cy5 fluorescence: while the background fluorescence was largely removed, no individual bands could be detected (fig. 5B). To check whether there is still any RNA left on the gel after the staining/de-staining procedure, the gel was subjected to Sybr Green II staining, which made visible several bands – three in the lane corresponding to the IVT reaction and one corresponding to the unlabelled RNA18 marker (fig. 5C). As seen previously, the largest band detected in the lane onto which the transcription products were loaded maps very close to the bottom of the well (but it can be clearly distinguished from it). Also, as Sybr Green II is not meant to be able to stain any proteins, it is very likely this is a genuine RNA product which had not been fluorescently tagged upon incubation with the click reactive dye.

While optimisation of this strategy might yield satisfactory results, it is hard to imagine it as an easy-to-use procedure since it would require repeated manipulation of the gel; as the aim of the biochemical assay is to obtain transcription profiles that exhibit the best separation possible between the RNA products, the use of large format sequencing-type gels is necessary, which makes click chemistry staining impractical. Yet, while not optimal for identifying the products of the transcription assay described here, using click chemistry in the manner of a stain could still prove useful for other applications, as staining is the most wide spread method of visualising nucleic acids post-electrophoresis.

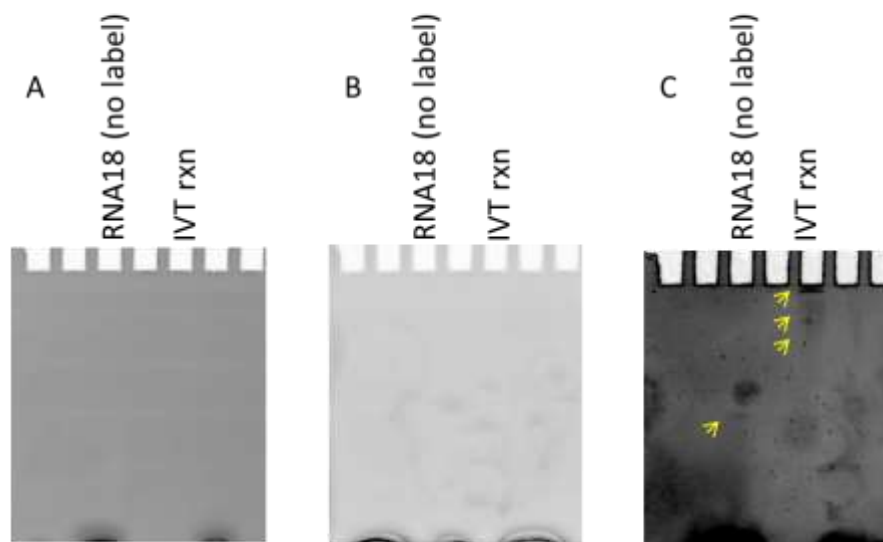


Fig. 5 – The IVT reaction (containing azide-modified RNA products) was electrophoresed on a 7M urea/20% ‘mini’ PAGE; at the end of the electrophoretic run, the gel was first stained with a 200nM DBCO-Sulfo-Cy5 solution in 1x TBE buffer (1 hour staining time), after which a Cy5 scan was performed (A). Post-staining, the gel was fixed and de-stained in 40% methanol (1 hour incubation time, then washed with mqH₂O) and imaged once more for Cy5 fluorescence (B), after which it was stained with Sybr Green II nucleic acid stain (according to the manufacturer’s protocol) and imaged for Sybr Green signal detection (C). The yellow arrowheads indicate bands identified as a result of the Sybr Green II staining.

5.3.3 Click chemistry labelling of RNA abortive products during electrophoresis

A second approach to label short RNAs using click chemistry relies on performing the click chemistry during electrophoresis, by adding the reactive dye to the running buffer; the dye will then react with the RNA products of interest which are partially separated in the gel matrix. This strategy excludes the need to manipulate the large format gel once the electrophoretic run is complete, might remove any additional washing steps (that could

lead to short RNA diffusion out of the gel matrix) and could prove to be a way of bypassing the accumulation of excess free dye in the gel.

To test whether this is indeed a viable option, a transcription reaction was set up as described in section 5.2.5, stopped by addition of one reaction volume of formamide, denatured and loaded on a denaturing PAGE gel (fig. 6); after the run was allowed to proceed for 30 minutes, the 1x TBE electrophoresis running buffer from the upper gel tank was replaced with a solution of DBCO-Sulfo-Cy5 in 1xTBE, which would allow the dye to enter the gel and migrate with the running buffer, which is indeed what can be seen when the gel is imaged for Cy5 fluorescence (fig. 6A).

To avoid any solubility problems that might arise should the labelling reaction occur in the well, I decided to introduce a lag between the start of the electrophoretic run (and the beginning of the RNA products' migration) and the addition of the dye. Since the region of the gel through which the dye in the running buffer had migrated was over-stained, I decided to try to visualise the RNA in a different manner.

Silver staining is a well-known staining procedure for protein detection; however, it can be also used for detection of both DNA and RNA nucleic acids (Mitchell et al. 1996; Boulikas & Hancock 1981) – the staining principle is the same and the procedure is known to be 2-5 times more sensitive than ethidium bromide. Moreover, as the silver stain signal is not dependent on the staining agent intercalating between the nucleic acid's bases, but relies on its chemical properties (that are the same between single and double-stranded RNA or DNA), there is no length requirement for the stain to be effective (Boulikas &

Hancock 1981). Given all this, I decided to use this method to detect the RNA products present in the gel.

Several bands became visible after the silver staining procedure – several in the IVT reaction lane and one corresponding to the RNA17 marker (fig. 6B, arrowheads); unsurprisingly, the major band detected in the IVT reaction lane maps to the bottom of the well (fig. 6B, 'IVT rxn' lane, top arrow) and is most probably given by the denatured RNA polymerase protein used in the transcription reaction – the protein is much too large to be able to enter the gel and hence is expected to aggregate in the well. In addition, three other bands assigned to RNA transcription products could also be detected (fig. 6B). Post-staining, the gel was also probed for Cy5 fluorescence: background fluorescence at this point was much reduced (compare fig. 6A to 6C), allowing for one band to be detected – this is clearly seen at the bottom of the well. I speculate this is probably given by azide-modified RNA products prevented to enter the gel by the obstruction posed by the large protein (fig. 6C, arrowhead in 'IVT rxn' lane).

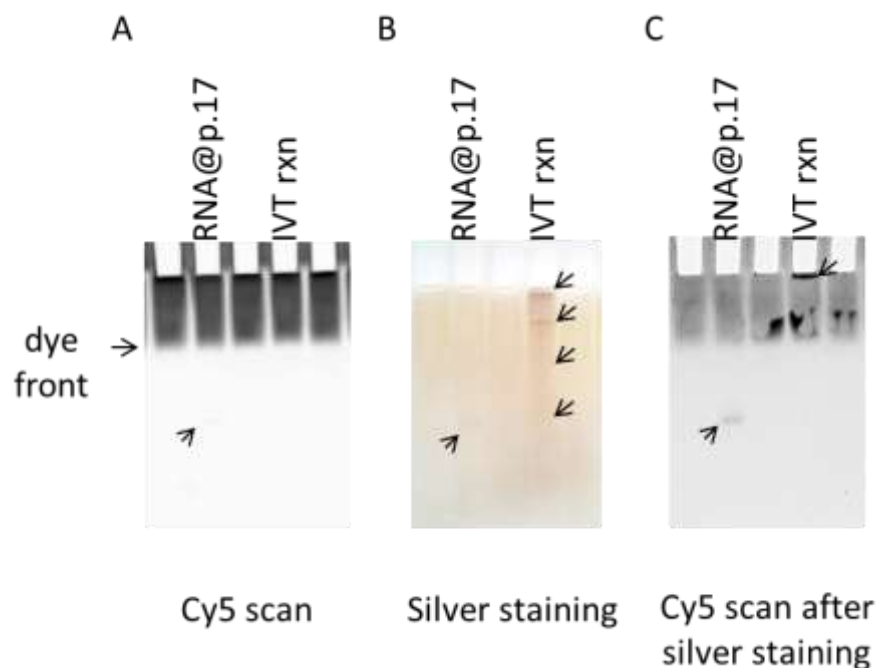


Fig. 6 – The unlabelled, 5' azide-modified RNA products resulting from the transcription assay were electrophoresed on a 7M urea/20% 'mini' PAGE gel that was allowed to proceed for 30 minutes at 200V, after which the buffer in the upper gel tank was replaced with a 300nM DBCO-Sulfo-Cy5 solution in 1xTBE; the run was resumed and allowed to proceed for 50 minutes (at the same voltage). Immediately after the run was concluded, the gel was imaged for Cy5 fluorescence (A), after which it was subjected to silver staining (as described in section 5.2.7) and imaged once more, both for the silver stain (B) and the Cy5 fluorescence after the staining procedure (C). In each image, the arrowheads indicate the position of the RNA17 marker or that of the detected species in the lane corresponding to the transcription reaction.

While for now this strategy has not led to efficiently labelling the short RNAs produced in the IVT assay, it doesn't necessarily mean it should be abandoned: indeed, adding the dye in the buffer from the inception of the electrophoretic run might yield better results. Alternatively, removing the protein from the reaction prior to electrophoresis might lead to more RNA entering the gel, leading to the click reaction producing a higher yield of labelled RNA products.

5.3.4 Click chemistry labelling of RNA abortive products before electrophoresis

Finally, the last approach taken for short RNA labelling was to perform the click reaction immediately after the transcription assay and before electrophoresis. As the concentration of the RNA products is unknown, the assumption that all the ApA available in the reaction is being used up was made; hence, the dye concentration employed was an excess relative to the 500 μ M ApA supplied in the reaction.

Initially, this approach was tested using the BCN-Cy3 dye, which is not sulfonated and carries a net positive charge that adversely influences its propensity to be used in electrophoresis analyses, forcing it to run out of the gel's wells. The experiment was repeated in the presence of DBCO-Sulfo-Cy5 dye (15x excess); post-labelling the sample was diluted and analysed by electrophoresis (fig. 7). The free dye runs as three separate bands (fig. 7A): two of these bands were seen to migrate in the upper half of the gel, while the third was detected close to the bottom; this latter band is most probably given by the unconjugated Cy5 dye, while the other two bands might correspond to the reactive DBCO-Sulfo-Cy5 species and one of its synthesis by-products. If the two free dye bands seen on the upper half of the gel in the lanes that contain only the free dye and those in which the IVT reaction was run are compared (fig. 7A), their intensities differ, with the lanes in which the click reaction was separated being stronger; this might suggest that some labelled RNAs exhibit the same mobility as the free dye.

Additional products can be detected in all the reactions in which the 'clicked' IVT was run (fig. 7B&C); all the identified products run beneath the 18nt marker, suggesting that these

are indeed abortive products. Caution needs to be exerted when comparing the RNA18 standard with the expected IVT products; it is likely that the marker and the clicked RNA products will exhibit different mobilities on the gel. The standard RNA18 is an amino-modified RNA labelled with ATTO647N (with the modification being attached to the RNA via a linker of six carbon atoms), while the IVT products carry the DBCO-Sulfo-Cy5 group, which is much bulkier.

In view of this and taken into account the expected abortive profile of the promoter used (explained in detail in section 5.3.1), it is possible that the major bands identified in fig. 7B&C are given by abortive products that are 5nt and 6nt long.

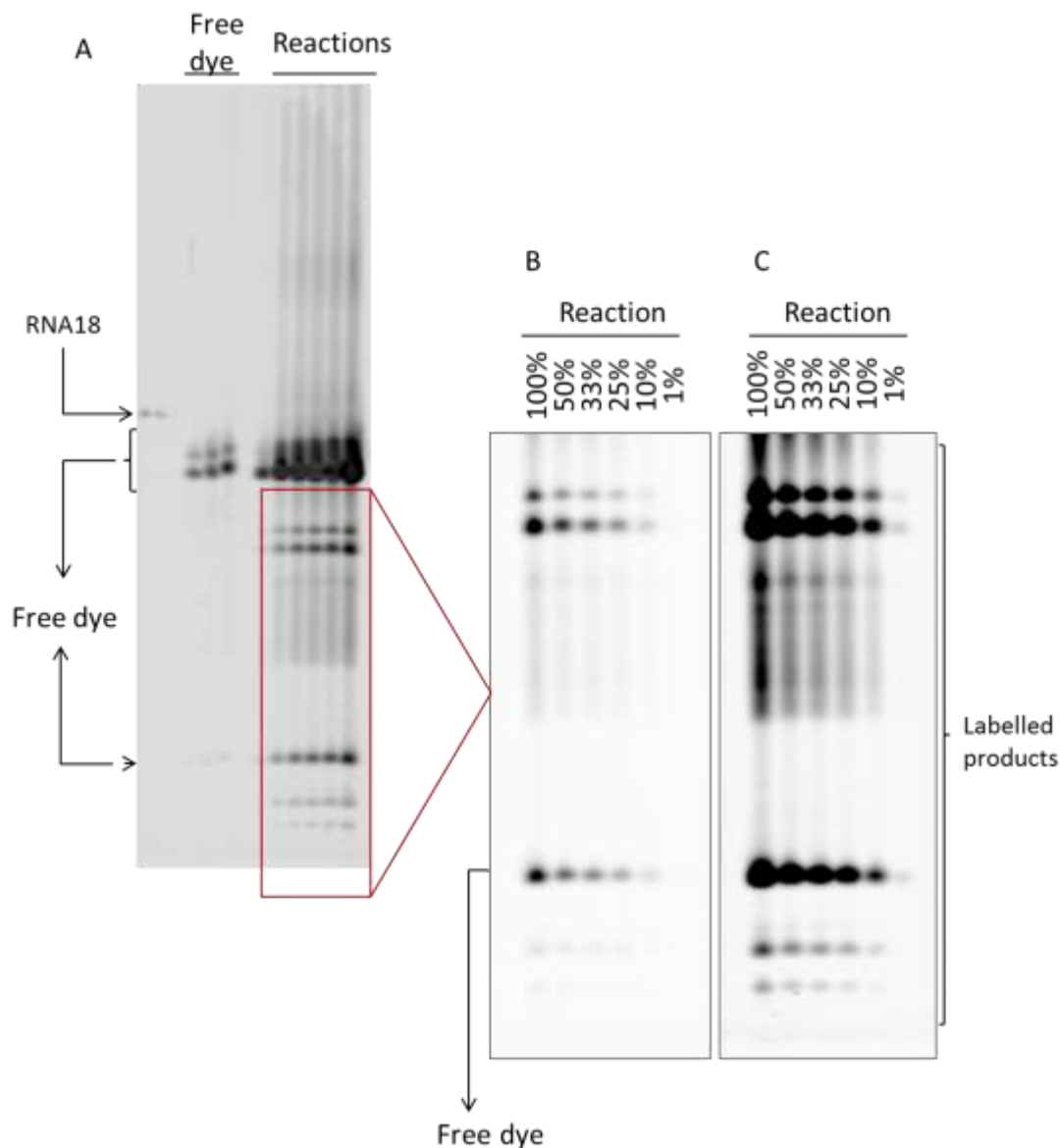


Fig. 7 – A) Electrophoresis analysis of transcription products after labelling with DBCO-Sulfo-Cy5 (15x excess); alongside the different dilutions of the reaction (100% - undiluted; 50%, 33%, 25%, 10%, 1%), three dilutions of free dye (100% - undiluted, 50%, 25% relative to the free dye sample) and two concentrations of a 18nt RNA marker (100nM and 50nM) were also electrophoresed. B) & C) Inset showing the area of the gel where the abortive products are expected to be found, at two different intensities. The image is produced by detecting Cy5 fluorescence as described in section 5.2.6.

Despite the dilutions, it is still obvious that the excess of dye is far too high and in the electrophoresis analysis, the intensity from the excess dye might be obstructing the

labelled RNA products the assay is interested in highlighting. Hence, another labelling reaction was set up, with a 1.8x dye excess; moreover, in an attempt to clean-up the excess dye post-labelling, the sample was challenged with azide-modified agarose that is able to react with the DBCO-Sulfo-Cy5 dye: before challenge, the sample appears intense blue (fig. 8A, left), but the colour fades and a dark blue precipitate readily appears upon incubation with the modified agarose (fig. 8A, right) suggesting that this route is an effective clean-up method. To test whether removal of the free dye has an equally dramatic effect when the samples are analysed by electrophoresis, a 'mini' gel employed – as this was meant as a proof of principle experiment, a smaller size gel was preferred due to its reduced running time and ease of manoeuvrability as compared to the sequencing size gels employed in the previous experiments described here. After clean-up, comparing with the lanes in which the free dye was run, it appears that some more products that could be labelled RNAs are detected (fig. 8B). As before, it is hard to assign sizes to these products or identify them precisely. Moreover, the clean-up seems to be more efficient after incubation of the sample at -20°C overnight, suggesting that incubation for 10 minutes is not sufficient to remove all the free dye present in the sample.

All the free dye used as a marker until now was recovered after protein labelling and purification by David Bauer (University of Oxford): this is DBCO-Sulfo-Cy5 that had already been reacted and is now stored in a buffer (50mM Tris-HCl pH 6.9, 5% glycerol, 0.5mM EDTA, 1mM β -mercaptoethanol) that is different than the buffer used in these labelling experiments.

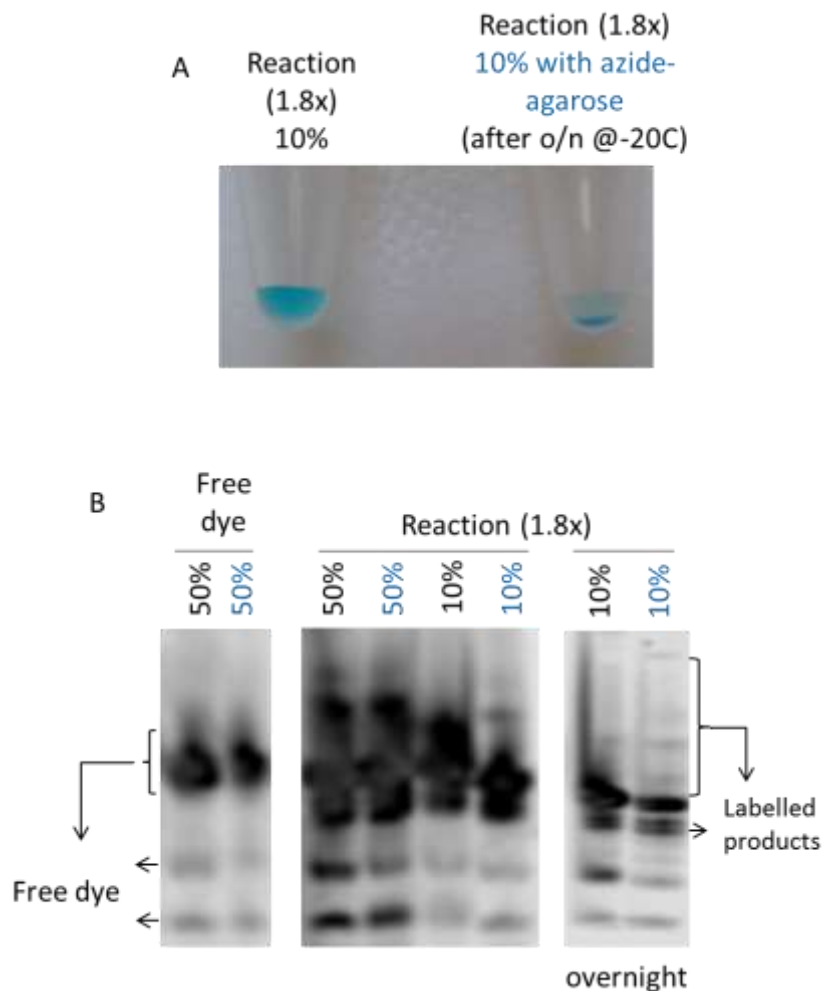


Fig. 8 – Effects of incubating the reaction, post labelling, with azide-agarose. A) A picture of a 10% dilution of the IVT sample, taken post labelling, before (left) and after (right) being challenged with azide-agarose (after addition of the azide-modified agarose, the sample was incubated at room temperature for 10 minutes, then stored overnight at -20°C). B) A comparison between free dye diluted 50% (relative to the free dye stock sample), unchallenged and challenged with azide-modified agarose (left on the gel, challenged sample in the lane labelled in blue) and two different reaction dilutions (50% and 10%; both unchallenged and challenged with azide-agarose for 10 minutes at room temperature, after which a fraction of the sample was loaded on the gel) as analysed on a 7M urea/20% 'mini' PAGE (in TBE). The remainder of the 10% sample (both challenged and unchallenged) was stored at -20°C overnight and analysed in the same way – the result is shown here for comparison (the image is a Cy5 scan).

A new IVT reaction was set up and its RNA products were reacted with a 6.6x excess of DBCO-Sulfo-Cy5 (equivalent to 50nmol) and challenged with azide-agarose (44nmol) as described in section 5.2.6; three different markers were also used: free dye (an unreacted DBCO-Sulfo-Cy5 sample diluted in TE buffer), a 25nt long 5' azide-modified RNA (RNA25), labelled using the click chemistry procedure used for the IVT RNA products (corresponding in sequence to the full length RNA transcript that can be generated from the template used in this transcription assay) and a 7nt long 5' azide-modified RNA (RNA7), labelled in the same way (corresponding to an abortive product generated in this assay and identified previously in Cordes et al. 2010).

Once the electrophoretic run was completed, the gel was imaged and multiple bands (in all the lanes) were visible under conditions suitable for detecting Cy5 fluorescence (fig. 9A). Taking the lane in which free dye was loaded as a comparison point (fig. 9A, grey boxes) labelled products were detected both in the IVT reaction lane and in the RNA25 one.

Following this initial analysis, the gel was stained with Sybr Green II, a nucleic acid stain specific for single-stranded RNA detection and which is sensitive enough to detect RNAs found in picogram amounts, after which it was imaged once more by detecting the signal from the Cy5 dye (fig. 9B – in red), from the Sybr Green II stain (fig. 9B – in green) and from the Cy5 dye excited by FRET from Sybr Green II (fig. 9B – in blue; the Sybr Green II stain is capable of homo FRET, meaning that it is able to transfer energy from one Sybr Green molecule to another, and ultimately to the Cy5 dye, whose emission can be detected). This analysis led to the following results:

- Both unlabelled (visualised with Sybr Green II) and Cy5-labelled species were detected in the RNA25 standard sample which had been allowed to react with the DBCO-Sulfo-Cy5 dye & then challenged with azide-agarose; the labelled RNA product appears shifted higher (as expected, detected by all the different scans; fig. 9B – in purple) in comparison with the unlabelled RNA product (fig. 9B – in green). In the lanes in which the RNA25 was not clicked, but either unchallenged or challenged with azide agarose, two products are also detected, suggesting that a reactive $n-1$ synthesis product is also present in the RNA sample. If the shifts between the Cy5-labelled and the unlabelled RNA, on the one hand, and the n and $n-1$ unlabelled products, on the other, are compared, it becomes immediately apparent that the labelled product will have been shifted higher, which agrees with the idea that the DBCO-Sulfo-Cy5 group, due to its size, changes the labelled RNA's mobility, slowing it down. Finally, no free dye could be detected in the lane corresponding to the labelling reaction, suggesting that the azide-agarose clean-up was very efficient.
- In the RNA7 standard sample, a signal could only be detected in the lane in which the RNA was allowed to react with the DBCO-Sulfo-Cy5. This is not surprising, as Sybr Green II is not likely to stain such short RNA strands (which is linear down to 50nt long RNAs, but variable from there on, according to the manufacturer's specifications). The signal identified corresponds to two bands that are slower than the signature band of the free dye; since in the RNA25 sample two bands could also be identified, and these two RNA standards were synthesised in the

same manner, it is likely that the signal corresponds to a labelled 7mer (standard) and a 6mer ($n-1$ synthesis product).

- In the lane corresponding to the IVT reaction, no RNA could be detected under Sybr Green II staining. Several explanations could account for this: first, it might be possible that the reaction did not occur; however, this is unlikely given my own experience with performing the assay, which has proven its robustness and reliability. Second, it might be possible that the amount of RNA products from the reaction is below the sensitivity limit of the stain (which is given by the manufacturer as 100pg of single-stranded RNA); indeed, 200ng of the standard RNA25 loaded on the same gel is easily visible under this staining regime. From previous assays (such as the one shown in fig. 4 and fig. 7), most of the IVT products generated are expected to be abortive RNAs, characterised by very short lengths, which are below the sensitivity limit of the stain; this is confirmed by the fact that no standard RNA7 could be detected with Sybr Green II (although the sample was loaded at the same concentration as RNA25 standard that was visible).
- Looking at the IVT products under Cy5 fluorescence, a signal can be detected from three different bands, of which only one corresponds to free dye (top band) as compared with the unreacted free dye sample loaded on the gel. The other two identified species exhibit mobilities which are very similar to those identified in the RNA7 standard subjected to click chemistry, bands which were assigned to a 7mer and a 6mer, respectively. In view of this and taken into account the previous

experiments described in fig. 4 and fig. 7, it is likely that these two species are indeed abortive products generated during the transcription reaction and Cy5-labelled using the click chemistry method (yet, to confirm the identity of these products, other methods will need to be employed, as discussed in the next section).

Taken together these results suggest that short RNA products, generated via an *in vitro* transcription reaction set up in the presence of an azide-modified initiating dinucleotide, can be labelled using click chemistry, with the most successful approach used to achieve this, was by performing the labelling reaction after the transcription reaction and before electrophoresis.

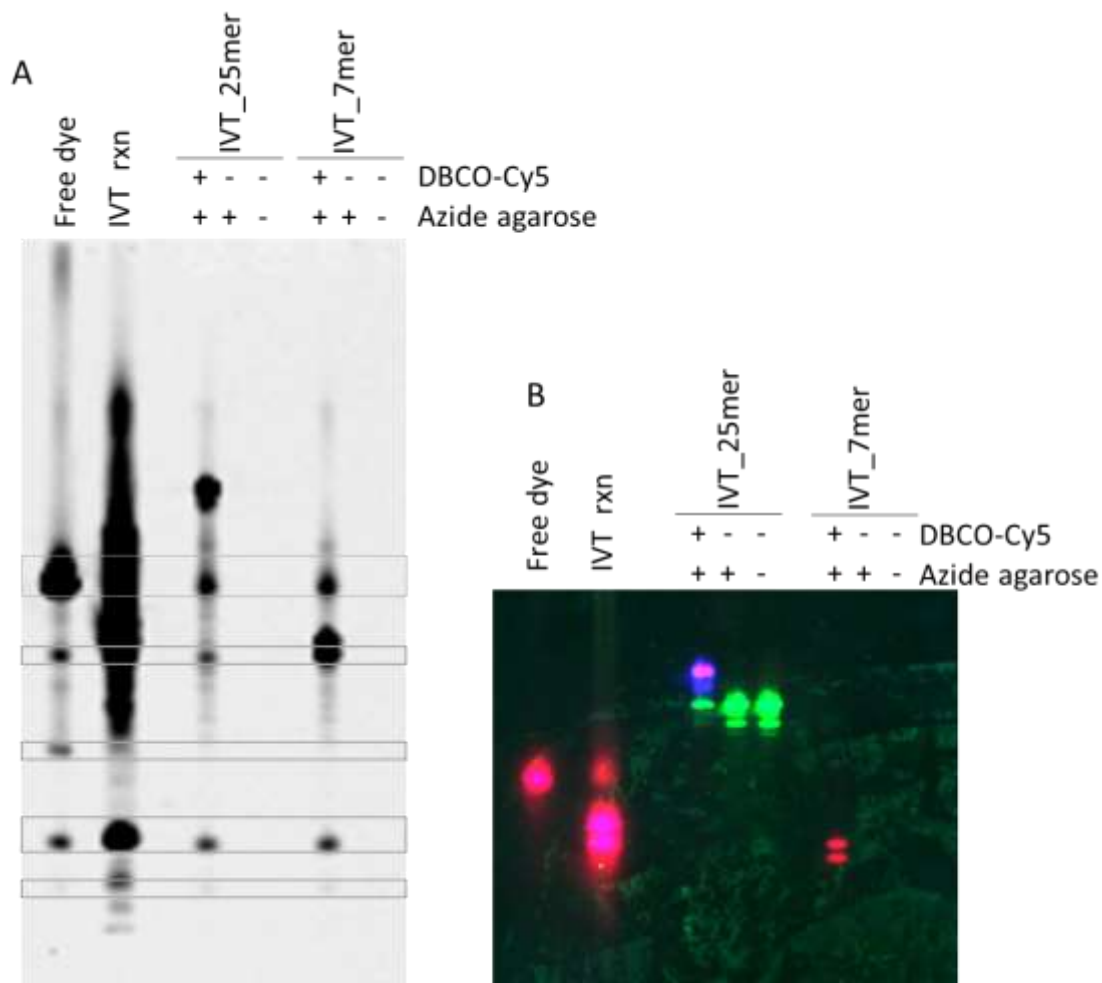


Fig. 9 – A) Labelled IVT reaction products analysed on a 7M urea/20% PAGE sequencing gel, alongside free dye and labelled RNA standards (labelled using click chemistry) of two lengths, detected via a Cy5 scan. The boxed areas underlie the bands corresponding to the free dye; the image was acquired immediately after the end of the electrophoresis run. B) A composite image consisting of three independently acquired images: Cy5 containing products are in red, RNA is shown in green (the signal is from Sybr Green II, an RNA specific stain, with two excitation wavelengths, a major one at 497nm and a minor one at 254nm); coincidence of the two signals was verified by a Sybr Green II excitation/Cy5 emission scan (shown in blue).

5.4 PERSPECTIVES

The aim of the work presented in this chapter was to devise a method that would allow very short RNAs, such as abortive transcription products, ranging between 2nt and 11nt, synthesised using an azide-modified initiating dinucleotide, to be labelled with a fluorescent dye via a copper-free azide-alkyne cycloaddition reaction. To the best of my knowledge, this is a novel approach that builds on previous reports that showed that a modified dinucleotide can be used to generate RNAs amenable to click through an *in vitro* transcription reaction (Samanta et al. 2013), that short RNAs can be labelled using click chemistry (albeit copper-catalysed cycloaddition; Paredes & Das 2011) and that micromolar concentrations of short RNAs (22-30nt) can be labelled using strain-promoted azide-alkyne cycloaddition (Winz et al. 2012).

To achieve this aim, I tested three different approaches: labelling post-synthesis, but pre-electrophoresis, which seems to be the most promising strategy; labelling during electrophoresis and post-electrophoresis.

Performing the click reaction post-electrophoresis (very similar to staining) seems to be the least optimal solution, as it would not only require extensive gel manipulation, but it would also create conditions favourable for short RNA diffusion in the gel, making it harder to detect the desired products. Allowing the click reaction to occur during electrophoresis might be possible, however, the conditions for this need to be further optimised.

The same can be said about labelling the reaction post-synthesis and before separating the products on a denaturing PAGE: further tweaking the excess of dye used in the

labelling reaction or possibly performing several clean-up steps might lead to better results. It is unfortunate that due to its short length, the RNA of interest cannot be easily separated from the free dye using ethanol or isopropanol precipitation or by using a reverse-phase chromatographic step (such as a C18 Sep Pak column, as used by Paredes & Das 2011), however I believe the azide-agarose clean-up step to be a suitable alternative.

While the use of the un-sulfonated BCN dye was considered a problem during the initial stages of this project, it might actually prove to be an interesting opportunity to explore, as a positively-charged free dye could prove advantageous during electrophoresis when rather than entering the gel, it would sweep out of the well, leaving the RNAs of interest to enter the gel's matrix.

A good strategy to confirm the identity of the labelled RNAs detected would be to excise them from the gel and investigate them further via mass spectrometry. Alternatively, setting up a transcription reaction in which the 5' azide-modified ApA initiating dinucleotide is used alongside radiolabelled GTP, would allow for identification of the RNA products via both the radioactive and the fluorescent tag; comparing the resulting transcription profile with the one obtained by using either the modified ApA or the radiolabelled GTP would yield useful information as to the identity (length) of the IVT products and their relative mobilities.

To conclude, the experiments presented in this chapter suggest that copper-free click chemistry labelling of short RNAs generated via *in vitro* transcription is a viable approach providing that the experimental conditions are further optimised.

5.5 CONTRIBUTIONS

The RNA7 and RNA25 standards were synthesized by Dr Afaf El-Sagheer (University of Oxford), who also provided me with the BCN-Cy3 dye.

5.6 BIBLIOGRAPHY

Agard, N.J., Prescher, J.A. & Bertozzi, C.R., 2004. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society*, 126(46), pp.15046–7.

Best, M.D., 2009. Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry*, 48(28), pp.6571–84.

Boulikas, T. & Hancock, R., 1981. A highly sensitive technique for staining DNA and RNA in polycrylamide gels using silver. *Journal of Biochemical and Biophysical Methods*, 5(4), pp.219–228.

Callaci, S. & Heyduk, T., 1998. Conformation and DNA binding properties of a single-stranded DNA binding region of sigma 70 subunit from Escherichia coli RNA polymerase are modulated by an interaction with the core enzyme. *Biochemistry*, 37(10), pp.3312–20.

Cordes, T. et al., 2010. Sensing DNA opening in transcription using quencher Förster resonance energy transfer. *Biochemistry*, 49(43), pp.9171–80.

Dojahn, C.M., Hesse, M. & Arenz, C., 2013. A chemo-enzymatic approach to specifically click-modified RNA. *Chemical communications (Cambridge, England)*, 49(30), pp.3128–30.

El-Sagheer, A.H. & Brown, T., 2010. New strategy for the synthesis of chemically modified RNA constructs exemplified by hairpin and hammerhead ribozymes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(35), pp.15329–34.

England, T.E. & Uhlenbeck, O.C., 1978. 3'-Terminal labelling of RNA with T4 RNA ligase. *Nature*, 275(5680), pp.560–561.

Hsu, L.M., 2009. Monitoring abortive initiation. *Methods (San Diego, Calif.)*, 47(1), pp.25–36.

Hsu, L.M., 2002. Promoter clearance and escape in prokaryotes. *Biochimica et biophysica acta*, 1577(2), pp.191–207.

Jao, C.Y. & Salic, A., 2008. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), pp.15779–84.

- Kapanidis, A.N. et al., 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science (New York, N.Y.)*, 314(5802), pp.1144–7.
- Kapanidis, A.N. et al., 2005. Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. *Molecular cell*, 20(3), pp.347–56.
- Kinoshita, Y., Nishigaki, K. & Husimi, Y., 1997. Fluorescence-, isotope- or biotin-labeling of the 5' -end of single-stranded DNA/RNA using T4 RNA ligase. *Nucleic acids research*, 25(18), pp.3747–8.
- Köhn, M. & Breinbauer, R., 2004. The Staudinger ligation-a gift to chemical biology. *Angewandte Chemie (International ed. in English)*, 43(24), pp.3106–16.
- Kurschat, W.C. et al., 2005. Optimizing splinted ligation of highly structured small RNAs. *RNA (New York, N.Y.)*, 11(12), pp.1909–14.
- Kwak, J.E. & Wickens, M., 2007. A family of poly(U) polymerases. *RNA (New York, N.Y.)*, 13(6), pp.860–7.
- Leibman, M. & Hochschild, A., 2007. A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *The EMBO journal*, 26(6), pp.1579–90.
- Li, N., Yu, C. & Huang, F., 2005. Novel cyanine-AMP conjugates for efficient 5' RNA fluorescent labeling by one-step transcription and replacement of [gamma-32P]ATP in RNA structural investigation. *Nucleic acids research*, 33(4), p.e37.
- Lingner, J. & Keller, W., 1993. 3'-end labeling of RNA with recombinant yeast poly(A) polymerase. *Nucleic Acids Research*, 21(12), pp.2917–2920.
- Martin, G. & Keller, W., 1998. Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA (New York, N.Y.)*, 4(2), pp.226–30.
- Mitchell, L.G., Bodenteich, A. & Merril, C.R., 1996. Use of silver staining to detect nucleic acids. *Methods in molecular biology (Clifton, N.J.)*, 58, pp.97–103.
- Murakami, K.S. & Darst, S.A., 2003. Bacterial RNA polymerases: the whole story. *Current Opinion in Structural Biology*, 13(1), pp.31–39.
- Murakami, K.S., Masuda, S. & Darst, S.A., 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science (New York, N.Y.)*, 296(5571), pp.1280–4.
- Paredes, E. & Das, S.R., 2011. Click chemistry for rapid labeling and ligation of RNA. *ChemBiochem : a European journal of chemical biology*, 12(1), pp.125–31.
- Phelps, K., Morris, A. & Beal, P.A., 2012. Novel modifications in RNA. *ACS chemical biology*, 7(1), pp.100–9.

- Pribnow, D., 1975. Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. *Journal of Molecular Biology*, 99(3), pp.419–443.
- Raffaella, M. et al., 2005. Holoenzyme Switching and Stochastic Release of Sigma Factors from RNA Polymerase In Vivo. *Molecular Cell*, 20(3), pp.357–366.
- Rao, H., Sawant, A.A., et al., 2012. Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chemical communications (Cambridge, England)*, 48(4), pp.498–500.
- Rao, H., Sawant, A. a, et al., 2012. Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chemical communications (Cambridge, England)*, 48(4), pp.498–500.
- Revyakin, A. et al., 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science (New York, N.Y.)*, 314(5802), pp.1139–43.
- Robb, N.C. et al., 2013. The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *Journal of molecular biology*, 425(5), pp.875–85.
- Rosemeyer, V., Laubrock, A. & Seibl, R., 1995. Nonradioactive 3'-End-Labeling of RNA Molecules of Different Lengths by Terminal Deoxynucleotidyltransferase. *Analytical Biochemistry*, 224(1), pp.446–449.
- Rostovtsev, V. V et al., 2002. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angewandte Chemie (International ed. in English)*, 41(14), pp.2596–9.
- Samanta, A., Krause, A. & Jäschke, A., 2013. A modified dinucleotide for site-specific RNA-labelling by transcription priming and click chemistry. *Chemical communications (Cambridge, England)*.
- Saxon, E. & Bertozzi, C.R., 2000. Cell Surface Engineering by a Modified Staudinger Reaction. *Science*, 287(5460), pp.2007–2010.
- Stark, M.R. et al., 2006. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA (New York, N.Y.)*, 12(11), pp.2014–9.
- Tornøe, C.W., Christensen, C. & Meldal, M., 2002. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *The Journal of Organic Chemistry*, 67(9), pp.3057–3064.
- Wickens, M. & Kwak, J.E., 2008. Molecular biology. A tail tale for U. *Science (New York, N.Y.)*, 319(5868), pp.1344–5.

Winz, M.-L. et al., 2012. Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic acids research*, 40(10), p.e78.

Young, B.A., Gruber, T.M. & Gross, C.A., 2002. Views of Transcription Initiation. *Cell*, 109(4), pp.417–420.

Zhao, B. et al., 2012. Simple and sensitive microRNA labeling by terminal deoxynucleotidyl transferase. *Acta biochimica et biophysica Sinica*, 44(2), pp.129–35.

Appendix: List of presentations and publications

Publications:

2014

Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter (A. I. Tomescu*, N. C. Robb*, N. Hengrung, E. Fodor and A. N. Kapanidis; *equal contributors) – *PNAS*, 111(32): p. E3335-42

2010

Sensing DNA opening in transcription using quenchable Förster resonance energy transfer (T. Cordes, Y. Santoso, A.I. Tomescu, K. Gryte, L.C. Hwang, B. Camará, S. Wigneshweraraj and A.N. Kapanidis) – *Biochemistry*, 49(43): p. 9171-80

Presentations:

2013

Biophysical Society 57th Annual Meeting (Philadelphia, USA) and **25th UK RNA Polymerase Workshop** (London, UK) – poster:

Characterizing influenza A RNA polymerase - promoter interaction using ensemble fluorescence spectroscopy (A.I. Tomescu*, N.C. Robb, N. Hengrung, E. Fodor and A.N. Kapanidis)

Department of Physics Graduate Students Seminar (Oxford, UK) – invited talk: *Biosensing and pathogen detection: spotlight on influenza*

Department of Physics Annual Students Seminar (Oxford, UK) – talk: *Investigating influenza A RNA polymerase - promoter binding using fluorescence techniques*

2012

3rd International Influenza Meeting (Muenster, Germany) – poster: *Promoter binding studies of the influenza A RNA*

polymerase (N.C. Robb*, A.I. Tomescu*, N. Hengrung, E. Fodor and A.N. Kapanidis, *equal contributors)

Department of Physics Annual Students Seminar (Oxford, UK) – poster: *Binding studies of the influenza RNA polymerase* (A.I. Tomescu, N.C. Robb, K. Gryte, N. Hengrung, E. Fodor and A.N. Kapanidis)

2010

22nd UK RNA Polymerase Workshop (Nottingham, UK) – talk: *Real time, single-molecule abortive transcription of RNAs*

69th Harden Conference: Structure, function and evolution of RNA Polymerases (Cambridge, UK) – poster: *Sensing DNA opening using single-molecule fluorescence quenching* (T. Cordes, Y. Santoso, A. Tomescu, K. Gryte, L. Chin Hwang, B. Camará, S. Wigneshweraraj and A.N. Kapanidis)