Protein expression and the Biacore™

Mafalda Santos
T cell biology Group

19th November 2019

Talk structure

1. Protein expression **basics**

i. Why express proteins **at all**?
ii. How to **decide** on an expression strategy
iii. **Judging** protein quality

2. Biacore **basics**

i. Principles
ii. Generating data
iii. **Analysing the data**: affinity, kinetics

Protein expression basics

mafalda.santos@rdm.ox.ac.uk

I. Why express and study proteins?

1. Proteins are of fundamental interest: biological systems are all about **protein recognition**
2. An understanding of biological phenomena very often depends on **knowing how proteins behave**
3. Can expect **hard answers** to scientific questions: *e.g.* Is this how my protein looks?
4. Modern science is reagent-driven so the choice of protein can **set the research agenda**
5. This can provide many opportunities for collaboration

II. The expression strategy

Your protein ⇨ Cytosolic?

⇩

Bacterial expression
(*e.g.* pET vectors)
- **fast**
- often **very large** amounts

II. The expression strategy

Your protein ⇨ Secreted or membrane bound?

⇩

1. Bacterial re-folds
- yields can be **low** (~1%)
- refold conditions generally **differ** for each protein
- sparse-matrix screens are available to help

II. The expression strategy

Your protein ⇨ Secreted or membrane bound?

⇩

2. Bacterial **secretory** systems
- *e.g.* pET-12a,b,c vectors
- yields often **very low**

II. The expression strategy

Your protein ⇨ Secreted or membrane bound?

1. Yeast (*e.g. Pichia)*
- **fast**
- very high yields
- **metabolic labelling** (NMR)
- deglycosylation possible
- **poor folding** of *e.g.* IgSF proteins

⇩

needs to be glycosylated or doesn't want to refold?

⇦

mafalda.santos@rdm.ox.ac.uk

III. The expression strategy

Your protein ⇨ Secreted or membrane bound?

⇩

needs to be glycosylated or doesn't want to refold?

⇦

2. Baculovirus
- can be **very slow**
- **modest** yields: 1-5 mg/L
- very good for some proteins *e.g.* **MHC II**

II. The expression strategy

Your protein ⇨ Secreted or membrane bound?

⇩

needs to be glycosylated or doesn't want to refold?

⇦

3. Mammalian cells
(*e.g.* CHO K1 cells or 293T cells)
- can be **fast**
- potentially **very high** yields (≤400 mg/L)
- sugars can be removed (*e.g.* Lec3.2.8.1 cells)
- **transient or stable**

Mammalian expression

Basic features

- expressed proteins are **generally designed to be secreted**, but can be put on the cell surface or made intracellularly
- soluble expression of membrane proteins is achieved by inserting a stop codon **immediately before** the TM domain, but maintaining **one** signal peptide at the N terminus
- proteins are glycosylated; **refolding unnecessary**
- the **more "intact" the protein, the better**
- fusion proteins, his-tagged proteins can be made

Transient expression

Advantages

- quick; transient expression takes 3-5 days
- excellent for **testing constructs**
- various fusion partners
- transfection with **$CaPO_4$** (cheap) or **lipids** (fast)

Disadvantages

- need to **repeat transfection** every time
- **beware** of Fc fusion proteins - Fc folds very efficiently, possibly taking mis-folded protein with it

Stable expression

Recommended: The GS system

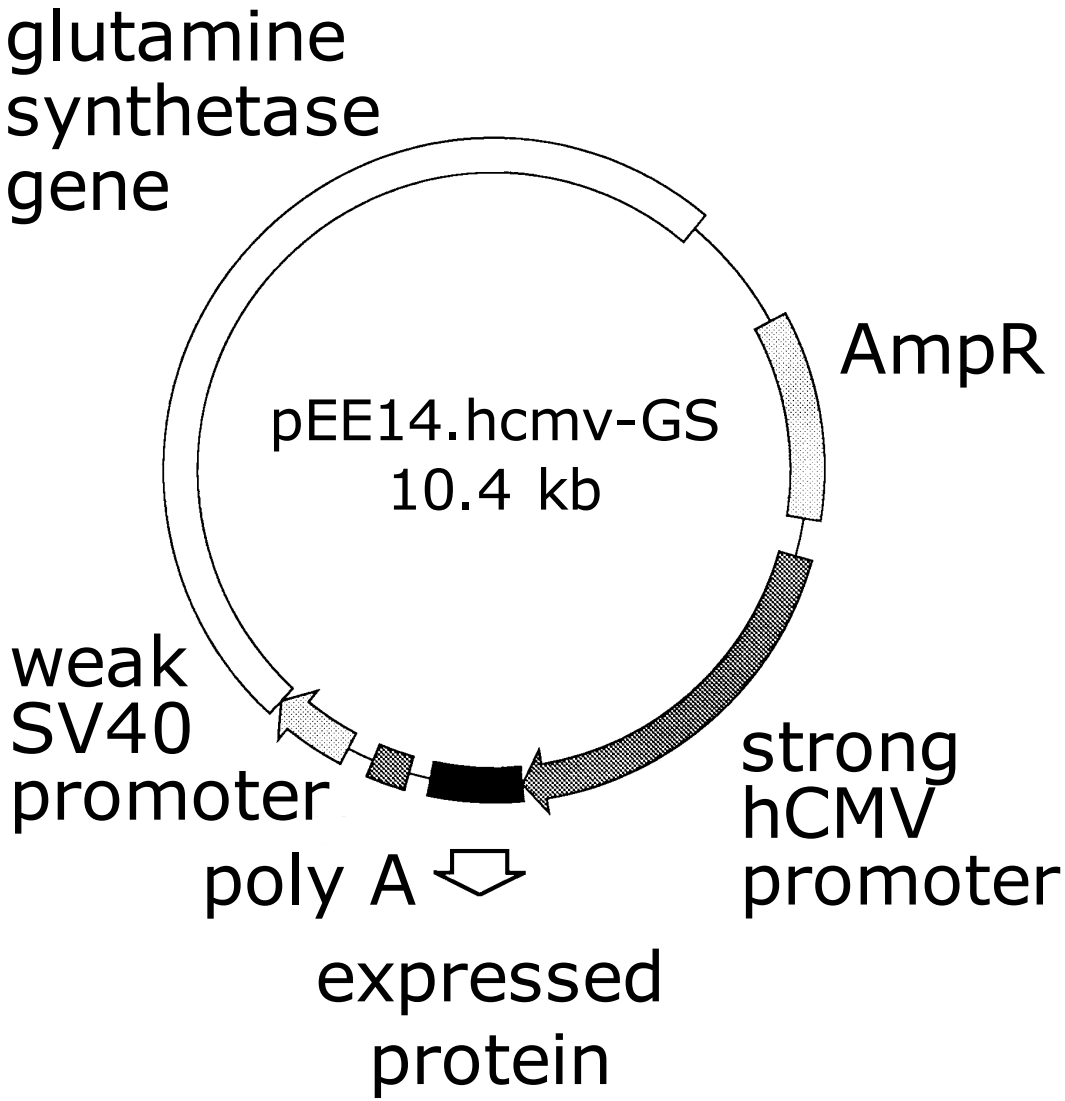
- CHO cells transfected with **$CaPO_4$** or **lipids**
- selection *via* the **glutamine synthetase** (GS) gene
- CHO cells have their own GS gene but can be killed with GS inhibitor, **methionine sulphoximine** (MSX)
- cells with extra GS from the plasmid **survive higher levels of MSX** than the mocks
  => more copies = better survival
- expression is driven by **strong hCMV promoter**

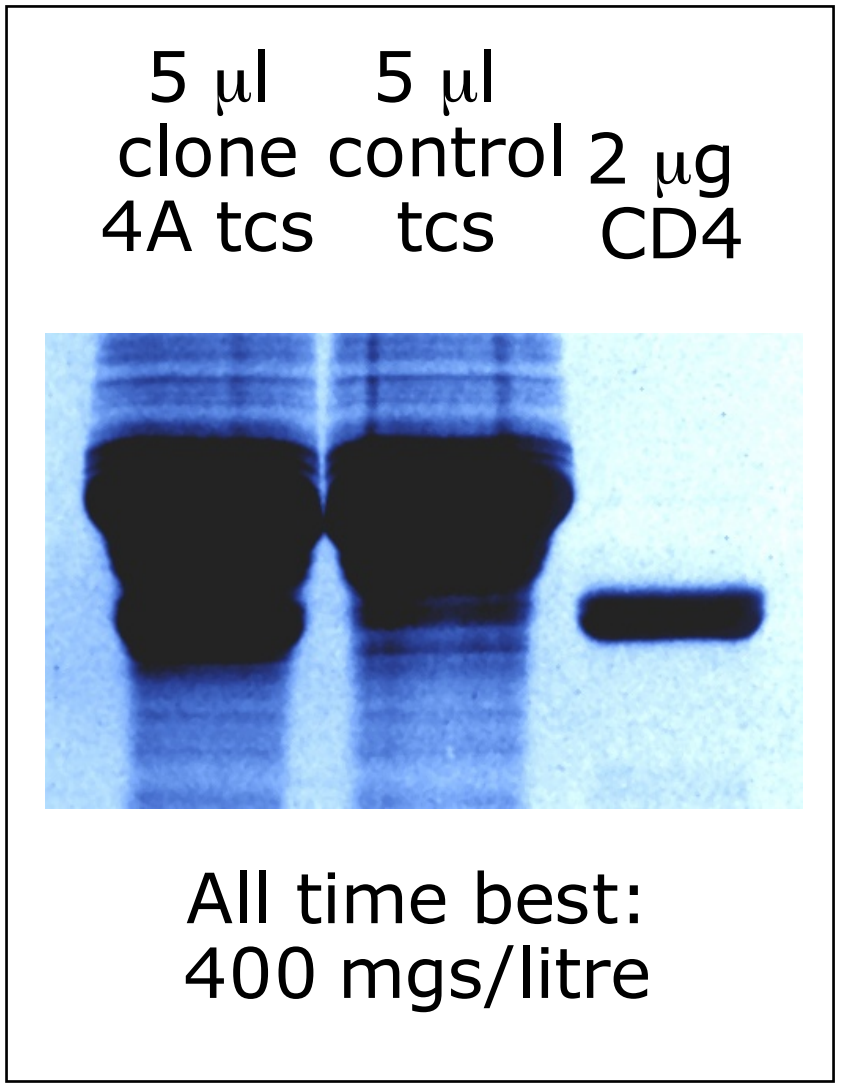
Stable expression

Recommended: The GS system

- selection takes **2 weeks**
- potentially **prodigious expression** levels
- can make enough protein to thoroughly **confirm that it's OK**
- mutant CHO cells can be used to alter glycosylation, *e.g.* Lec3.2.8.1 cells

The glutamine synthetase-based gene expression system





All time best:
400 mgs/litre

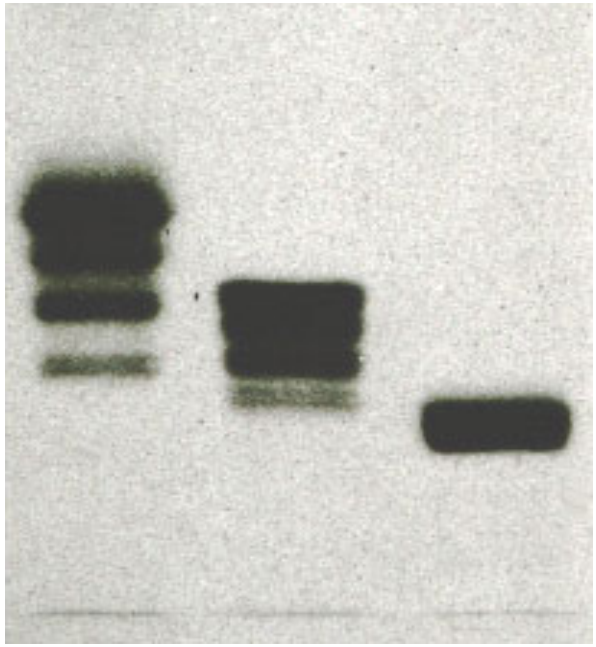
mafalda.santos@rdm.ox.ac.uk

Expression of rat sCD2 for structural studies in CHO mutant Lec3.2.8.1 cells

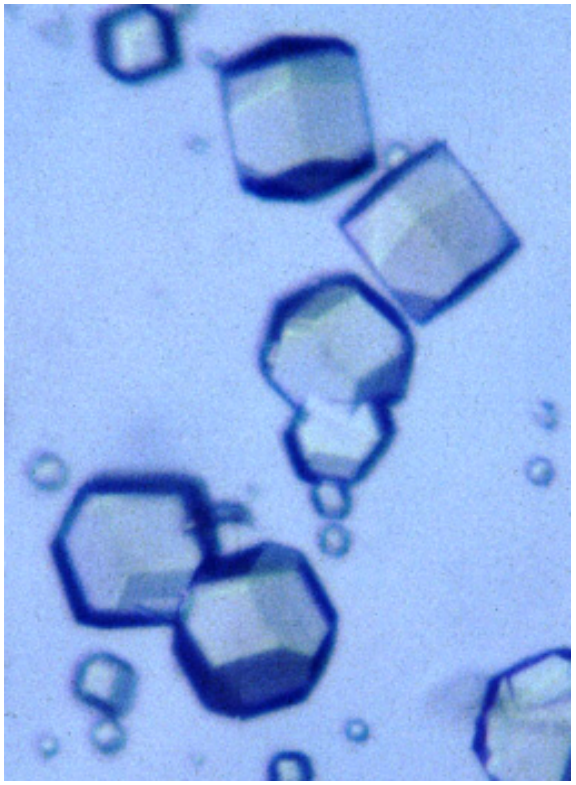
CHO-K1

Lec3.2.8.1

Lec3.2.8.1 + endo H



Deglycosylated sCD2 crystals



mafalda.santos@rdm.ox.ac.uk

Stable expression

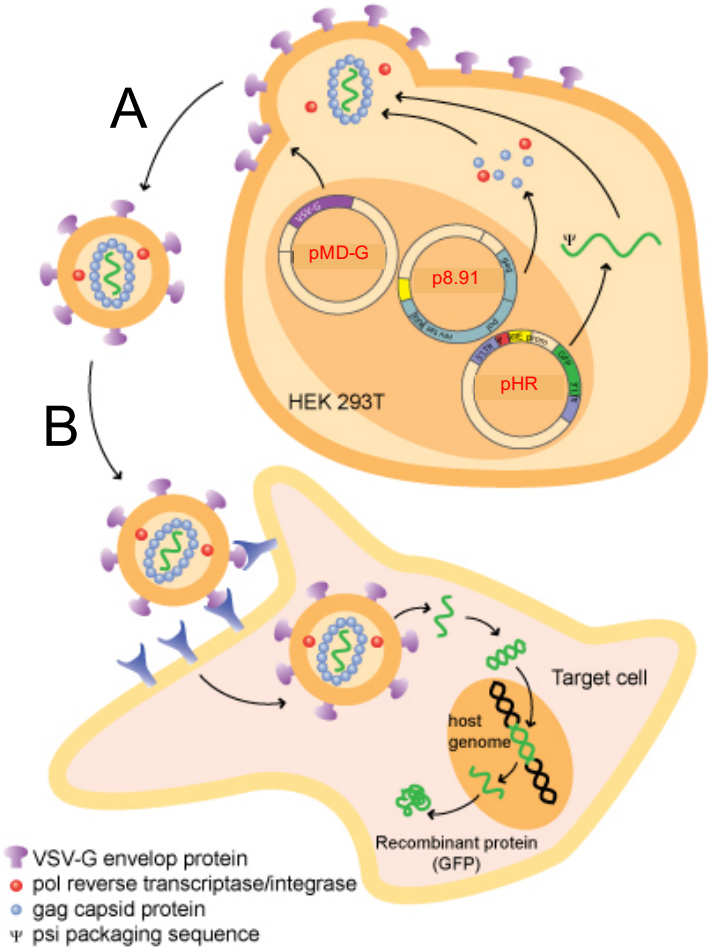
Recommended: Lentiviral Expression

- no selection required – faster process to obtain protein
- **high expression** levels of protein of interest
- protein of interest expressed fused with a fluorescent protein (GFP, citrine) - expression of membrane proteins

The lentiviral expression approach

Based on HIV-1, whose genome was split up into three different vectors:

- a **lentiviral cloning plasmid (pHR)**, containing packaging sequence and the gene of interest.
- **p8.91 plasmid**, encoding the *pol*, ga*g*, *rev* and *tat* viral genes.
- **pMD-G plasmid**, encoding the VSV-G envelope gene.



Adapted from "Lentiviral Vector Production and Cell Transduction Review"; *invivoGen 2011*

III. Is my protein any good?

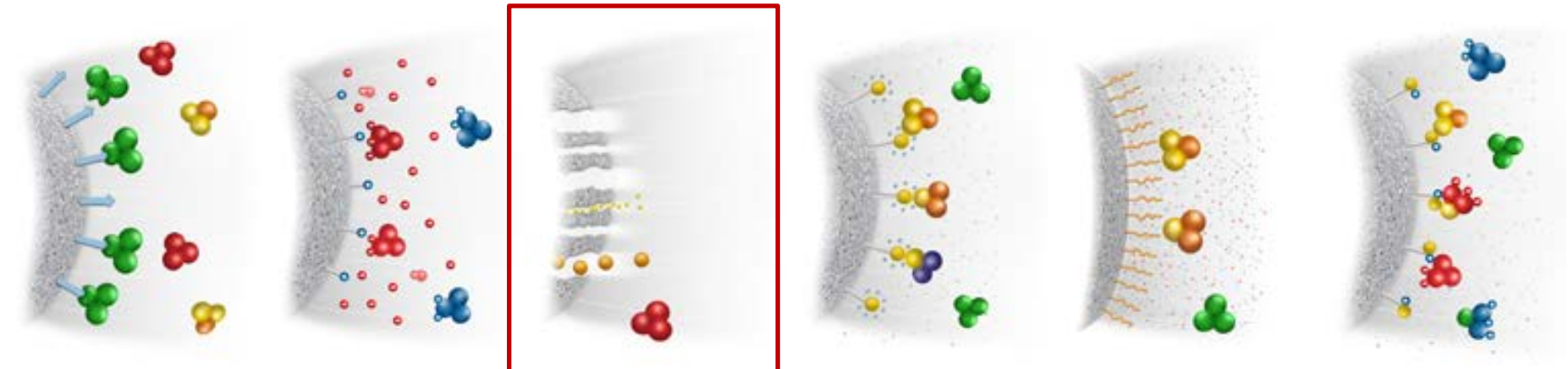
Whether you really ask this question is the **decisive factor** for a good Biacore experiment!

Good signs

- it's expressed at **high levels**
- if cys-containing, it runs at the **right size** on non-reducing SDS-PAGE (compare to reducing)
- the protein is stable/active for days/weeks **at 4ºC**
- the protein binds mAbs ***stoichiometrically***
- the protein is **soluble** at high concentrations

III. Is my protein any good?

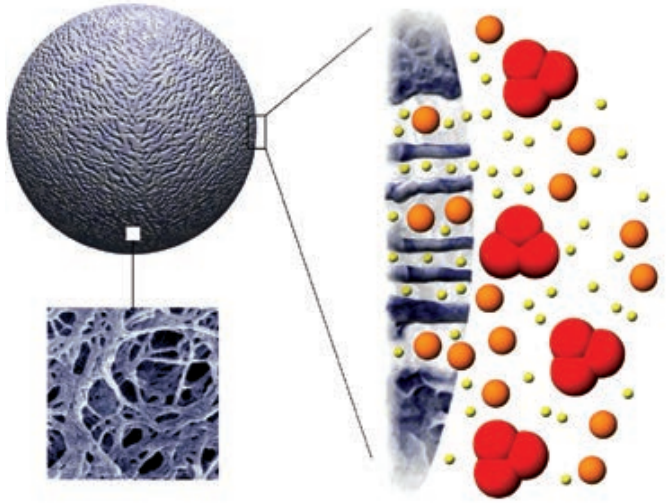
Property	Technique
Size	Size exclusion chromatography (SEC), also called gel filtration (GF)
Charge	Ion exchange chromatography (IEX)
Biorecognition (ligand specificity)	Affinity chromatography (AC)
Hydrophobicity	Hydrophobic interaction chromatography (HIC) Reversed phase chromatography (RPC)
Various (e.g., charge, hydrophobicity, and hydrogen bonding)	Multimodal chromatography (MMC)

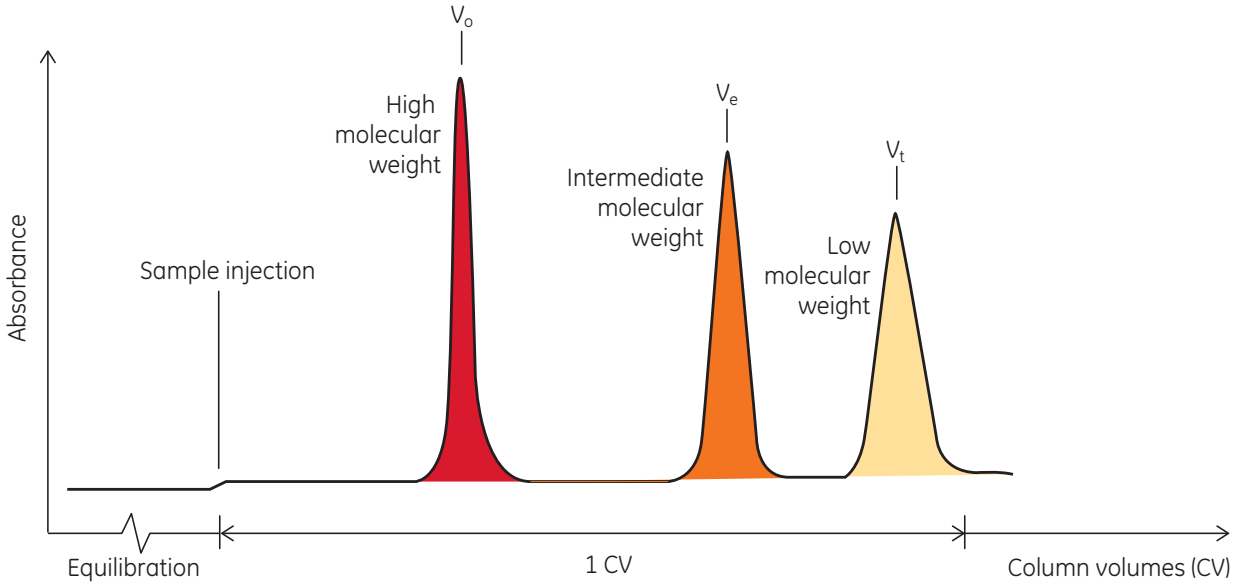


Affinity | Ion exchange | Size exclusion | Hydrophobic interaction | Reversed phase | Mulitmodal chromatography

III. Is my protein any good?

The size exclusion principle





III. Is my protein any good?

The absolute key for doing Biacore experiments and structural studies properly:

- **the protein is non-aggregated according to gel filtration.**
    - N.B. SDS-PAGE gels **do not** tell you this!

Protein expression and the Biacore

2. Biacore Basics

I. Biacore principles

II. Generating the data

III. Analyzing the data

The Biacore

What is it used for?

The **accurate** measurement of the binding properties of interacting molecules

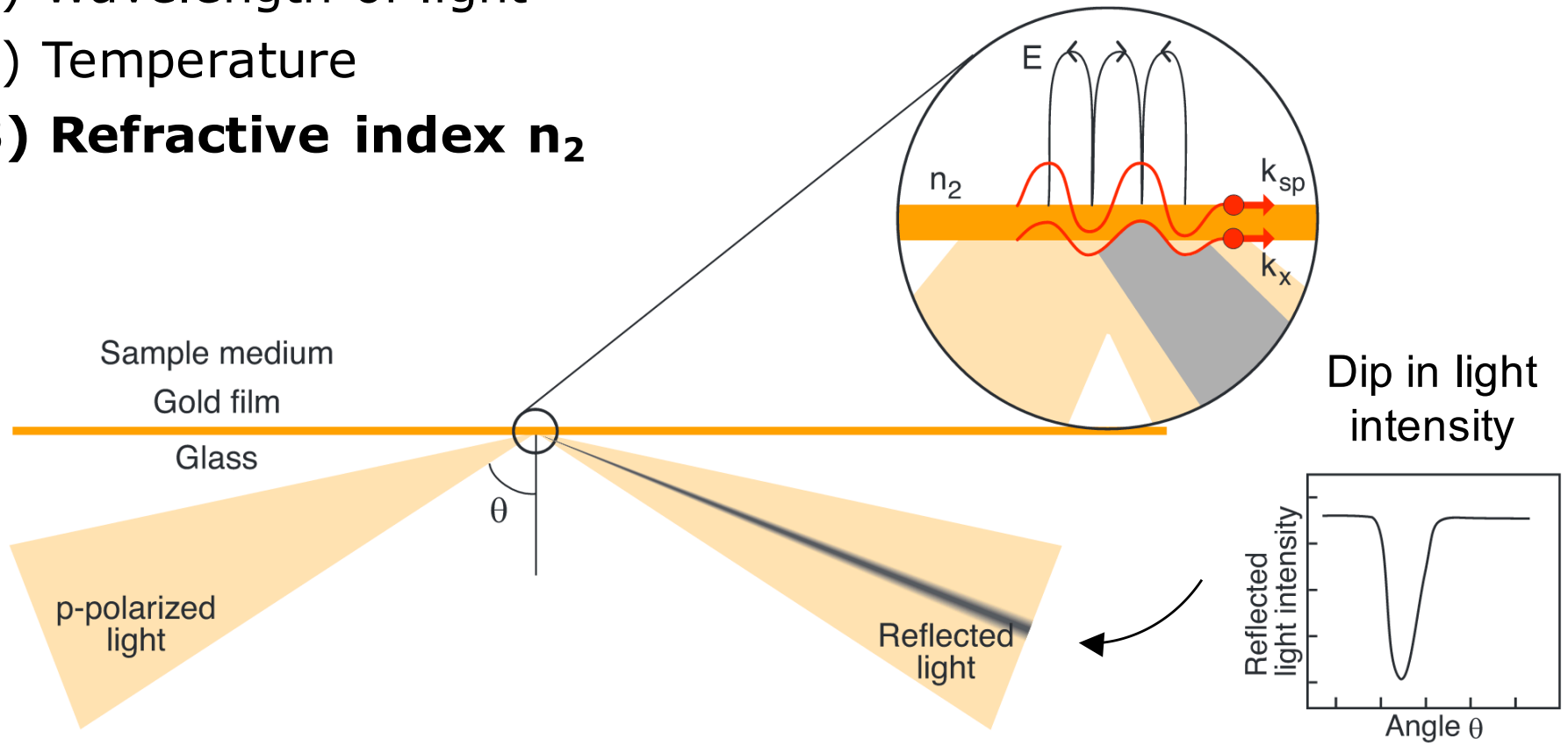
Why do we want to do this?

A **full understanding** of the functions of molecules often requires detailed insights into the nature of their interactions

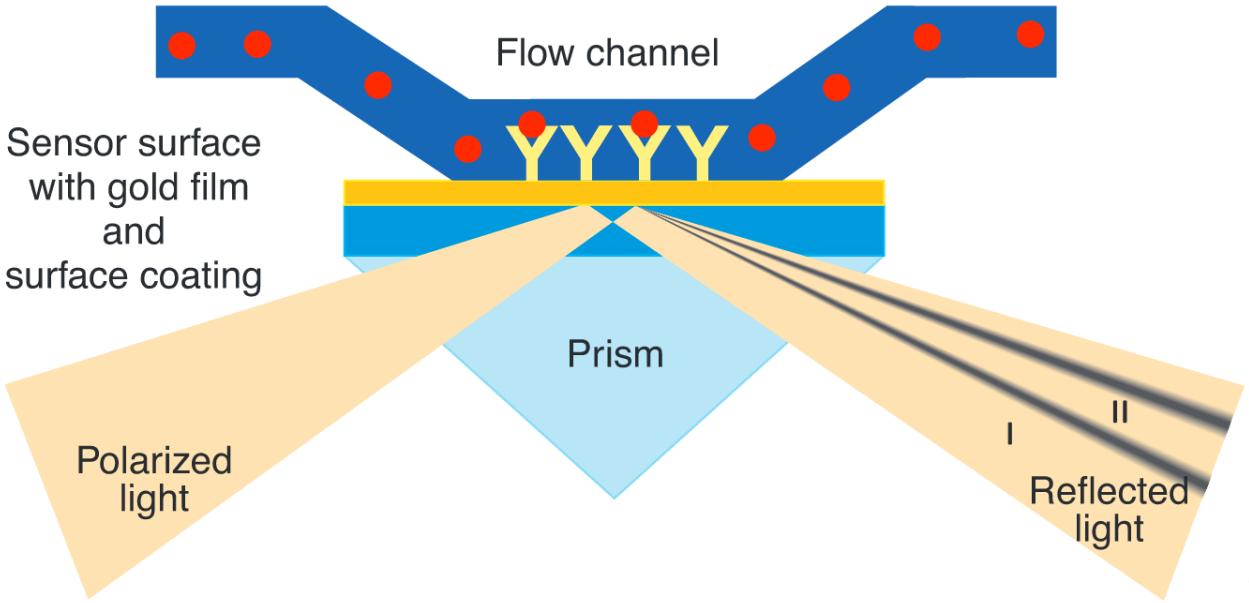
Surface plasmon resonance (SPR): the principle

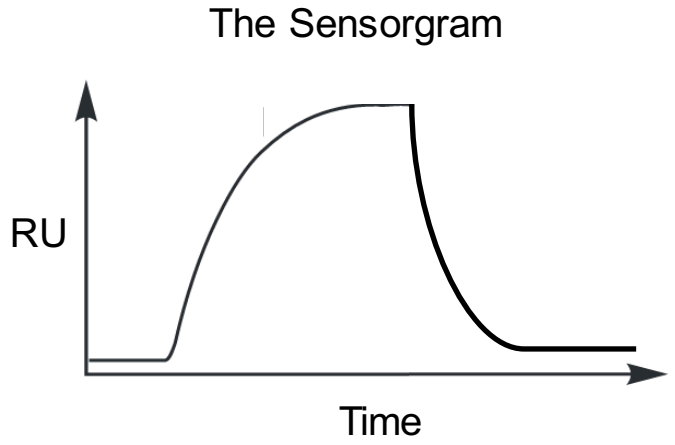
Angle of the "dip" is affected by:

1) Wavelength of light
2) Temperature
3) **Refractive index n_2**



SPR as implemented in the Biacore





Signal = response units (RU)

$\triangle 1$ RU = $\triangle 1$ pg protein/mm^2

Four crucial advantages for detecting binding:

- Interaction can be monitored in **real time**
- No wash step, so now ***very low affinities*** can be measured
- No need to **label** your proteins
- Need **very little** protein

Uses of the Biacore

- **Equilibrium** measurements:
  - Solution affinities (K_A or K_D)
- **Kinetics**: determination of k_a and k_d
- Analysis of **specificity** (*e.g.* drug screening)
- Competition assays (*e.g.* antibodies)
- In combination with **mutagenesis**:
  - Epitope mapping
  - Contribution of residues to binding energy
- Isolation of binding components from a mixture
  - Unknowns are identified by linked MS
- Can study binding of **proteins, small molecules, DNA, RNA**

Protein expression and the Biacore

2. Biacore Basics

I. Biacore principles

II. Generating the data

III. Analyzing the data

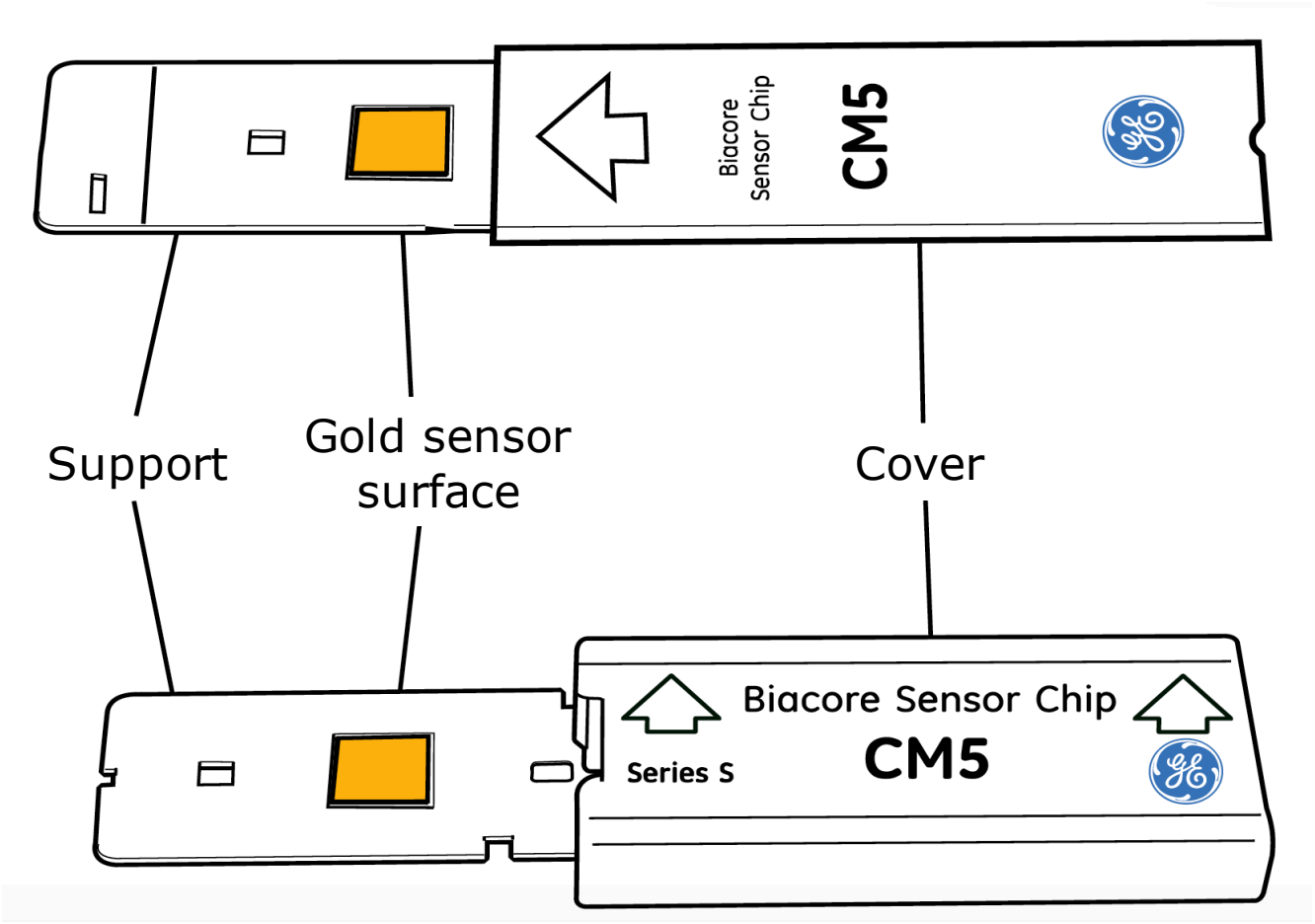
Protein expression and the Biacore

Biacore S200

Biacore 8K

mafalda.santos@rdm.ox.ac.uk

Sensor Chip technology

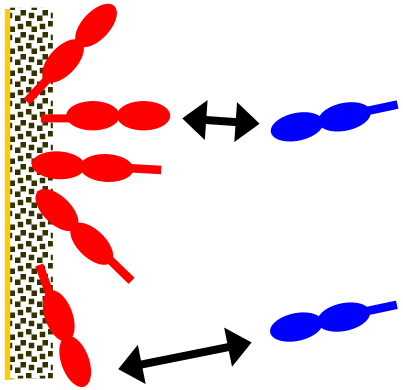


Biacore binding experiments
Step 1: Immobilize your "Ligand"

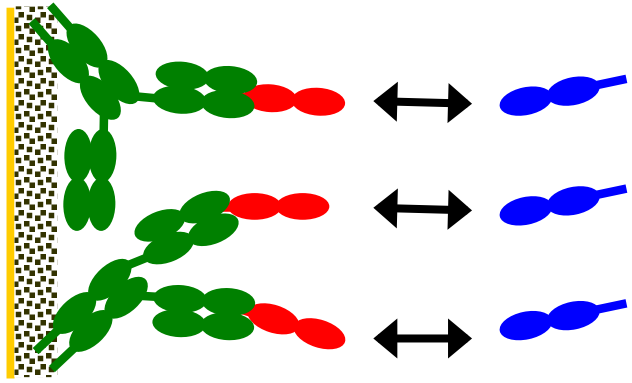
Two main options:

- **Direct**:
  Covalently bind your molecule to the chip,
  but beware protein inactivation
- **Indirect**:
  First immobilise something that binds your molecule with high affinity *e.g.* streptavidin, antibodies

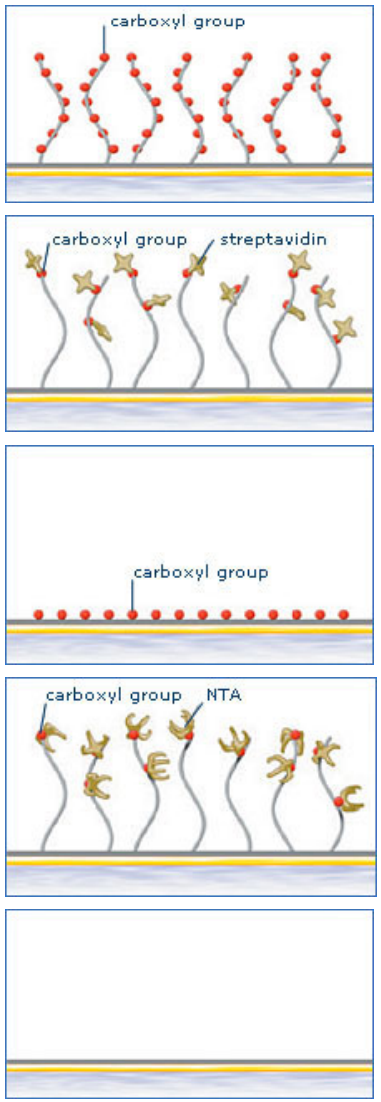
Direct



Indirect

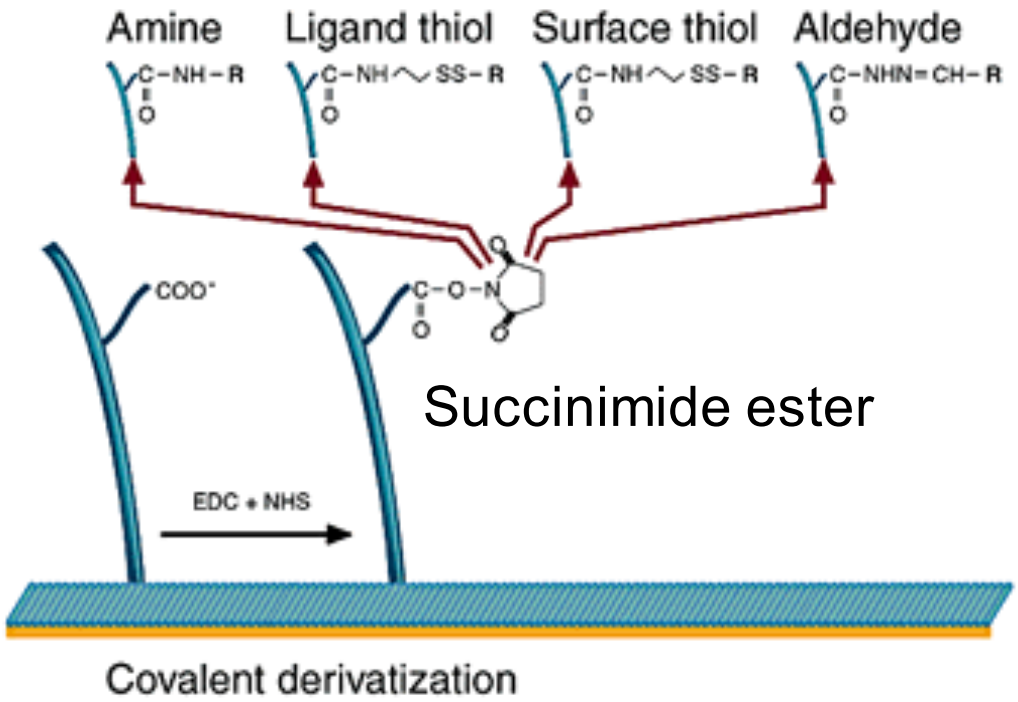


Sensor Chip technology



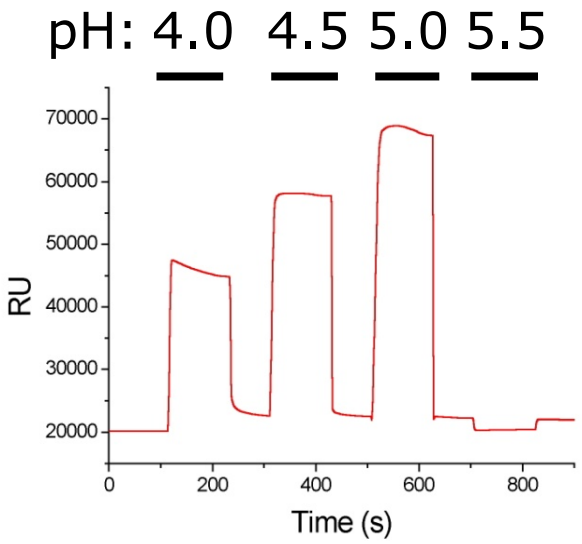
- CM5 (most common) – also CM3, CM4 and CM7:
  Ligand capture *via* **native groups**
- SA: Coated in streptavidin for capture of **biotinylated molecules**.
- C1: Like CM5 but without dextran matrix =>
  - Low immobilisation level
  - Bind large particles e.g. **cells and viruses**
- NTA: Capture of ligands by metal chelation
  - *e.g.* **His-tagged proteins**
- HPA: **Flat hydrophobic surface**, adding liposomes forms **lipid monolayers**, containing any molecules you inserted into the liposome

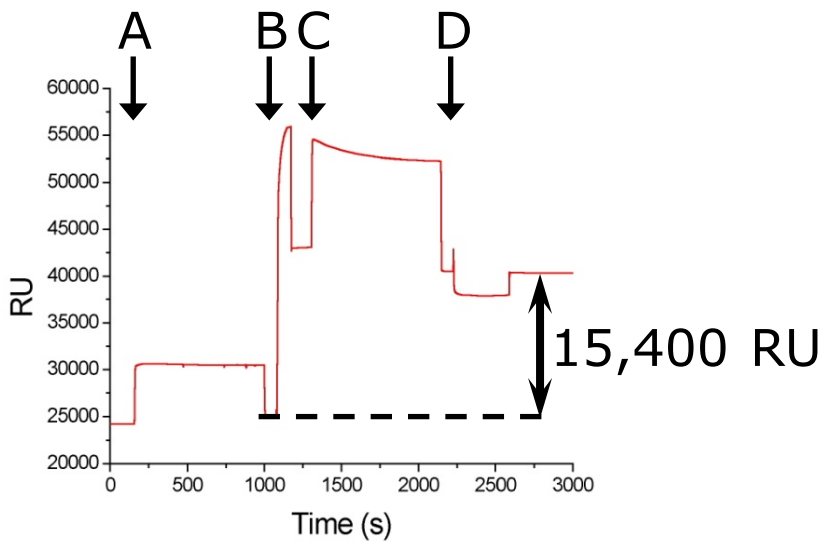
Direct coupling to the dextran surface of a CM5 chip



EDC = 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
NHS = *N*-Hydroxysuccinimide

Example of antibody immobilization





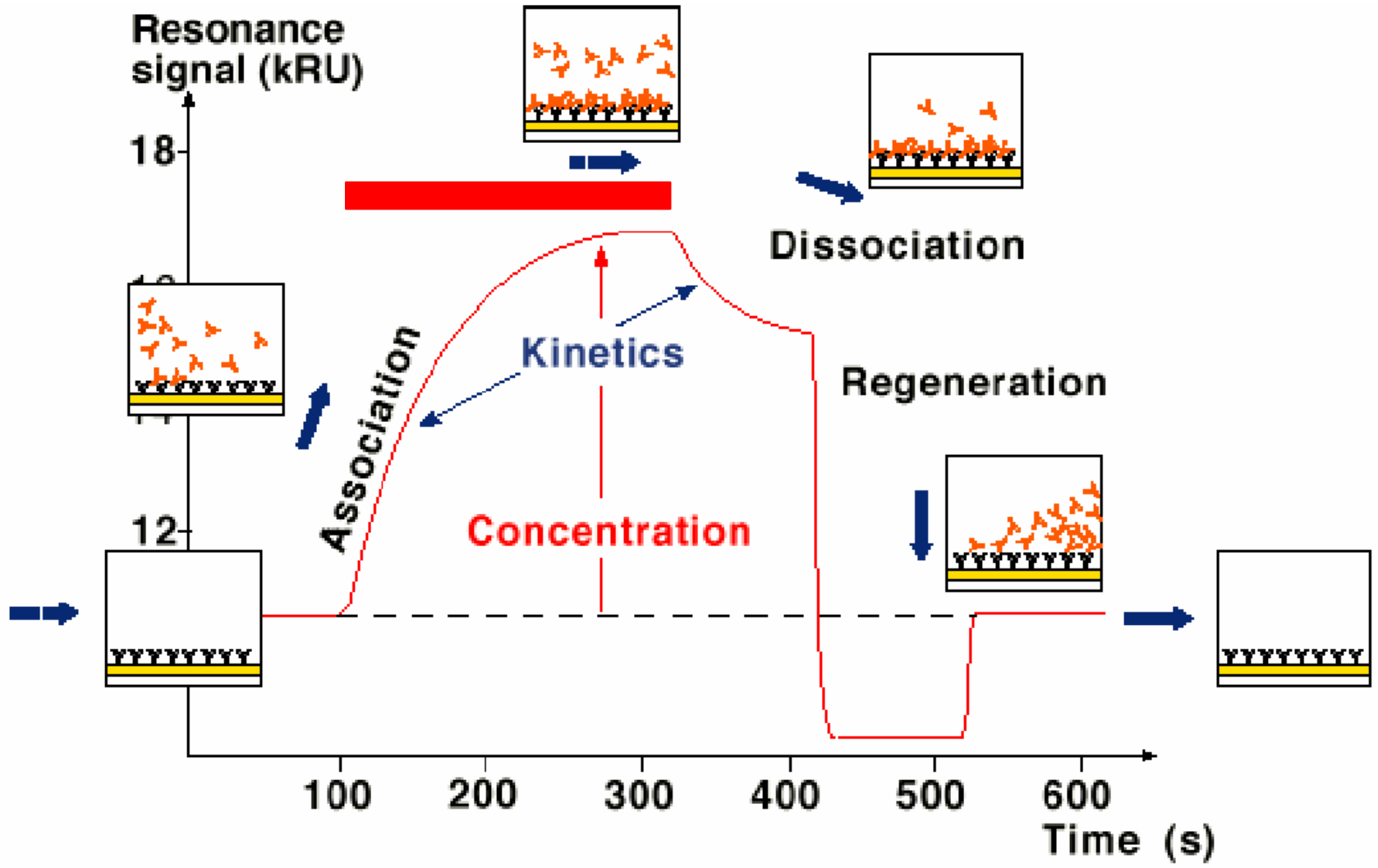
<u>Pre-concentration screen</u>

An antibody was diluted in buffers of **different pH** and injected over a non-activated chip. Maximum electrostatic attraction occurs at pH 5

<u>Immobilisation</u>

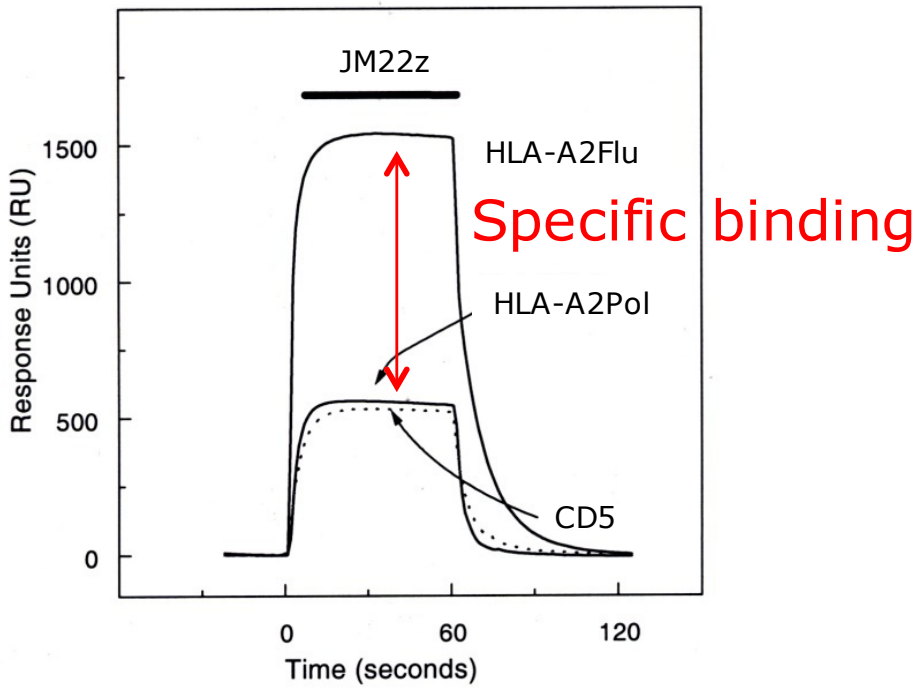
A. Inject 70µl 1:1 **EDC:NHS**
B. Inject 7µl **mAb** in pH5 buffer (in this case @370µg/ml)
C. Inject 70µl **ethanolamine**
D. Inject 30µl 10mM **glycine** pH2.5

Biacore binding experiments
Step 2: Observing binding



A note about controls

- Each chip has four "flow-cells"
- Immobilise different ligands in each flow-cell
- At least one control flow-cell is required to see "background"
- **"Specific binding"** is the response in flow-cell with the protein of interest minus response in a control flow-cell



2. Biacore Basics

I. Biacore principles

II. Generating the data

III. Analyzing the data

Analysing Biacore data
1. Affinity measurements

1. Measures how **favourable** an interaction is
2. Expressed as the **affinity constant**: K_A

3. For A + B ↔ AB

$$K_A = \frac{[AB]_{eq}}{[A]_{eq}[B]_{eq}} = \frac{1}{K_D}$$

- best thought of as the ratio of [products] *vs.* [reactants] **at equilibrium**
- note the units (M^{-1})
- higher affinity = higher K_A

Analysing Biacore data
1. Affinity measurements

4. Also expressed as dissociation constant: K_D
   - The inverse of K_A; **more intuitive**
   - Usually thought of as concentration of **A at which half of B is bound** ([B]=[AB]) at equilibrium

$$K_D = \frac{[A]_{eq}[B]_{eq}}{[AB]_{eq}}$$

   - Units are M
   - Higher affinity = lower K_D

Analysing Biacore data
1. Affinity measurements

- Could simply measure [A], [B] and [AB] **at equilibrium** and calculate K_D
- In practice this is difficult and the **following approach** is used
- Increasing fixed concentrations of analyte ([A]) are added to **a fixed small amount** of its ligand B and you measure the amount of bound A (*Bound*)
- Plot the results and fit the 1:1 **Langmuir equation** to the data to determine K_D and $Bound_{max}$

DERIVATION

$$K_D = \frac{[A][B]}{[AB]} \quad (1)$$

and

$$[B] = [B_{total}] - [AB]$$

$$= [AB]_{max} - [AB] \quad (2)$$

By substitution of (2) into (1) and rearranging, we get

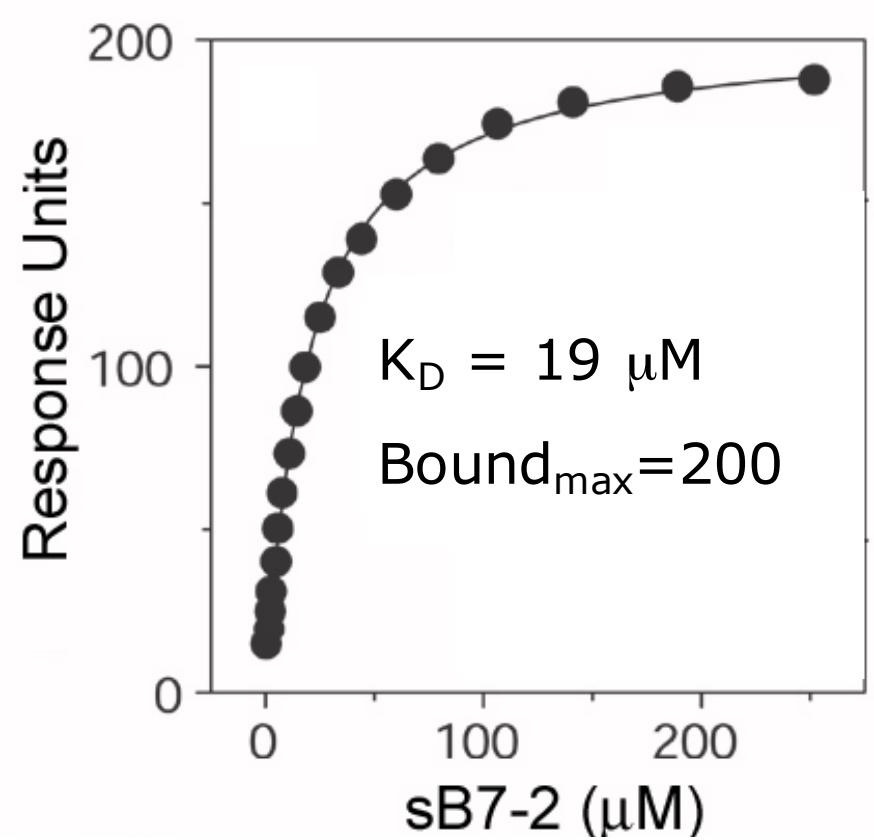
$$[AB] = \frac{[A][AB_{max}]}{[A] + K_D}$$

or

$$Bound = \frac{[A]Bound_{max}}{[A] + K_D}$$

mafalda.santos@rdm.ox.ac.uk

An example:
B7-2 binding to CD28



$$Bound = \frac{[A]Bound_{max}}{[A] + K_D}$$

- Data are circles
- Line is **non-linear fit** of the equation performed computationally (e.g. Origin)
- Gives the indicated values for K_D and $Bound_{max}$
- If the fit is good it indicates that binding follows the **simple 1:1 model**
- **Difficult to see if the fit is poor in this type of plot**

The Scatchard plot

- A plot of $\frac{Bound}{[A]}$ versus $Bound$
- **Linear** for a 1:1 interaction
- If curved it indicates wrong model and possible **problem with the experiment**
- Most commonly **concave up** (caused by experimental error e.g. protein heterogeneity)
- Far less common is to see **concave down** (usually caused by positive co-operativity)

DERIVATION

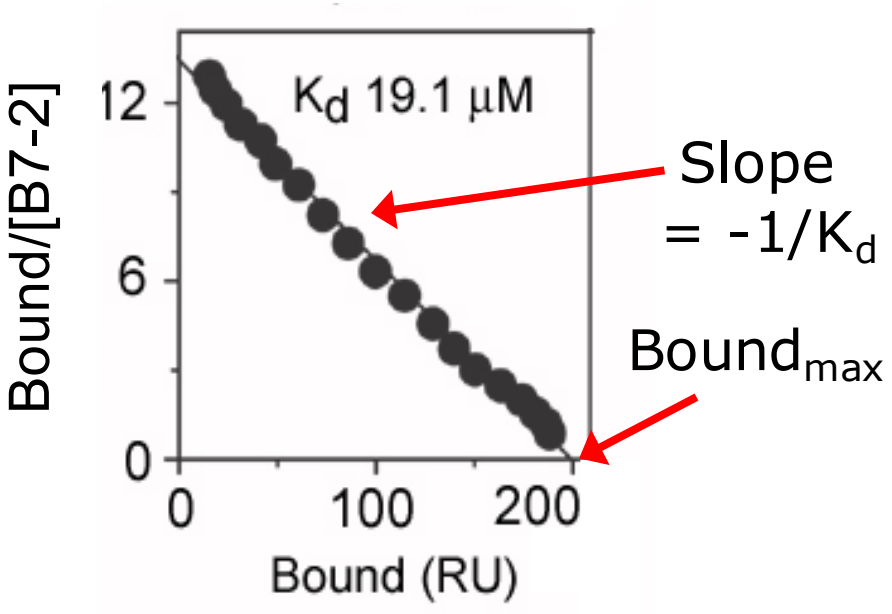
$$Bound = \frac{[A]Bound_{max}}{[A] + K_D}$$

$$Bound[A] + Bound \times K_D = [A]Bound_{max}$$

Divide both sides by $[A]K_D$ and rearrange, giving

$$\frac{Bound}{[A]} = -\frac{1}{K_D}Bound + \frac{Bound_{max}}{K_D}$$

Plot of $\frac{Bound}{[A]}$ versus $Bound$



Analysing Biacore data
2. Kinetics measurements

Biological systems are **rarely at equilibrium**, so the **rates** of binding and dissociation are critical

For a simple 1:1 interaction ($A + B \leftrightarrow AB$)

- Rate of association
  - $d[AB]/dt = k_{ass}[A][B]$
- Rate of dissociation
  - $d[AB]/dt = k_{diss}[AB]$

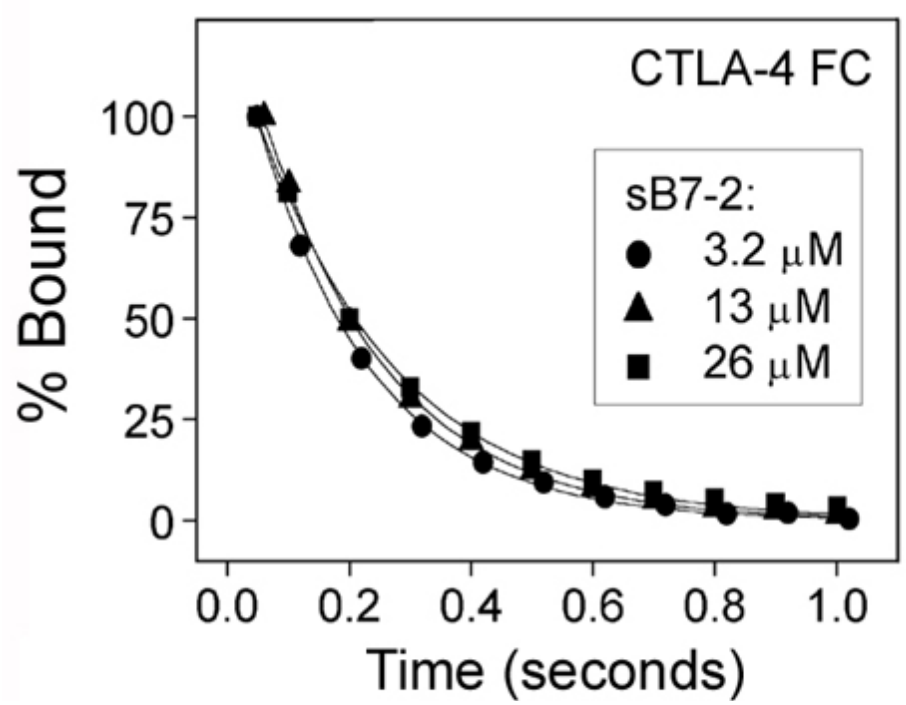
At equilibrium the rate of association **must equal** the rate of dissociation

$$k_{diss}[AB] = k_{ass}[A][B] \Rightarrow$$
$$k_{diss}/k_{ass} = [A][B]/[AB] = K_D$$

Dissociation

Dissociation is a **simple exponential process** so, using the relationship

- $[AB]_t = [AB]_o e^{-k_{diss}t}$
- k_{diss} can be determined directly by **curve fitting**



$k_{diss} = 5.1\ s^{-1}$

Association

In most experimental systems association cannot be studied **in the absence of simultaneous dissociation**

- Therefore you need to know $\mathbf{k_{off}}$ and **[A]** as well as measuring **[AB]** to calculate the k_{on}
- We use the following relationship

$[AB]_t=[AB]_{final}(1-e^{-kobst})$

where $k_{obs} = k_{ass}[A]+k_{off,}$

and globally fit the entire reaction to calculate k_{ass} computationally

Factors affecting kinetics

- The association rate constant does not vary that much
  - Association requires two proteins to collide in **correct orientation** and in the **correct conformation**
  - **Depends on diffusion** so will be similar for most proteins
  - The **basic rate** is about 10^5 $M^{-1}.s^{-1}$
  - Can be accelerated by **long range electrostatic forces**
    - Increased rate of collision
    - Steer binding sites into correct orientation

- The dissociation rate constant **varies considerably** and is responsible for most variation in affinity constants
  - It is determined by the **number and strength of bonds** in the contact interface
  - In turn depends on **size of interface** and the degree of **geometric and electrostatic complementarity**

Potential pitfalls in kinetic analysis

- Protein Problems: **Aggregates** (common)
  Concentration **errors**
  Artefacts of construct
- Choice of controls: **Bulk refractive index issues** due to different levels of immobilisation
  Use **both orientations** (if possible)
- Mass Transport: Rate of binding **limited by rate of injection**: $\mathbf{k_{on}}$ will be underestimated
- Rebinding: Analyte **rebinds** before leaving chip
  $\mathbf{k_{off}}$ will be underestimated

Last two can be spotted if measured k_{on} and k_{off} vary with immobilisation level

The HIU Biacore Facility

1. T200

- Very sensitive – can detect LMW analyte binding (~200 Da)
- Evaluation of results is RELATIVELY easy
- Easy programming is done using templates and wizards that are provided with the software

2. 8K

- High-throughput, high-sensitivity, eight needle system, high-performance SPR system
- Screening characterization, process optimization, and quality control of small molecules and biotherapeutics.

- Bookings: http://ww.imm.ox.ac.uk/internal
  > facilities booking system > HIU Biacores

Reading/watching

The Biacore Bible
By Anton van der Merwe

http://www.bioch.ox.ac.uk/aspsite/services/equipmentbooking/biophysics/spr.pdf

Principle of SPR in Biacore™ systems

https://www.youtube.com/watch?v=o8d46ueAwXI