Biolayer Interferometry: Protein-RNA Interactions

Stephen R. Martin[1], Andres Ramos[2], and Laura Masino[1]

e-mail: Laura.Masino@crick.ac.uk

[1]Structural Biology Science Technology Platform

The Francis Crick Institute

1 Midland Road, London NW1 1AT, U.K.

[2]Department of Structural & Molecular Biology

Darwin Building, Gower St

London WC1E 6BT, U.K.

Running Title: BLI: Protein-RNA interactions

Biolayer Interferometry: Protein-RNA Interactions

Abstract

RNA-binding proteins often contain multiple RNA-binding domains connected by short flexible linkers. This domain arrangement allows the protein to bind the RNA with greater affinity and specificity than would be possible with individual domains and sometimes to re-model its structure. It is therefore important to understand how multiple modules interact with RNA, because it is the modular nature of these proteins which specifies their biological function. This chapter is concerned with the use of biolayer interferometry to study protein-RNA interactions.

Key words: Biolayer interferometry, RNA, kinetics

1. Introduction

1.1 Protein-RNA interactions

Post-transcriptional gene regulation consists of a ubiquitous and essential network of protein-RNA based cellular processes that expands genomic diversity, and it is essential in the development and function of complex organisms. Not surprisingly, misfunction of the different RNA regulation steps has been associated with a range of pathologies, including different cancers, neurodevelopmental and neurodegenerative diseases, immuno-pathologies and viral infection [**1,2**].

RNA regulation is mediated by between 1000 and 2000 RNA binding proteins, a few hundred of which have been validated functionally [**3**]. In contrast, a human cell typically contains ten to 20000 different mRNAs [**4**], each binding many different RNA binding proteins. Most RNA binding proteins recognize large and diverse sets of RNA targets,

including both mRNAs and non-coding RNAs. An accurate recognition of these targets is essential to define the set of genes regulated by a protein, but also to define gene expression programs in different cellular locations at specific times [**1,2**]. Understanding, at a mechanistic level, how RNA binding proteins recognize the RNA targets is one of the main challenges in gene regulation.

Despite the large number of RNA binding proteins, recognition of the RNA targets is mediated by a relatively small number of different RNA binding domains, which are present in multiple copies of the same or different domains within one RNA binding protein [**5**]. In these proteins, RNA binding is mediated by a combinatorial action of more than one RNA binding domain. In this context, diversity of recognition stems from both the specificity of a domain in a given protein and from a range of different inter-domain coupling modes. Inter-domain coupling can increase affinity and specificity, re-shape the RNA structure and provide new opportunities for regulation [**6**].

While a global survey of RNA binding domains shows different sizes and RNA recognition properties, the most common of these domains are less than 100 amino acids in size. For many of these domains a structural characterization of a 'canonical' binding mode is available, together with, in some cases, the description of a few of the structural and RNA-binding variations on this mode [**5,6**]. However, information on the kinetics of binding and inter-domain coupling is available only in a small number of systems. This information is essential to model the binding of these domains in the cellular environment.

As an example of the RNA binding domains discussed above, the KH domain is a small (~70 amino acids) αβ fold found in a number of RNA regulatory proteins important in development, function and disease. The domain binds to single stranded nucleic acids with a varying degree of affinity and specificity. RNA binding is mediated by the interaction of the nucleic acid backbone with a negatively charged GxxG loop [**7**]. The details of this

interaction are different in different domains, but binding of the loop orients the nucleobases towards a hydrophobic groove in the protein, for sequence specific recognition [**7**]. In addition, individual KH domains can interact using a variety of surfaces. Importantly, despite the importance of KH-containing proteins in human health, kinetic and mechanistic information on multi-domain binding has only recently started to become available. This is partly due to the difficulties in obtaining high quality data on the kinetics of protein-RNA interactions at different affinities in the same experimental system.

1.2 Biolayer interferometry

Biolayer interferometry (BLI) is a label-free method that enables real-time analysis of biomolecular interactions occurring on 8 or 16 biosensors in 96- or 384 well plates [**8,9,10**]. White light travels down the biosensors and is reflected back to spectrometers from two places: an internal reference layer and the interface between the solvent and molecules immobilized on the sensor tip. This results in an interference pattern and the instrument measures the maximum wavelength of the pattern.

The interaction of a binding partner with a molecule immobilized on the biosensor tip gives an increase in the distance between the internal reference layer and material attached to the biosensor. This results in a wavelength shift in the maximum of the interference pattern which is monitored in real-time.

BLI can be used to analyze binding interactions of small molecules, proteins, antibodies, nucleic acids, viruses or whole cells. It can determine specificity, binding kinetics and affinity, and perform quantitation assays. The tips of the biosensors are derivatised with a range of different surface chemistries and can be used to analyze macromolecules with different tags (e.g. His-tag, GST-tag, biotin). In the case of protein-nucleic acid interactions,

biotinylated nucleic acids can be immobilized on streptavidin sensors and binding of the protein partner can be recorded. One significant advantage of BLI is that only molecules binding to or dissociating from the biosensor will change the interference pattern and generate an instrument response. Unbound molecules and changes in the refractive index of the solvent have no effect on the interference pattern. A further advantage is that the measurement is non-destructive and samples are recoverable.

1.3 Kinetic Theory

Although all the Octet instruments come with built-in software for curve analysis, it is of course advisable to fully understand the kinetic theory that underpins the technique. In addition, if in-house software for kinetic analysis is available, then complex instrument response curves can be downloaded and analysed using more sophisticated approaches than those available with the instrument.

In the simplest case the kinetic analysis of the biosensor data is based on the idea that the interaction between the soluble protein reactant (P) and an immobilized nucleic acid (N) may be described by the following scheme:

$$P + N \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PN$$

where k_{on} and k_{off} are the association and dissociation rate constants (units $M^{-1}s^{-1}$ and s^{-1} respectively). Under conditions where the extent of the reaction is governed by reaction kinetics rather than mass transport considerations, the differential equation for such a system is:

$$\frac{d[PN]}{dt} = k_{on}[P][N] - k_{off}[PN] \qquad (1)$$

Substituting [N] = [N_0] – [PN] (where [N_0] is the total (unknown) concentration of binding sites on the sensor) gives:

$$\frac{d[PN]}{dt} = k_{on}[P]([N_o] - [PN]) - k_{off}[PN] \quad (2)$$

In most cases the protein concentration will remain at its initial value ([P_0]) throughout the reaction because the total concentration of binding sites on the sensor is vanishingly small compared with the protein concentration in the well. Therefore:

$$\frac{d[PN]}{dt} = k_{on}[P_o]([N_o] - [PN]) - k_{off}[PN] \quad (3)$$

If there is no non-specific binding of the protein to the sensor then the instrument response must be directly proportional to [PN] and this equation may therefore be rewritten as:

$$\frac{dR}{dt} = k_{on}[P_o](R_{max} - R) - k_{off}R \quad (4)$$

where R denotes the response at time t and R_{max} is the maximal response that would be obtained if all available binding sites on the sensor were saturated (i.e., when [PN] = [N_0]). Integration of this equation gives:

$$R = \frac{k_{on}[P_o]R_{max}\{1-e^{-t(k_{on}[P_o]+k_{off})}\}}{k_{on}[P_o]+k_{off}} \quad (5)$$

Assuming that the maximum possible response (R_{max}) and the equilibrium response at the end of the association phase (R_{eq}) must be proportional to [N_0] and [PN] respectively, one may write:

$$\frac{k_{on}}{k_{off}} = \frac{[PN]}{[P][N]} = \frac{[PN]}{[P_o]([N_o]-[PN])} = \frac{R_{eq}}{[P_o](R_{max}-R_{eq})} \quad (6)$$

$$\text{and } R_{eq} = \frac{k_{on}[P_o]R_{max}}{k_{on}[P_o]+k_{off}} \quad (7)$$

Substitution in **Equation (5)** then gives

$$R = R_{eq}\left\{1 - e^{-t(k_{on}[P_o]+k_{off})}\right\} \quad (8)$$

and further substituting $k_{obs} = k_{on}[P_o] + k_{off}$ gives:

$$R = R_{eq}\{1 - e^{-tk_{obs}}\} \quad (9)$$

The time dependence of the biosensor response in the association phase is then expressed in terms of a pseudo-first-order rate constant k_{obs}, and R_{eq}, the response at equilibrium. Values for k_{on} and k_{off} can then, in favorable cases, be obtained as the slope and y-axis intercept of a plot of k_{obs} *versus* $[P_0]$ and the equilibrium dissociation constant (K_d) can be calculated as k_{off}/k_{on}.

A value for the K_d can also be obtained from the variation of the R_{eq} value with protein concentration. Dividing **Equation (7)** by k_{on} and substituting $K_d = k_{off}/k_{on}$ gives

$$R_{eq} = \frac{[P_o]R_{max}}{[P_o]+K_d} \qquad (10)$$

Non-linear regression analysis of R_{eq} values obtained at a series of protein concentrations should therefore yield estimates of K_d and R_{max}.

A much simpler expression applies to the protein dissociation that results when buffer is substituted for the protein solution as the liquid covering the biosensors (i.e. $[P_o] = 0$ in **Equation (4)**):

$$R = R_o e^{-tk_{off}} \qquad (11)$$

where R_o is the biosensor response prior to the start of the dissociation. R_o will not always be equal to R_{eq} as the association curves, particularly those recorded at low added protein concentrations, will not necessarily have reached equilibrium at the end of the association phase. Analysis of the dissociation phase can therefore, in favorable cases, give an additional independent measure of the dissociation rate constant k_{off}.

In relatively rare cases, it is possible to extract a self-consistent set of kinetic and thermodynamic parameters using the three approaches outlined above in **Equations (9), (10),** and **(11)**. It is more often the case that not all relevant parameters can be determined. For

example, low affinity interactions generally have high k_{off} values and their study necessarily requires the use of high protein concentrations. In such cases the association and dissociation phases are likely to be very fast and because the instrument only records data every 0.2 seconds it will not be possible to extract rate constants using **Equations (9)** or **(11)**. In favorable cases it may still be possible to determine a value for the K_d using **Equation (10)** as R_{eq} values can generally be obtained from the 'top-hat' instrument response curves that are observed when the association rate equals the dissociation rate and the overall response is flat. In the case of high affinity interactions, the dissociation rate is likely to be very slow and it is often not possible to determine a value for k_{off} using **Equation (11)** or from the intercept of a plot of k_{obs} *vs* $[P_0]$. The determination of a K_d for a high affinity interaction necessarily requires the use of very low protein concentrations and this can be problematic because the reaction will take a long time to reach equilibrium and extracting reliable R_{eq} values using **Equation (9)** may be difficult.

2. Materials

2.1 Instrumentation

Instruments are available from Molecular Devices (https://www.moleculardevices.com/). The Octet RED96 system that we use is an 8-channel instrument that is ideally suited for the characterization of protein-nucleic acid interactions (*see* **Note 1**).

2.2 Consumables

Two 96-well microplates are required for every assay with the Octet RED96: a sample plate for the experiment and a plate for pre-hydrating the sensors. The sample volume for the plates

is 200 µl per well and for the Octet RED96 the plates must be Greiner catalogue number 655209.

As noted above a wide range of biosensors is available from Molecular Devices. For most of our work we have used streptavidin coated biosensors (SA: Catalogue number 18-5021) to capture 5'biotinylated oligonucleotides.

2.3 Reagents

The manufacturer's recommended buffer for routine measurements is PBS (or HBS) containing 0.1 mg/ml BSA and 0.002% Tween-20. The buffer should be the same for all samples and in all experimental steps (*see* **Note 2**).

Biotinylated oligonucleotides can be purchased from various suppliers (e.g. Dharmacon and Integrated DNA Technologies) (*see* **Note 3**).

All protein samples should be of the highest possible purity and concentrations need to be accurately determined.

3. Methods

3.1 Standard Binding experiment

This section describes the method that is most often used in the determination of equilibrium dissociation constants and kinetic constants for the interaction of proteins with immobilized oligonucleotides.

1. Soak the biosensors in Experimental buffer using a sensor rack and a 96-well microplate. The biosensors should be pre-hydrated for at least 20 minutes before initiating the measurement.

2. Fill three columns in the sample microplate with Experimental buffer.

3. Fill a column in the sample microplate with the biotinylated oligonucleotide at the same concentration in each well (typically around 0.5 µg/mL) (*see* **Note 4**).

4. Fill a column in the sample microplate with protein at different concentrations. Ideally, the concentration range should be from 0.1 x K_d to 10 x K_d. In order to give adequate coverage of this range, it is frequently necessary to repeat the measurement with a different set of protein concentrations.

5. The typical standard binding experiment involves 5 steps in which sensors must be dipped into the different columns of the 96-well plate (*see* **Figure 1**). Program the computer to perform the following steps:

1) Baseline 1: The sensors are dipped in the first column of buffer for equilibration. The length of this step is typically 200 seconds.

2) Loading: The sensors are dipped in the column containing the biotinylated oligonucleotide. In experiments designed to determine a K_d using **Equation (10)** it is important that all sensors should give the same response in this step (*see* **Note 5**).

3) Baseline 2: The sensors are dipped in the second column of experimental buffer to remove any unbound RNA. These may be the same wells as those used in step 1) but we would advise using a separate column of wells (*see* **Note 6**).

4) Association: The sensors are dipped in the column containing the protein (*see* **Note 7**).

5) Dissociation: The sensors are dipped in the third column of experimental buffer. It is advisable to use a separate column of buffer and not the same as those used in steps 1) and 3) in order to avoid any cross contamination (*see* **Note 8**).

At the end of the experiment, samples in the 96-well plate can be recovered or reused for subsequent experiments, provided the reagents have not deteriorated and little or no evaporation has occurred, as this would lead to a concentration increase. Although we routinely discard the sensors at the end of the experiment, they can be used for additional cycles of measurements provided that the bound molecule completely dissociates in step 5). In the case of incomplete dissociation, the sensors can sometimes be regenerated and used in subsequent experiments (*see* **Note 9**).

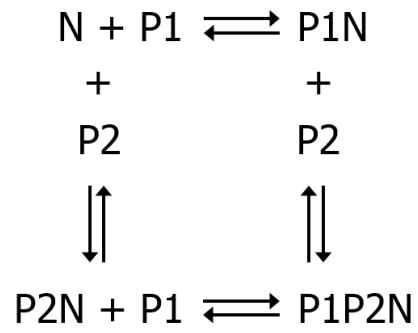
Two common problems that may be encountered are non-specific binding of the protein to the sensor surface and instrument drift, which may be a problem in those experiments which are designed to run for a long time. In initial experiments we routinely include two controls. We check for non-specific binding of the protein to the sensor surface using steps 1)-5) above but using a reference biosensor with no oligonucleotide loaded in step 2).

Instrument drift problems can be identified by using a reference biosensor with no immobilized oligonucleotide in step 2) and no protein present in step 4).

The inclusion of BSA and Tween-20 will generally reduce non-specific binding but will not always eliminate it completely (*see* **Notes 10 &11)**. In principle a small amount of non-specific binding can be corrected for by including reference sensors where no oligonucleotide is bound in step 2) but, because the amount of non-specific binding will depend on the protein concentration, this requires a separate reference sensor for each protein concentration used.

3.2 Ternary Complexes

More complex experimental designs are, of course, possible and are often very informative. For example, it is possible to study the formation of ternary complexes where two molecules can bind to different sites on an immobilized molecule [**11**]. This would be described by the following scheme:



where N is the immobilized nucleic acid and P_1 and P_2 are the first and second binding partners. In this case, the protocol would be:

1. Use the method described in section 3.1 to assess the binding of the individual proteins to form the binary complexes P_1N and P_2N.
2. Set up the 96-well microplate for the ternary complex binding experiment by filling the following columns:

Columns 1-3: Experimental buffer

Column 4: Biotinylated oligonucleotide

Columns 5-6: First binding partner P_1 at a fixed saturating concentration (at least 20 times the K_d for formation of the binary complex P_1N).

Column 7: First binding partner P_1 at the same fixed concentration and varying concentrations of the second partner P_2 (ideally in the range 0.1 x K_d to 10 x K_d for P_2 binding to P_1N, if known).

3. Program the computer to perform the following steps:

1-3) Baseline 1, Loading, Baseline 2: as in section 3.1

4) Association 1: The sensors are dipped in column 5, containing P_1. In this step the binary complex P_1N is formed.

5) Association 2: The sensors are dipped in column 7, containing P_1 and varying concentrations of P_2. In this step the ternary complex P_1P_2N is formed.

6) Dissociation 1: The sensors are dipped in column 6, containing P_1 but no P2. In this step, the dissociation of P_2 is measured.

7) Dissociation 2: The sensors are dipped in buffer column 3, to measure the dissociation of P_1.

3.3 Competition

It is also possible to perform various different types of competition experiments. If two protein molecules (say P_1 and P_2) compete for the same site on an immobilized molecule, then it is possible to study species P_2 displacing bound species P_1 (or *vice versa*), providing that the two species give a significantly different response when bound (*see* **Note 12**). This would be described by the following scheme:

$$N + P1 \rightleftharpoons P1N$$

$$N + P2 \rightleftharpoons P2N$$

where N is the immobilized nucleic acid and P_1 and P_2 are the first and second binding partners. In this case, the protocol would be:

1. Use the method described in section 3.1 to assess the binding of the individual proteins to form the binary complexes P_1N and P_2N.

2. Set up the 96-well microplate for the experiment by filling the following columns:

Columns 1-3: Experimental buffer

Column 4: Biotinylated oligonucleotide

Columns 5-6: First binding partner P_1 at a fixed saturating concentration (at least 20 times the K_d for formation of the binary complex P_1N, or higher)

Column 7: First binding partner P_1 at the same fixed concentration and varying concentrations of the second partner P_2, ideally in the range 0.1 x K_d to 10 x K_d for P_2 binding to N, if known. This concentration range might need to be changed if P_1 and P_2 bind with very different affinities.

3. Program the experiment to perform the following steps:

1-3) Baseline 1, Loading, Baseline 2: as in section 3.1

4) Association 1: The sensors are dipped in column 5, containing P_1. In this step the binary complex P_1N is formed.

5) Association 2: The sensors are dipped in column 7, containing P_1 and P_2. In this step some of P_1 will dissociate and some of P_2 will bind.

6) Dissociation 1: The sensors are dipped in column 6, containing only P_1. In this step, P_2 should dissociate and P_1 should rebind.

7) Dissociation 2 (optional): The sensors are dipped in buffer column 3, to measure the dissociation of P_1.

3.4 Data analysis

In recent years we have employed the analytical approach described in **Section 1.3** to determine kinetic (k_{on} and k_{off}) and thermodynamic (K_d) parameters for the interaction of different proteins with their DNA and RNA target sequences [**11,12,13,14,15,16**], and also showed that it is possible, starting from those data, to build a kinetic model for the interaction that provides information on mRNA regulation and RNA re-modelling [**15,16**].
IMP1/ZBP1 is a multi-functional RNA binding protein that regulates mRNA metabolism, transport and translation during development, and in cancer [**17**]. It contains six putative RNA-binding domains (two RRM and four KH) organized in three two-domain units. Interestingly, the binding of different RNA targets is mediated by the two KH di-domains, KH1KH2 and KH3KH4, in a target-dependent fashion. For example, binding of the c-Myc oncogene mRNA in highly proliferating cells requires the KH1KH2 di-domain [**18**], while the interaction of IMP1 with the β-actin mRNA in neurons requires only the KH3KH4 di-domain [**19**].

In a recent study we examined the interaction of KH1KH2 with an oligo recapitulating an IMP1 binding site (CACAGCAUACAUCCUGUCCGUC), which we named MYCRNA [**16**]. An important tool to dissect this interaction has been a KH domain mutant where nucleic acid binding is eliminated by the mutation of the two variable amino acids in the hallmark GxxG loop to Aspartate (GxxG-to-GDDG) [**12,13**]. This mutation does not affect the structure or the stability of the domain and allows one to examine RNA binding of the individual KH domains within an intact IMP1 KH di-domain structural context. We have used three protein constructs: wild-type KH1KH2, KH1KH2(DD) (the KH2 KO), and

KH1(DD)KH2 (the KH1 KO). BLI experiments using immobilized MYCRNA exposed to different concentrations of IMP1 allowed us to obtain the equilibrium dissociation constants (K_d) as well as the kinetic parameters for the interactions of two of these constructs. Typical experimental data for the wild-type and the KH1KH2(DD) construct are shown in **Figure 2** along with the kinetic and thermodynamic parameters for the interactions.

In a different study, we again used BLI to investigate the interaction of β-actin mRNA with a KH3KH4 di-domain from the chicken orthologue of IMP1, Zipcode binding protein 1 (ZBP1) [**15**]. |The IMP1 protein is conserved from Drosophila to human, in particular within the KH domains [**17**]. ZBP1 has the same RNA binding properties as the human protein and is often used as a proxy to study the IMP1-RNA interaction *in vitro*. As discussed above, the interaction with the β-actin mRNA is mediated by the KH3KH4 di-domain [**19**], which recognizes the 28 nucleotide β-actin 3' UTR Zipcode RNA element (ACCGGACUGUUACCAACACCCACACCCC) (*see* **Figure 3)**. In order to study the KH3KH4 interaction, in this study we used wild type protein plus two GxxG-to-GDDG ZBP1constructs, KH3KH4(DD) (the KH4 KO) and KH3(DD)KH4 (the KH3 KO). This is similar to what was discussed above for the KH1KH2-RNA interaction. The equilibrium interaction was studied by using immobilized 28-nucleotide Zipcode RNA exposed to different concentrations of ZBP1 KH3(DD)KH4 and KH3KH4(DD). The equilibrium dissociation constants for the Zipcode RNA:KH3KH4(DD) and RNA:KH3(DD)KH4 complexes were found to be ~1.5 μM and ~0.9 μM, respectively. Although the affinities of the two domains are similar, the kinetic constants are somewhat different. The association rate constant for KH3(DD)KH4 (1.4×10^5 $M^{-1}s^{-1}$) is five times faster than for KH3KH4(DD) (3.0×10^4 $M^{-1}s^{-1}$). Conversely, the dissociation rate constant for the KH3KH4(DD):RNA complex (0.046 s^{-1}) is three times slower than that for the KH3(DD)KH4 complex (0.13 s^{-1}).

The wild-type construct (KH3KH4) in which both domains can engage in the interaction binds to the RNA with an association rate constant (1.6 x 10^5 $M^{-1}s^{-1}$) that is similar to that for KH4 but the dissociation rate constant is very much smaller (0.0033 s^{-1}). The ratio of these constants gives a K_d of ~20 nM, indicating that the coupling of KH3 and KH4 binding is relatively weak, increasing the affinity of the individual interactions by only a factor of ~50 (*see* **Note 13**).

On the basis of the experiments described above we proposed a model (**Figure 4**) in which either domain of KH3KH4 can associate with its cognate sequence on the Zipcode to form a 1:1 complex. Each of the two possible complexes formed in this way can then proceed through a ‘‘ring-closure’’ step, in which the remaining unbound domain binds to its cognate RNA sequence [**20**]. Alternatively, a second KH3KH4 protein could bind to the unoccupied cognate sequence (*see* **Note 14**). The second scenario leads to the formation of a 2:1 protein-RNA complex, whereas the first leads to RNA remodelling.

Both pathways for formation of the closed complex involve a bimolecular step followed by what is in effect a conformational change. For such a mechanism the equilibrium dissociation constant (K_d) for formation of the closed complex is given by [**21**]:

$$K_d = \frac{K_dA.K_dB}{1 + K_dB} \tag{12}$$

In the case of the upper pathway $K_dA = k_{off}3/k_{on}3$ and $K_dB = kO4/kC4$

$$K_d = \frac{K_dA.{kO4}/{kC4}}{1+{kO4}/{kC4}} = \frac{K_dA.kO4}{kC4+kO4} \tag{13}$$

$$kC4 = \frac{kO4(K_dA - K_d)}{K_d} \tag{14}$$

The data available from our measurements allowed us to calculate the *kC4/kO4* ratio using **Equation (14)** but not the absolute values. However, a value of *kC4* (~ 9.3 $^{-1}$) was obtained from this equation by making the reasonable assumption that *kO4* is, in fact, the same as $k_{off}4$. In the case of the lower pathway $K_dA = k_{off}4/k_{on}4$ and $K_dB = kO3/kC3$. Assuming, as above, that *kO3* is the same as $k_{off}3$ a value of *kC3* (~ 2 s^{-1}) was calculated from the following equation.

$$kC3 = \frac{kO3(K_dA - K_d)}{K_d} \tag{15}$$

At the low protein concentrations used in our experiments both pathways for formation of the closed complex include a conformational change step that is very much faster than the initial bimolecular binding step. Under these conditions the rate expressions for reactions occurring exclusively by the upper and lower pathways in **Figure 4** are given by **Equations (16)** and **(17)** respectively (*see* **Note 15**)

$$k_{obs} = k_{on}3[Protein] + \frac{k_{off}3.kO4}{kC4 + kO4} \tag{16}$$

$$k_{obs} = k_{on}4[Protein] + \frac{k_{off}4.kO3}{kC3 + kO3} \tag{17}$$

For a reaction occurring exclusively by the upper pathway the k_{on} and k_{off} values would be 3 x 10^4 $M^{-1}s^{-1}$ and 0.00063 s^{-1}. For a reaction occurring exclusively by the lower pathway the k_{on} and k_{off} values would be 1.4 x 10^5 $M^{-1}s^{-1}$ and 0.0029 s^{-1}. These latter values are very close to the values observed in our experiments with the wild type KH3KH4 di-domain (1.4 x 10^5 $M^{-1}s^{-1}$ and 0.0033 s^{-1}), suggesting that the lower pathway dominates in both the association and

dissociation steps. For both pathways the overall dissociation constant for formation of the closed complex is ~ 20 nM, as it must be (*see* **Note 16**).

4. NOTES

Note1: The new Octet RED96e system is an enhancement to the Octet RED96 instrument that permits assays to be performed over a slightly wider temperature range (15-40°C), allowing for kinetic measurement of unstable proteins. An evaporation cover for microplates also results in minimal sample evaporation for up to 12 hours. The Octet RED384 is a 16-channel instrument that provides analytical performance similar to the 8-channel Octet RED96 systems.

Note 2: The BSA and Tween-20 are included to minimise non-specific binding and are not always needed. In practice many different buffer systems can be used and in some cases the buffer requires additional additives. For example, in the measurements with oligonucleotides we routinely use 10 mM Sodium Phosphate, pH 7.4, 50 mM NaCl, 0.5 mM TCEP, with 0.5 mg/ml BSA, 0.002% Tween-20, and RNAse inhibitor (RNAsin, Promega) at 40u-100u mL^{-1}

Note 3: Oligonucleotides should be de-protected by following the manufacturer's instructions, lyophilized, and resolubilised in the appropriate buffer. Final oligonucleotide concentrations are then calculated from absorption spectroscopy by using the Beer-Lambert law ($A = c \varepsilon l$, where A is the absorption, c the concentration in $mol \cdot L^{-1}$, ε the extinction coefficient in $L \cdot mol^{-1} \cdot cm^{-1}$, and l the pathlength in cm).

Note 4: The lowest concentration of immobilized oligonucleotide that gives enough signal in the protein association step should be selected as overloading the biosensor may lead to overcrowding and steric hindrance. It is generally the case that slow loading for a long time is preferable to fast loading in a short time.

Note 5: It can sometimes happen that not all sensors give the same response in this step. In this case, the measured association or dissociation amplitudes can be normalized for the different loading levels. With some oligonucleotides the response during the loading phase can be very small, making it difficult to ensure equal loading of all the biosensors. In some cases we have found that using a different salt concentration in step 2) increases the size of the response so that equal loading can be confirmed. The sensors then need to be returned to the experimental buffer in step 3).

Note 6: The baseline signal after loading in step 2) should be stable; that is there should be no leaching of the bound RNA. This is almost always the case with streptavidin (SA) biosensors but may not be with other sensor types. If leaching does occur then it is generally the case that reducing the concentration in the loading step reduces the extent of the leaching.

.

Note 7: In the ideal case the length of the association phase should be long enough to allow all response curves to approach close to equilibrium but this is not always possible, particularly in studies of high affinity interactions which require the use of low concentrations and therefore slow binding kinetics.

When repeating the measurement with a different set of protein concentrations in order to cover the appropriate concentration range, the loading level reached in step 2) must be the

same in each measurement as the instrument response, but not the kinetics of the response, is directly proportional to the loading level. If this is not the case, the instrument response (signal amplitude of the association or dissociation phase) can be normalized for the loading level (*see* **Note 5**).

Note 8: The manufacturers recommend that the duration of the dissociation phase should be long enough to give at least 5% dissociation. Although this may be reasonable in some cases, it is advisable to remember that the determination of the dissociation rate from such limited dissociation is based primarily on the assumption that the response following complete dissociation will be identical to that recorded prior to the association step (i.e., the response in step 3 in **Figure 1**). Any instrument drift or residual non-specific binding can therefore have a significant effect on the dissociation rate constant determined.

Note 9: Regeneration and subsequent reuse of biosensors offers the user considerable cost savings. The immobilized molecule must be stable under the regeneration conditions employed (generally high or low pH, high salt concentration or added detergent) and must retain binding capacity over several regeneration cycles. In addition, bound molecules must of course be completely removed by the regeneration process.

Note 10: In rare cases the amount of non-specific binding observed with a loaded biosensor can be greater than that seen on an unloaded one. The presence of a slow phase in the association step that never reaches equilibrium followed by incomplete dissociation is an indication that there are problems with non-specific binding of the protein to the sensor surface.

Note11: SSA biosensors are super-streptavidin sensors. They have higher density of streptavidin on the surface compared to the SA sensors. This allows for a higher binding signal and interactions over a larger surface area of the biosensor will be specific; thus reducing non-specific binding. Note however that SSA biosensors are much more expensive than SA.

Note 12: This approach is particularly useful if one of the proteins gives such a poor response that it is not possible to obtain a dissociation constant using the standard protocol described in **Section 3.1**

Note 13: The coupling is defined as weak because in the case of perfect coupling the affinity of the KH3KH4 construct would be equal to the product of the affinities of the RNA:KH3KH4(DD) and RNA:KH3(DD)KH4 complexes (~1.5 μM and ~0.9 μM), i.e., ~ 1.4 pM.

Note 14: The alternative binding pathway, i.e., binding of a second protein to the same RNA, would require a significantly higher affinity for the two interactions because the concentrations of protein and RNA used in our experiments are low compared with the K_ds for the binding of the individual domains.

Note 15: These equations only apply if the conformational change is very much faster than the bimolecular step over the range of protein concentrations being examined. If the bimolecular step is faster than the conformational change under all conditions, then there would be two kinetic phases with the fast phase varying linearly with protein concentration

and the slow process varying hyperbolically with protein concentration. For the upper pathway the observed rates of the fast and slow processes would be given by:

$$k_{obs}(F) = k_{on}3[Protein] + k_{off}3$$

$$k_{obs}(S) = \frac{kC4[Protein]}{K_d3 + [Protein]} + kO4$$

Note 16: The Gibbs free energy (ΔG) of a reaction depends only on the free energy of the products (the final state) minus the free energy of the reactants (the initial state). The ΔG of a reaction is therefore independent of the path (or molecular mechanism) of the transformation. The free energy change, and therefore the K_d, must therefore be the same for both pathways.

5. REFERENCES

[1] Licatalosi DD, Darnell RB (2010). RNA processing and its regulation: global insights into biological networks. *Nature Reviews. Genetics*, *11*: 75–87.

[2] Morris AR, Mukherjee N, Keene JD (2010), Systematic analysis of posttranscriptional gene expression. *WIREs Syst Biol Med*, *2*: 162–180.

[3] Gerstberger S, Hafner M, Tuschl T. (2014). A census of human RNA-binding proteins. *Nat Rev Genet*, *15*: 829–845.

[4] Marinov KG, Williams BA, McCue K, Schroth GP, Gertz J, Myers GM, Wold BJ (2014) From single-cell to cell-pool transcriptomes: Stochasticity in gene expression and RNA splicing. *Genome Res. 24*: 496–510.

[5] Gronland GR, Ramos A (2018). The devil is in the domain: understanding protein recognition of multiple RNA targets. *Biochem Soc Trans. 45*: 1305-1311.

[6] Lunde BM, Moore C, Varani G (2007). RNA-binding proteins: modular design for efficient function. *Nature Reviews. Molecular Cell Biology*, *8*: 479–90.

[7] Nicastro G, Taylor IA, Ramos A (2015) KH – RNA interactions : back in the groove. *Curr. Opin. Struct. Biol.* 30: 63–70 (2015).

[8] Ciesielski GL, Hytönen VP, Kaguni LS (2016) Biolayer Interferometry: A Novel Method to Elucidate Protein-Protein and Protein-DNA Interactions in the Mitochondrial DNA Replisome. *Methods Mol Biol*. 1351:223-31.

[9] Lou X, Egli M, Yang X (2018) Determining Functional Aptamer-Protein Interaction by Biolayer Interferometry. *Curr Protoc Nucleic Acid Chem*. 67: 25.1-7.25.15.

[10] Sultana A, Lee JE (2015) Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Curr Protoc Protein Sci*.79:19.25.1-19.25.26.

[11] Cukier CD, Hollingworth D, Martin SR, Kelly G, Diaz-Moreno I, Ramos A. (2010) Molecular basis of FIR-mediated *c-myc* transcriptional control *Nature Struct Mol Biol* 17:1058-1064

[12] Hollingworth D, Candel AM, Nicastro G, Martin SR, Briata P, Gherzi R, Ramos A (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Research* 40:6873-6886.

13] Candel AM, Hollingworth D, Nicastro G, Martin SR, Briata P, Gherzi R, Ramos A (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *FEBS Journal* 279 (Suppl 1):478-478.

[14] Nicastro G, García-Mayoral MF, Hollingworth D, et al (2012) Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. *Nature Struct Mol Biol* 19:1282-1286.

[15] Nicastro G, Candel AM, Uhl M, et al (2017) Mechanism of β-actin mRNA recognition by ZBP1. *Cell Reports* 18:1187-1199.

[16] Dagil R, Ball NJ, Ogrodowicz RW, et al (2019) IMP1 KH1 and KH2 domains create a structural platform with unique RNA recognition and re-modelling properties. *Nucleic Acids Res* 47:4334-4348

[17] Yisraeli JK (2005) VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. *Biol Cell* 97:87-96

[18] Wächter K, Köhn M, Stöhr N, Hüttelmaier S. (2013) Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. *Biol Chem*. 394:1077-1090

[19] Patel VL, Mitra S, Harris R, et al (2012) Spatial arrangement of an RNA zipcode identifies mRNAs under post-transcriptional control. *Genes Dev* 26:43–53.

[20] Chao JA, Patskovsky Y, Patel V, et al (2010) ZBP1 recognition of β-actin zipcode induces RNA looping. *Genes Dev* 24:148–158.

[21] Eccleston JF, Martin SR, Schilstra MJ (2008) Rapid kinetic techniques. *Meth Cell Biol* 84:445-477.

Figure 1 – Computer simulation of the five steps required in a 'typical' BLI experiment (see text for details). The curves were simulated with k_{on} = 4 x 10^5 $M^{-1}s^{-1}$, k_{off} = 0.006 s^{-1} and protein concentrations ranging from 4 to 500 nM

Figure 2 – BLI data for the interaction of IMP1 constructs with MYCRNA

(a) wild-type KH1KH2: serial dilutions from 0.25 μM (0.25, 0.13, 0.06, 0.03 μM)

k_{on} ~ 1 x 10^6 $M^{-1}s^{-1}$, k_{off} ~ 0.047 s^{-1}, K_d ~ 47 nM

(b) KH1KH2(DD): serial dilutions from 1 μM (1, 0.5, 0.25, 0.13 μM).

k_{on} ~ 2.7 x 10^5 $M^{-1}s^{-1}$, k_{off} ~ 0.48 s^{-1}, K_d ~ 1.76 μM

Figure 3 – RNA binding by the protein regulator IMP1/ZBP1. a) Domain organization of IMP1. b) Inter-domain arrangement and RNA binding by the ZBP1 KH3KH4 di-domain structural unit. The surface representation of the bound KH3KH4 protein (grey) and the ribbon representation of the protein backbone (blue) are shown. The two bound cognate RNA sequences from the well characterized β-actin mRNA target (CACA for KH3 and CGGAC for KH4) are displayed using a stick representation coloured by atom type. A dashed line has been traced to represent the connection between these two sequences, which does not make contact with the protein. The image has been built by superimposing the NMR structures of the KH3KH4DD-CACA and KH3DDKH4-CGGAC complexes.

Figure 4 - Kinetic model for the interaction of KH3KH4 constructs from ZBP1 with a 28-nucleotide Zipcode RNA (ACCGGACUGUUACCAACACCCACACCCC).

$k_{on}3$ and $k_{off}3$ were determined from experiments with KH3KH4(DD), $k_{on}4$ and $k_{off}4$ were determined from experiments with KH3(DD)KH4, k_{on} and k_{off} were determined from

experiments with wild-type KH3KH4. The remaining constants were estimated as described in the text.