Protein reporters to study in vivo protein interactions and aggregation

Montserrat Morell Fernández

INDEX

Index

PART 2

CHAPTER 3

ABBREVIATIONS

Abbreviations

AD	Alzheimer's disease
AFM	Atomic force microscopy
Amp	Ampicillin
AP-MS	Affinity purification and mass spectrometry
ATP	Adenosine Triphospate
BFP	Blue Fluorescent Protein
BIFC	Bimolecular Fluorescence complementation
BRCA1	Breast cancer type 1 susceptibility protein
BRET	Bioluminescence Resonance Energy Transfer
CE	Capillary Electrophoresis
CHC	Central Hydrophobic Cluster
Cm	Chloramphenicol
CR	Congo Red
CYFP	C-terminal fragment of Yellow Fluorescent Protein
Dab	2,4-diaminobutyric acid
DHFR	Dihydrofolate reductase
DTT	Dithiothreitol
FACS	Fluorescence-activated cell sorting
FC	Flow Cytometry
FCCS	Fluorescence Cross-correlation Spectroscopy
FL	Fluorescent light
fMTX	Methotrexate labeled with Alexa
FRET	Fluorescence Resonance Energy Transfer
FSC	Forward scatter
FT-IR	Fourier Transform Infrared spectroscopy
GFP	Green fluorescent protein
GuHCl	Guanidinium Hydrochloride
h	Hour
IBs	Inclusion Bodies
IPTG	isopropyl-β-D-thiogalactopyranoside
Kan	Kanamycin
K_d	Dissociation constant

LIF	Laser-induced fluorescence detection
LRET	Luminescence Resonance Energy Transfer
MALDI-TOF	Matrix-assisted laser desorption/ionization Time-of-flight
MAPPIT	Mammalian protein-protein interaction trap
MeArg	*N*-methyl-arginine
min	minutes
MS	Mass Spectrometry
MTX	Methotrexate
NYFP	N-terminal fragment of Yellow Fluorescent Protein
o.n.	overnight
PCA	Protein Complementation Assays
PD	Parkinson's disease
Pip	4-amino-4-carboxy-piperidine acid
PK	Proteinase K
PMTs	Photomultipliers
poliQ	Polyglutamine expansions
RP-HPLC	Reverse-phase high pressure liquid chromatography
r.p.m	Revolutions per minute
Rluc	Renilla luciferase protein
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SRP	Surface plasmon resonance
SSC	Sideward scatter
TAP	Tandem affinity purification
TEM	Transmission Electron Microscopy
TFA	Trifluoroacetic acid
Th-T	Thioflavine T
TIRF	Total internal reflection fluorescence
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride
Wt	Wild-type
Y2H	Yeast-two-hybrid
YFP	Yellow Fluorescent Protein
α-Syn	α-Synuclein
β–gal	β–galactosidase

Note about the format

For a better comprehension, this thesis has been divided in two different parts, with two chapters each. The chapters correspond to five publications. However, the publications are not reproduced in the text. Their format has been unified and a general section of Experimental Procedures has been added. Besides, specific methodologies for each work are included in the corresponding chapter.

INTRODUCTION

Introduction

In recent years, information on the genomes of many organisms has been made available. From these sequences, hypothetical proteins have been derived. However, the physico-chemical properties, structure and function of most of these proteins still have to be solved.

Once a new protein has been discovered, one important and relevant aspect is its function. It has to be taken into account that many cellular functions are executed by protein complexes acting like molecular machines. The term 'functional proteomics' derives from the hypothesis that the association of proteins would suggest their common involvement in a biological function, analogous to the 'guilt by association' concept in criminal investigation. Therefore, one of the follow-ups of the whole-genome sequencing is a variety of large-scale studies aimed to find out how the proteins encoded within them interact and the consequences of their interactivity.

During many years, the methods developed to study protein interactions were performed *in vitro* and were only able to study strong bindings. However, the majority of interactions that take place in the cell are transient and weak. Therefore, it has become a key issue in functional proteomics to develop strategies that are able to detect and study this type of interactions *in vivo.*

The deciphering of a large number of interaction networks has allowed the acquirement of a comprehensive view of the cell and its biochemical processes. Besides, many human diseases are related to aberrant protein-protein interactions like the loss of a crucial interaction or the establishment of an abnormal binding between endogenous proteins or with pathogen proteins. Protein interactions have been recognized as challenging but attractive targets for chemical drugs. In particular, the inhibition of protein interactions by small drug-like chemical drugs has been intensively studied and the outcome of these investigations suggests that this strategy could lead to new and effective treatments for human diseases. This accounts for an urgent need of techniques that allow studying the specific inhibition of protein interactions *in vivo*.

Another way to get a comprehensive view of the cellular processes is structural proteomics. These large initiatives are trying to solve the tree-dimensional structures of the complete protein set encoded by a specific genome. However, these investigations usually encounter problems in the protein production step because many polypeptides become aggregates when they are expressed at high levels in heterologous hosts. Bacterial cells are the default factories for recombinant protein productions: they can reach high cell densities, produce recombinant proteins in high yields and the growth media are quite inexpensive. Unfortunately many proteins do not fold properly when they are overexpressed in prokaryotic environments and form intracellular aggregates named inclusion bodies. Moreover, the process of protein aggregation is not exclusive of prokaryotic cells. It also occurs inside eukaryotic cells and it is implicated in increasing number of degenerative human diseases like Alzheimer or Parkinson. Recent studies suggest that the protein aggregates formed inside prokaryotic cells resemble those responsible of human diseases. Therefore, it has been suggested that they could provide a simple, yet powerful system for studying the formation and prevention of toxic aggregates, such as the ones responsible for a number of degenerative diseases. Deciphering the relationship between prokaryotic and eukaryotic aggregates has become an important issue in the field because it may imply the existence of evolutive conserved mechanisms to cope with protein aggregation in different organisms.

According to the emerging relevance of protein aggregation processes in both biotechnology and biomedicine, there is an increasing interest in protein solubility screening methods that allow foreseeing genes, chemical compounds or culture conditions that could modulate intracellular protein aggregation in order to develop tools and strategies that prevent the deposition of misfolded proteins inside the cell.

EXPERIMENTAL PROCEDURES

M.1 Laboratory equipment and reagents

M.1.1 Laboratory equipment

Atomic force microscope Asylum Research, model MFP-3D

Autoclave Matachana, model 500

Balances analytical and preparative Sartorius

Centrifuge bench-top Eppendorf, model 5415C

Centrifuge Heraeus, model Megafuge 20R

Flow cytometer Beckton Dickinson, model FACSCalibur

Flow cytometer and cell sorter DakoCytomation, model MoFlo

Fluorescence spectrophotometer Varian, model Cary Eclipse

FT-IR Spectrometer Bruker, model Tensor 27

Freezer -80 °C Forma Scientific

Gel electrophoresis equipment Bio-Rad, model Miniprotean 3

Heating blocks Labnet

Imager VersaDoc MP Bio-Rad

Laminar-flow hood Telstar, model CAM 1400-I.

Magnetic stirrers SBS

Mass spectrometer Bruker, model Ultraflex

Microscope Leica, model DMRB

Microscope confocal Leica, model TCS SP2 AOBS

Multilabel plate reader Perkin Elmer, model Victor^{3}V

MultiMode atomic force microscope, Nanoscope IV electronics

pH meter Radiometer, model pHM83

Shaking incubator Infors

Speedvac evaporator Thermo Savant, model UVS400A

Thermal cycler, automated for PCR Bioer, model Xp cycler

UV transilluminator Spectroline, model TC-365A

UV/visible-light spectrophotometer Varian, model Cary 400 Bio UV-Visible

Vortex Shaker Ika, model MSI

Water bath Polyscience

M.1.2 Reagents

Generally, the reagents have been purchased in Sigma Aldrich and Merck.

M.2 Molecular biology methods

M.2.1 Strains

M.2.1.1 Escherichia coli strains

. *Escherichia coli* strain XL1 blue (Stratagene)

Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F ´*proAB* lacI^{q}ZΔ*M15* Tn*10* (Tet^{r})].

This strain has been used in order to obtain high efficiency in plasmid transformation.

. *Escherichia coli* strain BL21(DE3)[1]

Genotype: *E. coli* B F^{-} *dcm ompT hsdS*($r_B^- m_B^-$) *gal* λ(DE3)

This strain has been used for high-level protein expression using T7 RNA polymerase-based expression systems.

M.2.1.2 Saccharomyces cerevisiae strains

. *Saccharomyces cerevisiae* strain FY834

Genotype: *MAT* a *ura*3-52 *leu2*-1 *trp1*-63 *his3*-200 *lys2*-202

This strain has been used in preliminary assays and in experiments related with the overexpression of chaperones.

. *Saccharomyces cerevisiae* strain BY4741

Genotype: *MAT* a *his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*

This strain is the one used to create strains with a deletion in one chaperone. It is provided by Euroscarf (European *Saccharomyces cerevisiae* Archive for Functional analysis)

.*Saccharomyces cerevisiae* strain *erg6*Δ

Genotype: *MAT* a *his*3Δ1 *leu*2Δ0 *met*15Δ0 *ura*3Δ0 *erg6*Δ*::kanMX4.*

This strain (based in the BY4741 parental background) has been used in the drug testing experiments. It has a mutation that affects cell permeability, specifically, in the gene codifying C-24 sterol methyltransferase Erg6p.

M.2.2 Vectors

M.2.2.1 Escherichia coli vectors

. pBAT4 expression vector[2]

Plasmid used for protein expression in *E.coli*. This vector has a promoter inducible by ITPG and ampicillin resistance. Its size is 4.1 kbp

. pET 28a(+) expression vector (Novagen)

Plasmid used for protein expression in *E.coli*. This vector has a promoter inducible by ITPG and kanamycin resistance. Its size is 5.3 kbp.

. pGEMT-Easy cloning vector (Promega)

Plasmid used for cloning PCR products in *E.coli*. This vector is a linear plasmid that has a 3´terminal thymidine to both ends. These single 3´-T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids. Its size is 3 kbp.

M.2.2.2 Saccharomyces cerevisiae vectors

. pESC vectors (Stratagene)

The pESC vectors are a series of epitope-tagging vectors designed for expression and functional analysis of eukaryotic genes in the yeast *S. cerevisiae*. These vectors contain the GAL1 and GAL10 yeast promoters and the yeast 2μ origin, which enables autonomous replication of the plasmids in *S. cerevisiae*. They have a selectable marker gene (*HIS3, TRP1, LEU2, or URA3*) to select and maintain the expression vector in yeast cells.

M.2.3 Growth media

M.2.3.1 Escherichia coli media

. LB (Luria-Bertani) media

Dissolve the following compounds to 800 ml H_2O:

10 g Bacto-tryptone

5 g yeast extract

10 g NaCl

Adjust pH to 7.5 with NaOH.

Adjust volume to 1 liter with ddH_2O and sterilize by autoclaving.

M.2.3.2 Saccharomyces cerevisiae media

. YDP (Yeast Extract Peptone Dextrose) media

Dissolve the following compounds to 800 ml H_2O:

10 g of BactoYeast extract

20 g of BactoPeptone

20 g Dextrose

Adjust volume to 1 liter with ddH_2O.

Sterilize by autoclaving.

. SC (Synthetic Complete) drop-out media

Dissolve the following compounds in 1 liter ddH_2O

6.7 g Yeast nitrogen base without amino acids

100 ml of the appropriate sterile 10x Drop Out Solution

Adjust the pH to 5.8 if necessary, and autoclave. Add the appropriate sterile carbon source, usually dextrose (glucose) to 2%.

. Drop Out solution

To make one liter of 10x –Leu/–Trp/–URA/–His Drop Out solution, combine the components listed in the table below taking into account the amino acid that has to be omitted from the Drop Out. For example, if the Drop Out is –Leu, do not add this component to the final solution.

Table M.1 Components of the Drop Out solution

Constituent	Final mg/ml	Stock solution for 100 ml dH2O	ml stock for 1l media
Adenine sulfate	20	200 mg*	10
Uracil	20	200 mg*	10
L-tryptophan	20	1 g	2
L-histidine-HCL	20	1 g	2
L-arginine-HCL	40	1 g	4
L-methionine	20	1 g	2
L-tyrosine	50	200 mg	25
L-leucine	60	1 g	6
L-isoleucine	60	1 g	6
L-lycine-HCL	50	1 g	5
L-phenylalanine	50	1 g*	5
L-aspartic	100	1 mg	10
L-glutamic acid	100	1 g*	10
L-valine	150	3 g	5
L-threonine	200	4 g*	5
L-serine	400	8 g	5

Some stocks should be stored at room temperature (indicated by an asterisk) to prevent precipitation, while the others should be refrigerated. Afterwards, the solution has to be autoclaved.

M.2.4 Polimerase chain reaction

One of the most important methodological inventions in molecular biology has been the polymerase chain reaction (PCR). It is a fast and easy method to amplify DNA *in vitro*. In this work, PCR has been used for three purposes:

- Introduction of restriction endonuclease recognition sites in both ends of the amplified DNA.

- Introduction specific mutations in the sequence.

- Obtaining the DNA codifying a fusion of two proteins.

It has to be taken into account that the amplified fragment has always to be sequenced to confirm that the polymerase has not introduced any mutation.

One of the key steps in any PCR reaction is the primer design. There are some factors that have to be considered:

- The specificity of PCR depends strongly on the melting temperature (Tm) of the primers (the temperature at which half of the primer has annealed to the template). Usually good results are obtained when the Tm's for both primers are similar (within 2-4°C) and above 60°C. The Tm for a primer can be estimated using the following formula:

$$T_m = 2°C \times (A + T) + 4°C \times (C + G)$$

- Another variable to look at is the inclusion of a G or C residue at the 3' end of primers. It helps to ensure correct binding at the 3' end due to the stronger hydrogen bonding of G/C residues. The primers should be examined for complementarities because they can form primer-dimers. Also the hairpin formation has to be checked because it may reduce significantly PCR performance. The primer's oligonucleotide GC-content should be between 40-60%.

All the primers used in this work have been synthesized by Roche or Invitrogen.

M.2.4.1 Construction of a fusion protein

In order to create the DNA that encodes for a protein fusion, PCR reactions have been used. The strategy followed in the case of a fusion between the C-terminus of protein 1 and N-terminus of protein 2 is depicted in the figure M.1.

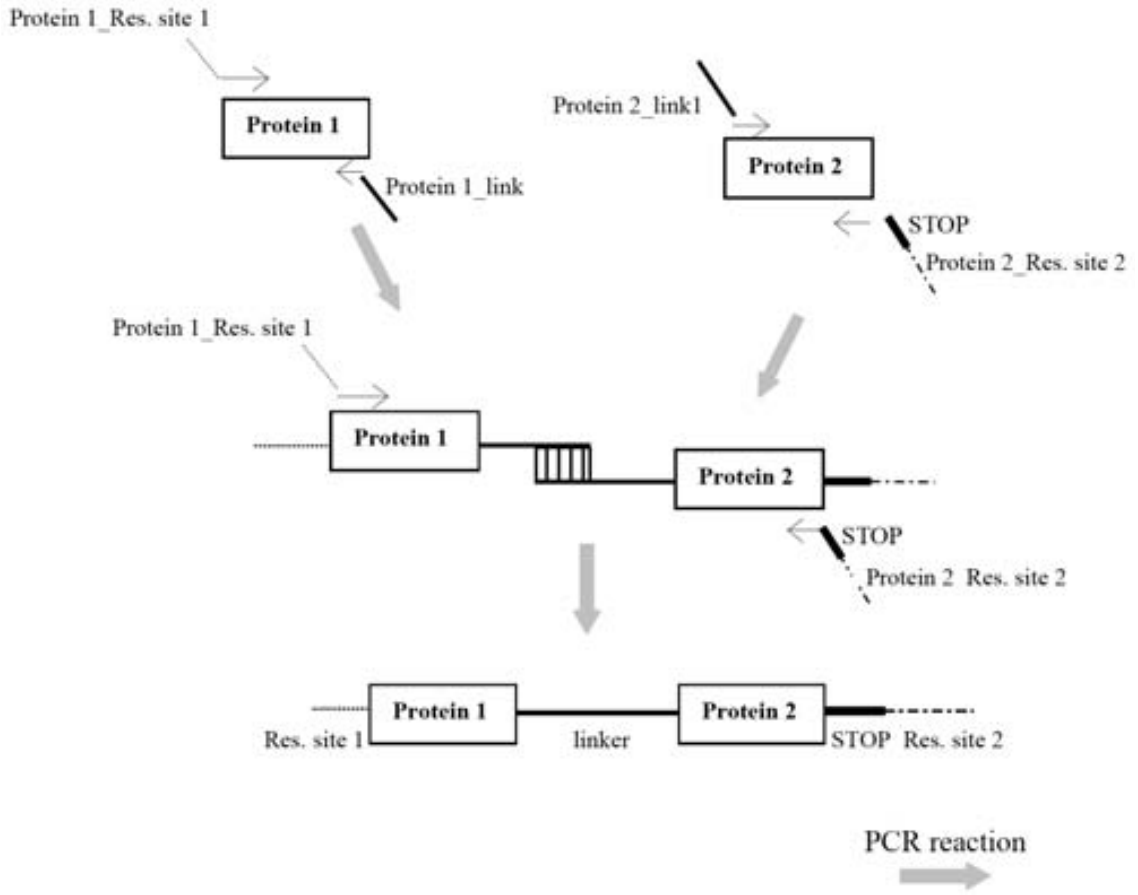


Figure M.1. Scheme of the strategy followed to obtain a protein fusion of two proteins (protein 1 and protein 2).

First, the forward and reverse primers for each protein have to be designed. The forward primer for the protein 1 and the reverse primer for the protein 2 codify a restriction enzyme site and a fragment that matches with the start or the end of the gene, respectively. The reverse primer of the protein 1 is designed with a portion that coincides with the end of gene 1 and the beginning of the linker. Besides, the forward primer for the protein 2 comprises the start of gene 2 and the linker that is also present in the reverse primer 1. Afterwards, when the PCR reaction is set up, a couple of linear amplifications are done in order to generate a template, and then it is purified. The real amplification is performed using the 5' primer for gene 1 and the 3' primer for gene 2.

M.2.5 DNA plasmid extraction

It has been used a commercial kit (GFX™ Micro Plasmid Prep Kit from GE) designed for the rapid extraction and purification of plasmid DNA from *E. coli*. It is based on a modified alkaline lysis procedure. No toxic organic solvents are used; instead, chaotropic salts denature protein contaminants and promote selective binding of DNA to a glass fibre matrix, prepacked in a column. Isolated DNA is suitable for direct use in PCR amplification, restriction digest analysis, subcloning, transformation and labeling.

M.2.6 Digestion of DNA with restriction endonucleases

Restriction endonucleases recognize short DNA sequences and cleave double-stranded DNA at specific sites within or adjacent to the recognition sequences. Restriction endonuclease cleavage of DNA into discrete fragments is one of the most basic procedures in molecular biology. The activity of these enzymes depends on the pH value, ionic strength and the temperature. Each enzyme has specific optimum conditions for its activity.

In principle, 1 U restriction endonuclease completely digests 1 µg of purified DNA in 60 min using the recommended assay conditions. The volume of restriction endonuclease added should be less than 1/10 the volume of the final reaction mixture, because glycerol present in the enzyme storage buffer may interfere with the reaction.

The enzymes used in this work have been purchased in Roche or New England Biolabs.

M.2.7 Separation and purification of DNA fragments

M.2.7.1 Agarose gel electrophoresis

In this type of electrophoresis, the DNA molecules are separated by size. An electric field forces the fragments to migrate through the gel. Due to the net negative charge of the phosphate backbone of the DNA chain, DNA will migrate to the positive pole. The DNA is forced to move through a sieve of molecular proportions that is made of agarose: larger pieces of DNA move slower than small ones. After the separation is completed, the DNA fragments are visualized using a dye specific for DNA, such as ethidium bromide.

The agarose concentration in the gel will vary depending on the size of the fragments that have to be analyzed (Table M.2)

Table M.2 Agarose concentration depending on the size of DNA fragments to be separated.

% Agarose	Range of separation (kbp)
0.7	0.8 - 10
0.9	0.5 - 7
1.2	0.4 - 6
2.0	0.1 - 2

. Running buffer (10xTAE (Tris/acetate/EDTA) electrophoresis buffer)

Dissolve the following compounds in 1 liter ddH_2O:

48.40 g Tris base

11.42 ml acetic acid

50 ml 0.2 M EDTA, pH 8.0

a) Preparation of the agarose gel:

1) Weigh out 0.5 g of agarose into a 250 ml conical flask. Add 50 ml of TAE if the gel is small. If it is large, add 150 ml of TAE. Swirl to mix.
2) Microwave for about 2 minutes to dissolve the agarose.
3) Leave it to cool for 5 minutes down to about 60°C.
4) Add 1µl of ethidium bromide (10 mg/ml) and swirl to mix.
5) Pour the gel slowly into the tank.
6) Insert the comb.
7) Pour 1x TAE buffer into the gel tank to submerge the gel to 2–5 mm depth.

b) Electrophoresis

1) Prepare the samples by adding an appropriate amount of loading buffer into each tube. Add 0.2 volumes of loading buffer, i.e. 2 µl into a 10 µl sample.

. Loading buffer (10x)

Glycerol 50%

Orange-G dye (small spatula tip)

2) Load the first well with molecular weight markers. In this work different markers have been used depending on size of DNA fragments: DNA molecular weight marker III (size range 0.12-21.2 kbp) and molecular weight marker X (size range 0.07-122 kbp) from Roche.
3) Continue loading the samples in the wells.

4) Close the gel tank, switch on the power-source and run the gel at 5 V/cm.

c) Image acquirement

The images of the agarose gels have been obtained using Geldoc XR Imaging system from Bio-Rad.

M.2.7.2 Purification of DNA fragments from agarose gels

For this purpose, a commercial kit from Qiagen (QIAEX II system) has been used. Purification of DNA fragments with this kit is based on solubilization of agarose and selective adsorption of nucleic acids onto silica-gel particles. In this way, DNA is separated from salts, agarose, polyacrylamide, dyes, proteins, and nucleotides.

1) Excise the DNA band from the agarose gel with a clean, sharp scalpel.
2) Weight the gel slice in a tube. Add the buffer (provided by the kit) following the manufacturer's instructions.
3) Resuspend the silica-gel by vortexing for 30 sec. Add 10 μl the silica-gel to the sample and mix.
4) Incubate at 50ºC for 10 min to solubilize the agarose and bind the DNA.
5) Centrifuge the sample for 30 s and carefully remove the supernatant with a pipet.
6) Wash the pellet.
7) Air-dry the pellet for 10-15 minutes.
8) Add 20 μl of ddH_2O to elute the DNA.
9) Centrifuge for 30 sec. Carefully pipete the supernatant that contains the DNA into a clean tube.

M.2.8 Cloning of DNA fragments into vectors

In this work, to be able to clone a DNA insert into a cloning or expression vector, both were treated with two restriction enzymes that created compatible ends. The procedure comprises the following steps:

1) Digestion of the DNA insert and vector with the appropriate restriction enzymes.
2) Purification of both DNA fragments by agarose electrophoresis.
3) Ligation reaction. This process involves the formation of phosphodiester bonds between adjacent 5'-phosphate and 3'-hydroxyl residues, which can be catalyzed by the bacteriophage T4 DNA ligase. The efficiency of the ligation reaction depends on: the

absolute DNA concentration (minimum 100 ng of the digested vector) and the molar ratio between of insert to vector (2:1). Set up each ligation mixture in the following way:

200 ng of vector DNA
400 ng of insert DNA
10x Ligase buffer
1 μl T4 Ligase

Bring volume to 10 μl with nuclease-free water and incubate the ligation reactions at room temperature for 3 hours (or at 4°C overnight).

M.2.9 Preparation and transformation of *E.coli* cells

M.2.9.1 Preparation of E.coli competent cells

The most common protocol used to prepare *E.coli* competent cells is the calcium chloride method[3]:

1) Pick a single colony from a plate of cells that have been freshly grown for 16–20 h at 37°C, and transfer it to 5 ml of LB medium. Incubate the culture for 16–20 h at 37°C with moderate shaking.

2) Inoculate 1 ml of the culture into 100 ml of LB medium in a sterile 500 ml flask. Grow cells at 37°C for about 3 h at 250 r.p.m. For efficient transformation, it is essential that the number of viable cells do not exceed 10^8 cells/ml.

3) When the OD_{650nm} is 0.6, transfer the cells aseptically to two 50 ml sterile tubes. Leave the tubes on ice for 10 min.

4) Harvest the cells by centrifugation at 2400x*g* for 10 minutes at 4°C.

5) Pour off the supernatant, resuspend each pellet of cells in 25 ml of ice-cold 50 mM $CaCl_2$ and store on ice during 10 minutes.

6) Recover the cells by centrifugation at 2400x*g* for 10 min at 4°C. Discard the supernatant, and resuspend each pellet in 5 ml of an ice-cold solution of 50 mM $CaCl_2$.

7) Cells can be stored at 4°C in a solution of 50 mM $CaCl_2$ for 16–20 h. The efficiency of transformation increases four- to six-fold during this time of storage.

8) Afterwards, centrifuge cells at 2400x*g* for 10 min at 4°C. Discard the supernatant, and resuspend each pellet in 5 ml of 50 mM $CaCl_2$ + 15% glycerol.

9) Use immediately for heat shock transformation, or aliquot 100 μl to 1.5 ml tubes and store at -80° C.

In order to obtain the maximum transformation efficiency the method used to prepare competent cells was the called Inoue method[4]. This protocol differs from other procedures in that the bacterial culture is grown at 18°C rather than the conventional 37°C.

M.2.9.2 Transformation of E.coli by heat shock

1) Thaw the competent cells on ice.
2) Add maximum 20 µl of the ligation mixture or 5 ng of plasmid DNA to each tube.
3) Incubate the tubes on ice for 30 minutes.
4) Heat shock for 2 minutes at 37°C and chill immediately on ice for 5 minutes.
5) Add 950 µl of room temperature media LB and incubate at 37°C for 1 hour.
6) Transfer the cultures to 1.5 ml tubes and spin for 1 min at 6000 r.p.m.
7) Remove 800 µl of the supernantant and resuspend the pellet.
8) Plate out the suspension on LB agar plate containing the appropriate antibiotic.
9) Incubate the plates overnight at 37°C.

M.2.10 Preparation and transformation of yeast competent cells

M.2.10.1 Preparation of yeast competent cells[5]

1) Inoculate the yeast strain into 25 ml of liquid 2x YPD medium and incubate overnight at 200 r.p.m. and 30°C.

2) After 12–16 h of growth, determine the titer of the yeast culture. This can be done using a spectrophotometer. Pipette 10 µl of cells into 1 ml of water in a spectrophotometer cuvette, mix thoroughly by inversion and measure the OD at 600 nm (a suspension containing $1x10^6$ cells/ml will give an OD_{600nm} of 0.1).

3) Add $2.5x10^9$ cells to 500 ml of 2x YPD in a culture flask. The titer of this solution should be $5x10^6$ cells/ml.

4) Incubate the flask in the shaking incubator at 30°C and 200 r.p.m. until the cell titer is at least $2x10^7$ cells/ml. This should take about 4 h.

5) Harvest the cells by centrifugation at 3000x*g* for 5 min, wash the cells in 0.5 volumes of sterile water, resuspend in 0.01 volumes of sterile water, transfer to a suitable sterile centrifuge tube and centrifuge the cells at 3000x*g* for 5 min at 20°C.

6) Resuspend the cell pellet in 0.01 volumes of filter sterile frozen competent cell (FCC) solution (5% v/v glycerol, 10% v/v DMSO). Use good quality sterile DMSO.

7) Dispense 50 μl samples into an appropriate number of 1.5 ml tubes.

8) Place the tubes into a 100 tubes styrofoam rack with lid. It is best to place this container upright in a larger box (Styrofoam or cardboard) with additional insulation such as Styrofoam chips or newspaper to reduce the air space around the sample box. This will result in the samples freezing slowly, which is essential for good survival rates.

9) Put the large Styrofoam container in a -80°C freezer overnight. The Styrofoam rack containing the frozen yeast cells can then be removed from the freezing container and stored at -80°C.

M.2.10.2 Transformation of yeast competent cells

1) Thaw cell samples in a 37°C water bath for 15–30 s.

2) Centrifuge at 13000x*g* for 2 min and remove the supernatant.

3) Make up frozen competent cell (FCC) transformation mix for the planned number of transformations plus one extra. Include an extra tube for a negative control tube for no plasmid DNA. Add this to the pellet and vortex mix vigorously to resuspend the cell pellet.

Table M.3 Components of the transformation mix solution for yeast.

Transformation mix components	Volume (ml)
PEG 3350 (50% (w/v))	260
LiAc	36
Single-stranded carrier DNA (2mg/ml)	50
Plasmid DNA plus sterile water	14
Total volume	360

4) Add 70 μl of DMSO. Mix well by gentle inversion or swirling.

5) Incubate at 30°C for 30 min with shaking (200 r.p.m).

6) Heat shock at 42°C in a water bath for 15 min.

7) Chill cells on ice for 1–2 min.

8) Centrifuge the tubes at 13000x*g* for 30 s and remove the supernatant.

9) Pipette 1.0 ml of sterile water into the transformation tube.

10) Plate and spread 200 μl of the cell suspension onto the appropriate SC selection medium. Cells should be plated less densely when possible because plating density negatively affects transformation efficiency.

11) Incubate the plates at 30°C for 3–4 days and recover the transformants.

M.2.11 Protein recombinant production

M.2.11.1 Protein production in bacteria

All the plasmids used in this work have an expression system based on the T7 RNA polymerase promoter using IPTG as inducer.

1) Transform *E. coli* BL21(D3) competent cells with the plasmids containing the desired protein and plate them on appropriate selective LB plates.

2) Grow overnight at 37°C.

3) Next day, pick a single colony from the LB agar plate in order to set up an overnight culture into 1 ml liquid LB containing appropriate antibiotics

4) Incubate at 37°C overnight at 250 r.p.m.

5) In the morning, centrifuge 1 ml of each culture for 5 min at 1300x*g* and remove the supernatant. Resuspend the cell pellet in fresh LB (1 ml).

6) Dilute the culture 1:100 in LB containing appropriate antibiotics. For example, one aliquot of 50 μl should be diluted to a final volume of 5 ml.

7) Incubate the culture at 37°C and 250 r.p.m. until the A_{600nm} = 0.6.

8) Induce protein expression by adding IPTG to a final concentration of 1 mM.

9) Incubate the culture at the appropriate temperature and 250 r.p.m. overnight. The usual temperature is 37°C but if the expressed protein have a tendency to aggregate (like the protein fusions related to BIFC method) cells are grown at 18°C. The following day, a SDS-PAGE can be performed to ensure the protein production.

M.2.11.2 Protein production in yeast

1) Transform *S. cerevisiae* competent cells and plate them on the appropriate SC minimum media Drop Out plate. The used Drop Out media will always depend on the plasmid used for yeast transformation.

2) After three days, pick one transformed colony and set up a culture into appropriate SC minimum Drop Out media using raffinose (final concentration 2%) as carbon source.

3) Incubate at 30°C during 24 hours.

4) Harvest 1 ml of the cell culture and wash it three times with the SC media.

5) Dilute the culture 1:150 in SC minimum Drop Out media using galactose (final concentration 2%) as carbon source.

6) Incubate the culture at 30°C for approximately 24 hours.

7) The cells are then ready for further biochemical analysis.

M.3 Protein analysis methods

M.3.1 SDS polyacrilamide gel electrophoresis (SDS-PAGE)

SDS-PAGE stands for Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) and it is a method used to separate proteins according to their molecular weights. SDS, an anionic detergent, is used to bring proteins to their primary (linearised) structure and coat them with uniform negative charges.

There are two gels, the resolving and the stacking gel. The stacking gel is of very low acrylamide concentration and it is used to form the wells into which the sample is loaded. The low acrylamide concentration also allows most proteins to be concentrated at the dye front, so that samples with different protein concentration could be compared.

a) Preparation of the polyacrilamide gel

In the following table, there are indicated the different compositions of the stacking and resolving gel depending on the polyacrilamide concentration.

Table M.4 Components of stacking and resolving gel

Resolving gel[a]			**Stacking gel**[a]	
Components	12,5%	15%	Components	4%
Polyacrilamide[b]	2.5 ml	3 ml	Polyacrilamide[b]	0.4 ml
Tris-HCl 1 M (pH 8,0)	3 ml	3 ml	Tris-HCl 1 M (pH 8,0)	0.5 ml
SDS 10%	80 ml	80 ml	SDS 10%	40 ml
ddH_2O	2.36 ml	1.86 ml	ddH_2O	3.03ml
TMED	12 ml	12 ml	TMED	6 ml
Ammonium persulfate	40 ml	40 ml	Ammonium persulfate	20 ml

[a]Volumes for one 1.5 mm gel

[b]40% Acrilamide/bisacrilamide 37:5:1 (Bio-Rad)

b) Electrophoresis

1) Place gel in holder/electrode, and then transfer to running tank. Fill with 1x running buffer (keep inside and outside buffer chambers separated).

2) Mixing protein samples with sample loading buffer and heat the mixture at 100°C during 10 minutes.

Table M.5 Sample loading buffer

Loading buffer (6x)	
Tris-HCl 1 M (pH 6,8)	1.75 ml
SDS	0.5 g
Glycerol	1.5 ml
ddH_2O	1.5 ml
Bromophenol blue	2.5 mg
2-Mercaptoethanol	0.25 ml

Table M.6 Electrophoresis running buffer

Electrophoresis buffer	
Tris	30.25 g
SDS	10 g
Glycine	144 g
ddH_2O	until 1 l

3) Load the samples on the gel. In one of the wells, also load a protein standard ladder (i.e. BenchMark protein ladder or SeeBlue Prestained of Invitrogen).

4) Run at 100V-200 V until the dye front reaches the bottom of the gel.

c) Gel staining

1) After the electrophoresis, stain the gel with Coomassie Blue (see Table M.7) at room temperature during 30 minutes with slow agitation.

2) Wash the gel with the destaining solution during 2-3 hours with slow agitation.

Table M.7 Staining and destaining solution

Stain solution	
Coomassie Blue	0.15%
Methanol	30%
Acetic Acid	8%

Destain solution	
Acetic acid	10%
Methanol	50%

f) Image acquirement.

The images have been obtained using Versadoc Imaging system from Bio-Rad.

M.3.2 Blotting methods

M.3.2.1 Western Blotting

Western Blotting or immunoblotting allows determining, with a specific primary antibody, the relative amounts of the protein present in different samples.

a) Sample preparation

Samples are prepared from tissues or cells that are homogenized in a buffer that protects the protein of interest from degradation. Proteolysis remains the major concern, as many methods of solubilization release intracellular proteases that can degrade the target polypeptide. Steps to avoid degradation include keeping the sample at 0°C and/or adding protease inhibitors to the lysis buffer. A mixture of the most widely used protease inhibitors are commercially available.

The preparation of cell extracts depends on the type of samples: from bacteria or yeast cells.

M.3.2.1.1 Bacterial samples

In bacterial samples, the procedure is very simple because cells can be disrupted by sonication.

1) Incubate the cells minimum 16 hours after IPTG addition.
2) Harvest cells by centrifugation and resuspend the pellet in extraction buffer.

Table M.8 Extraction buffer for bacteria

Extraction buffer components	Concentration
Tris-HCl, pH 8.0	50 mM
NaCl	50 mM
Tween-20	0.05%
DTT	5 mM
PMSF	1 mM

3) Afterwards, sonicate the cells with 10 short burst of 10 sec followed by intervals of 30 sec for cooling. It is important to keep the suspension at all times on ice.
4) Remove cell debris by ultracentrifugation at 4°C for 30 min at 13000x*g*.

M.3.2.1.2 Yeast samples

The preparation of yeast extracts presents the added challenge of having to disrupt a particularly resistant cell wall. Its removal by mechanical disruption with glass beads is very effective.

1) After induction of the protein expression, grow yeast until an OD_{600} of 2-3.
2) Harvest the cells by centrifugation and resuspend the pellet in extraction buffer.

Table M.9 Extraction buffer for yeast

Extraction buffer components	Concentration
Tris-HCl, pH 8.0	10 mM
NaCl	150 mM
Tween-20	0.05%
Glycerol	10%
EDTA	5 mM
DTT	1 mM
Anti-protease cocktail for yeast	

3) Afterwards, cells are lysed by vortexing with glass beds.

4) Remove cell debris by ultracentrifugation at 4°C for 30 min at 13000x*g*.

b) Determination of sample concentration

In order to load the same quantity of proteins in each gel, the protein content in each sample was quantified using the Lowry assay[6].

1) Prepare the following stock solutions:

Table M.10 Stock solutions for Lowry assay

Solution	Components
Lowry A	2% Na_2CO_3 in 0.1 M NaOH
Lowry B	1% $CuSO_4$ in ddH_2O
Lowry C	2% sodium potassium tartrate

Table M.11 Lowry stock reagent composition

Solution	Volume
Lowry A	49 ml
Lowry B	0.5 ml
Lowry C	0.5 ml

2) Prepare a series of dilutions of bovine serum albumin in the same buffer containing the samples, to give concentrations from 30 to 150 μg/ml.

3) Prepare the Lowry solution by mixing the solutions as indicated in Table M.10/11.

4) Dilute the samples if it is necessary and work in triplicates.

5) Mix 500 μl of your sample/standard with 700 μl of the Lowry solution.

6) Vortex briefly to mix and incubate for 20 min at room temperature.

7) Prepare the Folin's Reagent Phenol reagent: 2N (Folin - Ciocalteau reagent) diluted 1:1 in ddH_2O.

8) Incubate 30 min in the dark at room temperature.

9) Measure the absorbance at 750 nm and calculate the protein content of the samples from the calibration curve.

c) SDS-PAGE electrophoresis

Follow the instructions of the previous section.

d) Transfer of proteins from gel to membrane

1) Immerse the gel in an appropriate transfer buffer and allow it to equilibrate for 10–15 minutes.

Table M.12 Transfer buffer composition

Components	Concentration
Glycine	39 mM
Tris-HCl	48 mM
SDS	0.04%
Methanol	20%
Adjust to pH 9.0	

2) If working with a PVDF membrane, wet the membrane in 100% methanol for 15 seconds.

3) Equilibrate the membrane for at least 5 minutes in the transfer buffer and soak filter paper in the transfer buffer for at least 30 seconds.

4) Assemble the transfer stack. To ensure an even transfer, remove air bubbles by carefully rolling a clean pipette over the surface of each layer in the stack.

5) Transfer proteins according to blotting apparatus manufacturer's instructions.

6) Remove the blot from the transfer system and briefly rinse the membrane in ddH_2O to remove gel debris.

d) Immunodetection

This type of protein identification uses a specific antibody to detect and localize a protein blotted. The specificity-antigen antibody binding permits the identification of a single protein in a complex sample.

1) Rinse the blot with water and then place the blot in blocking buffer and incubate for 1 hour with gentle agitation at room temperature. Blocking buffer is a solution of 1.5% (w/v) blocking agent (i.e. non fat dry milk) in wash buffer (see Table M.13)

Table M.13 Wash buffer composition

Components	Concentration
Tris-HCl	10 mM
NaCl	150 mM
Tween-20	0.1%

2) Prepare primary antibody solution by diluting the antibody in blocking buffer.

3) Place the blot in the diluted primary antibody solution and incubate for at least 1 hour with gentle agitation at room temperature. Ensure that the solution moves freely across the entire surface of the membrane.

4) Wash the blot with fresh wash buffer three times with gentle agitation for 5–10 minutes.

5) Prepare secondary antibody solution by diluting the antibody in blocking buffer.

6) Place the blot in the diluted secondary antibody solution, and incubate for 1 hour with gentle agitation.

7) Wash the blot with wash buffer three times with gentle agitation for 5–10 minutes.

e) Chemiluminescent detection

It uses an enzyme to catalyze a reaction producing visible light.

1) Prepare the solution of the chemilumiscent substrate (i.e. Immobilon Western HRP substrate from Millipore) following manufacturer's instructions.

2) Place the blot protein side up in a container and add the substrate onto the blot.

3) Incubate the blot for 5 minutes at room temperature.

4) Drain the excess substrate.

5) Cover the blot with a clean plastic wrap and remove any air bubbles.

f) Image acquisition

The image is acquired directly in a digital imaging system (i.e. Versadoc imaging system from Bio-Rad).

M.3.2.2 Filter trap assays.

This type of blotting is a common method to detect aggregated proteins. The proteins are applied onto a membrane. A dissolved sample is filtered through the membrane by applying vacuum. Proteins are adsorbed to the membrane and the other sample components are pulled through by vacuum.

1) The pellet fraction (13000x*g* for 10 min) of the disrupted yeast is resuspended in PBS with a protease inhibitor.

2) For cell lysis and DNA cutting, it is treated with 2% SDS at room temperature during 15 min and sonicated for 10 s.

3) The sample is filtered through a cellulose acetate membrane (Osmonics, 0.22-μm pore size) using a dot blot filtration unit.

4) The SDS-insoluble aggregates retained on the filters are detected by incubation with the appropriate antibody following the instructions given in the previous section.

M.4 Imaging and measuring cell fluorescence

The first step is the elimination of the growth media. It is essential to remove LB or SC minimum media because they could cause auto-fluorescence in subsequent steps. The day after the protein expression, centrifuge the cells at 1.300x*g* for 5 min and remove the supernatant. Wash the cells by resuspending in two volumes of 50 mM Tris–HCl pH=7, followed by centrifugation (1300x*g* for 5 min). Carry out three washes in total.

After the final centrifugation, cells can be resuspended in one volume of PBS and used immediately for the imaging. Alternatively, the cells can be fixed as described in the following box and analyzed at a later date.

BOX 1 | Fixing cells for analysis

This step ensures a constant fluorescence signal during analysis

1) Resuspend the cells in an equal volume of 0.1% (v/v) formaldehyde, incubate at room temperature for 15 min.

2) Centrifuge at 1300x*g* for 5 min and discard the supernatant.

3) Resuspend the cellular pellet in an equal volume of PBS or Tris–HCl, pH=7.

Fixed cells can be stored at 4°C until analysis. The maximum recommended storage period is 5 days. After this time, fluorescence signal starts to decrease.

Cells can be analyzed by microscopy and/or fluorescence spectrophotometry.

M.4.1 Microscopic analysis

1) Deposit a drop of cell solution on a microscope slide and allow it to dry.
2) Place two to three drops of Fluoprep reagent over the dried preparation.
3) Immediately cover the sample with the cover slip. Gently press the cover slip to remove air bubbles and excess of mounting medium. The cover slip will be fixed firmly within minutes.
4) Examine the slide under the microscope. Microscopy images can be taken with a fluorescence microscope equipped with a camera (or similar equipment) at a magnification of x160. If YFP or GFP is used as a reporter protein, then a GFP excitation filter (450–495 nm) and a GFP emission filter (515–560 nm) are both required.

M.4.2 Spectrophotometric analysis

1) Dilute freshly washed and fixed or unfixed cells to A_{600nm} = 0.03 in Tris–HCl 50 mM, pH=7.0

2) Set up a fluorescence spectrophotometer to excite at 480 nm and record the fluorescence emission spectra in the range 510–600 nm if the fluorescent protein is YFP. The optimal values of parameters such us excitation/emission, slit width, gain or recording speed should be determined experimentally because they depend both on the samples and on the particular spectrophotometer. Emission wavelength and emission spectra range are also dependent on the fluorescent proteins used.

3) Take measurements for each sample, in triplicate, at 25°C with continuous stirring. Usually the minimum required sample volume is 800 μl. Blank the system with Tris–HCl 50 mM before each reading.

Alternatively, only the fluorescence of the intracellular media can be analyzed. In this case, cells are lysed using Novagen Bug Buster protein extraction reagent with benzonase nuclease and protease inhibitory cocktail set III (Calbiochem). Insoluble proteins are afterwards removed by centrifugation, and the fluorescence of the soluble part is measured.

M.5 Flow cytometry and cell sorting experiments

Flow cytometry is a powerful method that allows the analysis of entire cell populations based on the characteristics of single cells flowing through an optical and/or electronic detection device. Modern flow cytometers are able to analyze thousands of cells every second in 'real time' and, when coupled to cell sorters, can actively separate and isolate cells with specific properties. Multiple characteristics including cell count, cell size or response to fluorescent probes reporting different cellular functions may be collected simultaneously by this method. In addition, the ability to analyze large number of cells averages experimental variability and provides more consistent and reproducible results.

M.5.1 Flow cytometry analysis

A first step to fixate the cells is always advisable because it ensures a constant fluorescence signal during analysis and prevents contamination of the flow cytometer owing to bacterial growth. However, do not fix the cells if cell sorting experiments are to be performed afterwards, as this will abolish bacterial growth and consequently will prevent plasmid isolation and sequencing.

1) Dilute the cells in PBS (the most commonly used suspension buffer) to achieve an optimal cell density. Usually cells must be resuspended at a density of 10^5–10^7 cells/ml to prevent the narrow bores of the flow cytometer and its tubing from clogging up. In our case, working with *E. coli*, $OD_{600nm} = 0.05$ is usually appropriate for analysis.

It is important to filtrate the PBS using a 0.22-μm filter.

2) Establish flow cytometer parameters by comparing the light scattering and fluorescence properties of *E. coli* expressing the fluorescent protein, GFP or YFP (positive control, exhibiting fluorescence) and *E. coli* exhibiting no fluorescence (negative control). First analyze the negative control. Place these bacteria in a sample tube. Establish the event rate to 1000 events per second to minimize coincidence and improve the resolution of the cell population. High events rates can be corrected either by dilution or by decreasing the instrument flow rate.

3) Three parameters -forward scatter (FSC), sideward scatter (SSC) and green fluorescence (FL1, fluorescent light (FL) used for green fluorescent detection)- are monitored during flow cytometer analysis. Set photomultipliers (PTMs) of FSC and SSC to logarithmic amplification to ensure a recognizable population for gating on scale. Gating is an important procedure in FC to selectively visualize the cells of interest while

eliminating results from unwanted particles (e.g. cell debris or dead cells). Cells could be gated according to physical characteristics; subcellular debris and clumps can be differentiated from cells by size, estimated by FSC.

4) Generate a dot plot of SSC versus FSC. Adjust PTMs voltatge to place the bulk of recorded light scattering particles in the centre of SSC/FSC dot plot. Use the SSC channel as the trigger for data collection.

5) Accumulate cells and set a gate around the bacterial population. For fluorescence measurements, the PMT voltage for the FL1 detector has to be adjusted with no fluorescent *E. coli* cells, so that the entire bacterial population is present within the first decade of the logarithmic scale for this fluorescence channel. For this purpose, generate a histogram representing log FL1 fluorescence on the x-axis and the number of events (cell count) on the y-axis to analyze fluorescence in the bacterial population. Set this negative fluorescent signal in the first logarithmic decade by changing FL1 PMT voltage. Acquire information for 20000 cells of the specified gate. The flow cytometer settings used for defining the bacterial population in this study are described below.

Table M.14 Flow cytometer settings

Thresold (F1)	Detector setting		
416	FSC	SSC	FL1
	E02*	447	645

*E02 is an adjustment of the FSC in FACSCalibur that multiplies FSC signal by 100

6) After acquiring the all measurements, remove residual bacteria from previous samples by back-flushing the sample tubing and running sheath fluid through the sample lines for 15 sec. This step prevents cross-contamination between bacterial samples and is particularly important if cell sorting is to be carried out.

7) Next, analyze positive control cells expressing the fluorescent protein. Use the same procedure and cytometer settings outlined before. At this point, a histogram marker can be placed around the positive control to designate positive events and differentiate them from negative events using the CellQuest Pro software or equivalent suite. Statistical information for that region is provided by this program, for example, mean fluorescence.

8) Finally, analyze the fluorescence of cells displaying some degree of fluorescence. Once the histogram has been obtained, various markers can be created to designate the different populations detected and statistical information is provided for each one.

M.5.2 Cell sorting experiments

If a fluorescence-activated cell sorting (FACS) device is available, cell subpopulations exhibiting increased mean fluorescence emission in a mixture of different cell populations can be electronically deflected into separate collection tubes and sorted at high purity using slow sorting rates.

1) Check the sterility and the alignment of the cytometer and prepare the cell sorter using autoclaved sheath fluid and 10-μm fluorescent calibration beads. The sorter must be pre-sterilized with bleach, then 70% ethanol. Check that the observation point is not obscured. Afterward, verify that side streams are not split and adjust vibration phase. Use beads to calibrate the cell sorter. Afterwards, sterilize the sample fluid path by running through it 1% bleach for 5 min, followed by 70% ethanol for 5 min to wash away the bleach; then sterile sheath fluid for 5 min to wash away the ethanol. Between each fluid change, place the machine on run mode and let several drops pass into the waste to remove residual bleach or alcohol.

It is important to select a sheath fluid with a composition that enables further biochemical assays (azide-free) and also provides sufficient ions to hold a droplet charge during the sorting process.

2) Run the bacterial sample on the flow cytometer. On the resulting histogram (counts versus FL1), define the sorting gates (populations with high fluorescence signal that you want to isolate). If the sorted population represents a small percentage of the total cell population recheck the optical alignment of the instrument.

3) Check the efficiency of sorting by reanalyzing a portion of the sorted sample. The sorted population should have the same mean fluorescence value as the original subpopulation but a significantly increased percentage of fluorescent cells.

4) To grow the sorted bacteria for further analysis, use an inoculating loop to streak an LB plate (with the appropriate antibiotics) with the sorted bacteria.

M.6 References

1. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60-89 (1990).
2. Peranen, J., Rikkonen, M., Hyvonen, M. & Kaariainen, L. T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. *Anal Biochem* **236**, 371-373 (1996).
3. Mandel, M. & Higa, A. Calcium-dependent bacteriophage DNA infection. *J Mol Biol* **53**, 159-162 (1970).
4. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96**, 23-28 (1990).
5. Gietz, R.D. & Schiestl, R.H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**, 1-4 (2007).
6. Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**, 265-275 (1951).

PART 1

CHAPTER 1

Detection of Weak Protein-Protein Interactions by Bimolecular Fluorescence Complementation (BIFC)

1.1 Introduction

During the past decade, the complete sequencing of genomes has provided us with a comprehensive inventory of predicted proteins for many different species. Life sciences face now the decipherment of the structure and functions of the encoded proteins.

One of the most useful approaches to understand the function of a specific protein is the identification of the molecules that interact with it. It has to be taken into account that proteins almost never act in an isolated manner: they interact with other proteins in order to perform essential roles in many important cellular processes. Apart from their ability to form stable multiprotein complexes, proteins associate transiently with their targets to modify, regulate by steric effects, or translocate them to different cellular compartments. Moreover, since biological processes are orchestrated and regulated by dynamic signalling networks of interacting proteins, the analysis of protein complexes and protein-protein interaction networks –and the dynamic behavior of these networks as a function of time and cell state- are of central importance in biological research. Furthermore, this task is highly challenging because of the diverse physico-chemical properties of proteins and the very different characteristics of protein-protein interactions (for example, equilibrium dissociation constants comprising several orders of magnitude; the different abundance of proteins in the cell and in its compartments, etc.)

1.1.1 Methods to study protein-protein interactions

Different methods have been developed to study protein interactions. In fact, protein binding can be analyzed from many perspectives (Figure 1.1). In order to discover the interaction partners of one protein, the required method should enable the screening of large numbers of candidates. Ideally, the system should operate *in vivo* to maintain the cellular context where bindings occur. In addition, once a specific interaction has been identified, the molecular and biophysical properties of the complex need to be characterized. Key aspects such as the oligomeric state of the interaction partners, the stoichiometric ratio in the complex, the affinity of the partners for each other or the kinetic rate constants should be addressed. In the following section the main techniques for identifying and characterizing protein interactions *in vivo* and *in vitro* will be reviewed.

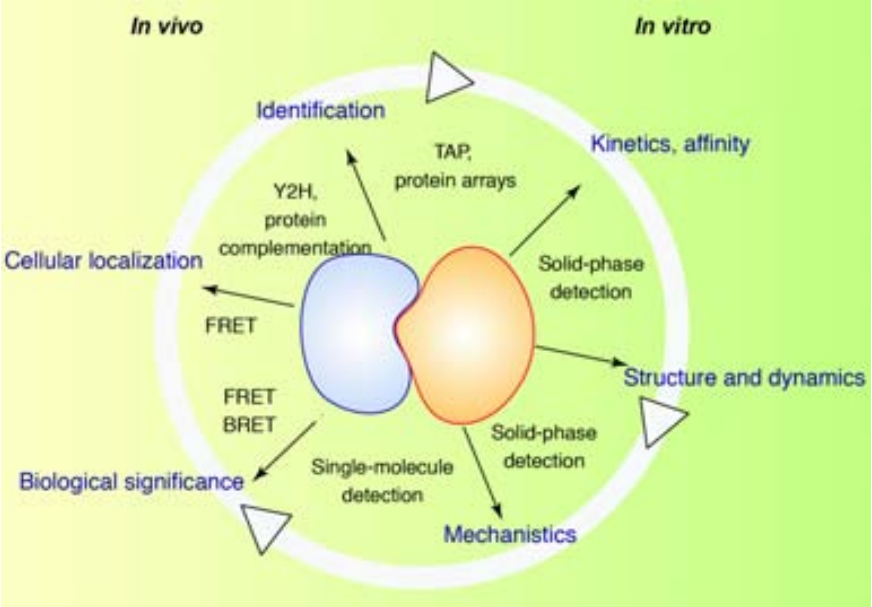


Figure 1.1 Different levels of characterization of protein-protein interactions *in vivo* and *in vitro*.[1]

1.1.1.1 Methods to study protein interactions in vitro

1.1.1.1.1 Bimolecular display technologies.

Display technology refers to a collection of methods for creating libraries of modularly coded biomolecules that can be screened for desired properties. The combination of an ever-increasing variety of libraries of modularly coded protein complexes together with the development of innovative approaches to select a wide array of desired properties has facilitated large-scale analyses of protein interactions.

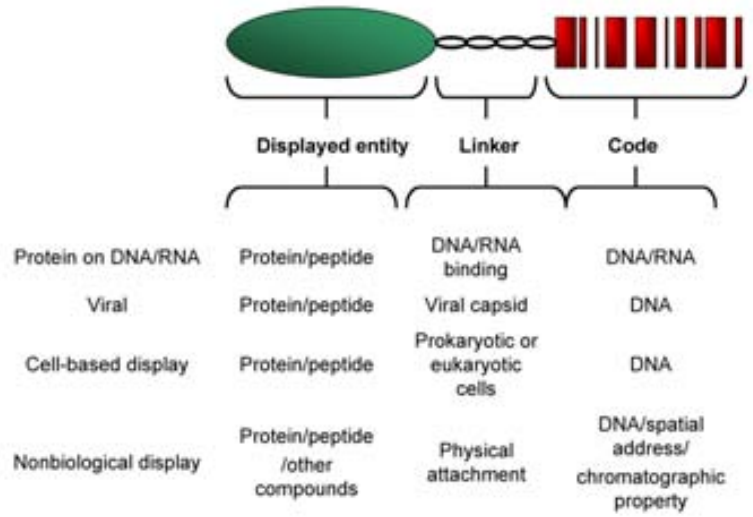


Figure 1.2 Schematic diagram of a typical display module and a list of four major display systems[2].

Nowadays, display technologies have comprised two major groups: biological display systems that employ biological host/biological reactions; and non-biological display systems that use chemical and engineering techniques[2]. Regardless of the format, a display library consists of modularly coded molecules, each of which contains three components: displayed entities, a common linker, and the corresponding individualized codes (Figure 1.2). These technologies use different types of displayed entities, linkage formats and coding strategies. One of the most used coding format is the viral (o phage) display (Figure 1.3)[3]. Specifically, it is the expression of peptides, proteins or antibody fragments at the surface of bacterial viruses (phage). This is accomplished by the incorporation of the nucleotide sequence encoding the protein to be displayed into a phage or phagemid genome as a fusion to a gene encoding a phage coat protein. This fusion ensures that as phage particles are assembled, the protein to be displayed is presented at the surface of the mature phage, while the sequence encoding it is contained within the same phage particle.

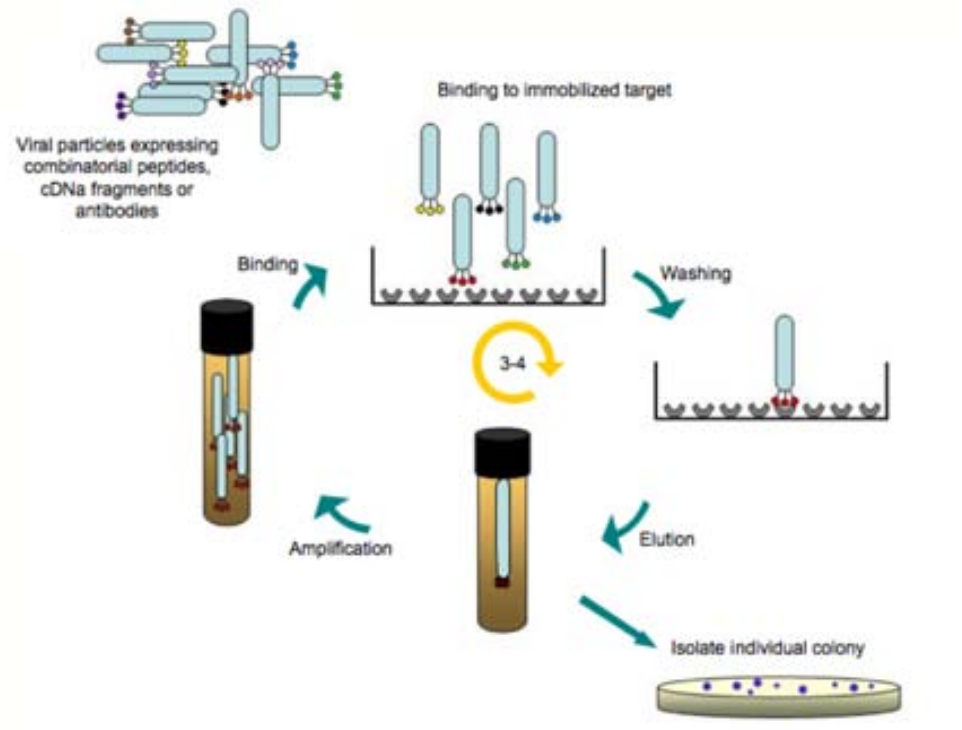


Figure 1.3 Phage display selection experiment: phage libraries are conveniently screened by isolating viral particles that bind to targets, plaque-purifying the recovered phage, and sequencing the phage DNA inserts. Usually, three or four rounds of affinity selection are sufficient to isolate binding phage.

This physical link between the phenotype and genotype of the expressed protein and the replicative capacity of phage are the structural elements that underpin all phage display technology. To achieve desired display advantages, several different viral systems

have been used to display peptides, including lysogenic filamentous phages[3], lytic lambda phage[4], T7 bacteriophage and T4 bacteriophage[5].

Phage display presents several advantages for the study of protein-protein interactions like the diversity of variant proteins that can be represented (phage display antibody libraries with diversities as high as 10^{10} are routinely constructed)[6] or the sensitivity of amplification. Due to the use of panning cycles, the final partners that are selected have a strong affinity. The limitations of this method include its *in vitro* applicability, system-specific negative selections, incompatibility with the expression host system or restricted scope of selection parameters.

1.1.1.1.2 Solid phase detection

These techniques detect the interaction of a soluble ligand with a receptor immobilized on the surface of a physicochemical transducer. In recent years, label-free solid-phase detection has been remarkably enhanced due to its combination with other techniques such as mass spectrometry[7], surface sensitive fluorescent detection[8] or total internal reflection fluorescence (TIRF) spectroscopy[9].

Surface Plasmon Resonance (SPR) is one of optical detection techniques explored and optimized with respect to bimolecular interactions analysis. In this technique, ligands are first immobilized on a sensor chip surface using appropriate coupling chemistries. Analytes are subsequently bound to the immobilized ligands by injection into microfluidic flow cell.

The SPR optical unit consists of a source for a light beam that passes through a prism and strikes the surface of a flow cell at an angle, such that the beam is totally reflected (Figure 1.4). Under these conditions, an electromagnetic component of the beam propagates into the aqueous layer and can interact with mobile electrons in the gold film at the surface of the glass. At a particular wavelength and incident angle, a surface plasmon wave of excited electrons (the plasmon resonance) is produced at the gold layer and is detected as a reduced intensity of the reflected light beam.

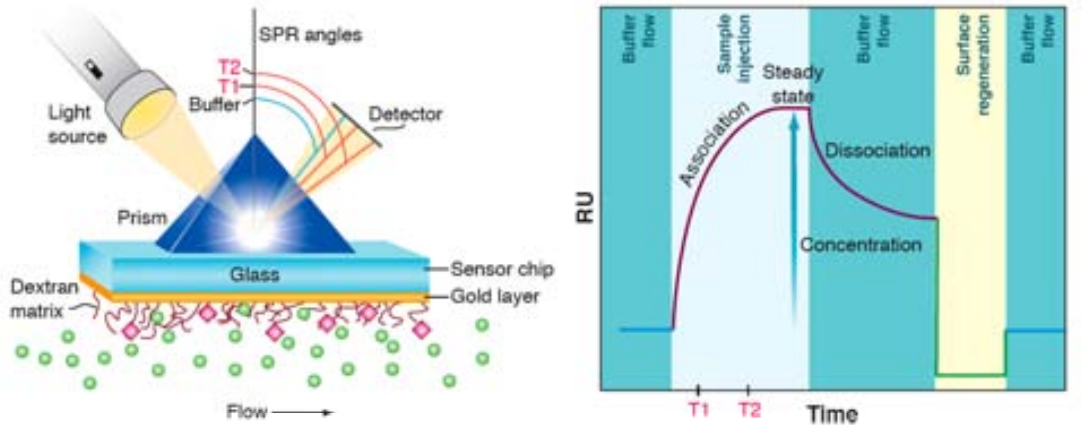


Figure 1.4 Surface plasmon resonance. At left, an SPR optical unit and a sensor chip detect the analytes (green spheres) in the flow solution, which passes by the ligands (pink diamonds) linked to the dextran matrix. The blue SPR angle defines the position of the reduced-intensity beam. Time points T1 and T2, shown in the schematic sensorgram (right), correspond to the two red SPR angles, which shift as analytes binds to ligands over time. The complex dissociates upon reintroduction of the buffer[10].

The SPR angle is sensitive to the composition of the layer at the surface of the gold[11]. A baseline SPR angle is first determined by washing buffer over the surface with a fixed amount of ligand attached (blue angle). Then, some analyte is added to this flow of buffer. The binding of the analyte to the ligand causes a change in the SPR angle because it is directly proportional to the amount of bound analyte. One important advantage of these techniques is that both equilibrium and interaction kinetics can be analyzed because protein interactions are detected in real time.

1.1.1.1.3 Protein arrays

The concept of the protein arrays is the obvious continuation of the DNA array approach. A defined set of proteins are immobilized onto solid supports in a spatially resolved manner and analyzed at high density[12]. They can be classified into two categories depending on the spotted proteins: protein profiling arrays and functional protein arrays. Protein profiling arrays usually consist of multiple antibodies printed on glass slides and are used to measure protein abundance and/or alterations. Protein functional microarrays can be made up of any type of protein and have diverse applications such as screening for biochemical activities or analyzing interactions with proteins, lipids, nucleic acids or small molecules (Figure 1.5).

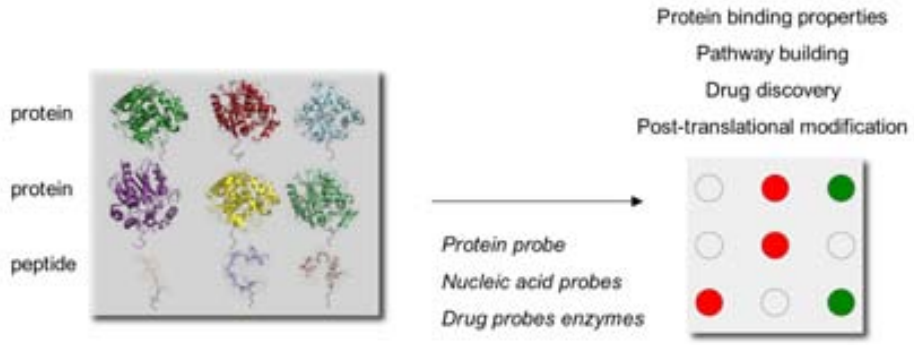


Figure 1.5 Functional protein arrays. Native proteins or peptides are individually purified or synthesized using high-throughput approaches and arrayed onto a suitable surface. These chips are used to analyze protein activities, binding properties and post-translational modifications.

One of the most challenging aspects of this technology is the functional immobilization of a large number of diverse proteins with different physico-chemical properties. A variety of surfaces and immobilization chemistries can be used in the printing of protein microarrays. Glass slides have been widely used because they have low fluorescence background and are compatible with most assays[13-15]. To attach the proteins, the glass surface has to be modified to achieve the maximum binding capacity. One method is the coating with a nitrocellulose membrane or poly-L-lysine. In this case, the proteins can be passively absorbed to the modified surface through non-specific interactions[16]. Nevertheless, the noise level is usually higher due to nonspecific adsorption/absorption. To achieve more specific protein attachment, reactive surfaces on glass that covalently cross-link proteins have been created[17, 18]. However, it is plausible that the proteins are attached to the surface in a random fashion, which may alter their native conformation reducing its activity or hindering its accessibility to probes.

Perhaps the best mean is through highly specific affinity protein interactions[19]. Proteins fused to a high affinity tag at their amino or carboxy terminus are linked to the chip surface; hence, all the proteins should orient uniformly away from the surface. Using this method, immobilized proteins are more likely to remain in their native conformation, while analytes have easier access to active sites of proteins.

Typically, the read-out of protein arrays is indirect and uses fluorescence probes due to their sensitivity, stability and availability of fluorescent scanners tailored for microarray use[13, 19] (Figure 1.6).

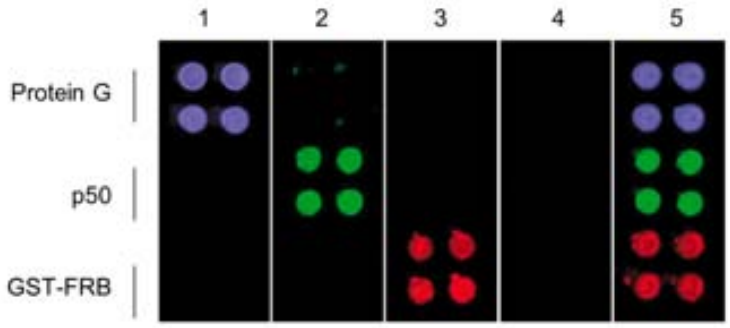


Figure 1.6 Detection of protein–protein interactions on a model protein function microarray[12]. Three different proteins were spotted in quadruplicate on each of five glass microscope slides. 1) Slide probed with IgG labeled with BODIPY-FL (a fluorophore); blue spots indicate an interaction of IgG with immobilized protein G. 2) Slide probed with Cy3-labeled IκBα; green spots indicate an interaction of IκBα with immobilized p50. 3-4) Slides probed with Cy5-labeled FKBP12 in the presence (3) or absence (4) of 100 nM rapamycin; red spots indicate the rapamycin-dependent interaction of FKBP12 with immobilized FRB. 5) Slide probed with a mixture of all three labeled proteins and 100 nM rapamycin.

An advantage of using arrays is the control of the experiment conditions: pH, temperature, ionic strength or the presence/absence of cofactors. In this sense, protein function microarrays will benefit from improved methods of fabrication, processing and analysis; but at the present, the great obstacle is the production of large collections of pure, recombinant proteins.

1.1.1.1.4 Single-molecule detection techniques

During the past decade a development in methodologies for studying interactions on the single-molecule level has occurred. There are some distinct advantages of these techniques over conventional ensemble methods: interaction assays can be carried out with very low quantities and protein concentrations, which often correspond to the natural cellular level; and the dynamics of interactions at equilibrium can be studied. Currently, two main approaches for studying protein interactions on the single-molecule level have been followed: atomic force microscopy (AFM) and fluorescence techniques.

AFM has opened a new window towards characterizing the forces involved in protein interaction using force spectroscopy[20].

Several fluorescence detection approaches have been successfully applied to probing protein interactions on the single-molecule level[21]. Among them, fluorescence correlation spectroscopy is a powerful method for studying the interaction dynamics of protein complexes in solution. Specifically, interactions can be detected as changes in diffusion properties[22].

A robust approach for studying protein interactions is achieved by analyzing fluorescence cross-correlation of proteins labeled with different fluorophores[23, 24]. However, a serious limitation of fluorescence-based single-molecule imaging is the relatively fast photobleaching of organic fluorophores, which typically limits the observation time for an individual fluorophore to a few seconds.

1.1.1.2 Methods to study protein interactions in vivo

1.1.1.2.1 Yeast-two hybrid system

The yeast two-hybrid assay[25] provides a genetic approach to the identification and analysis of protein–protein interactions. It relies on the modular nature of many eukaryotic transcription factors, which contain both a site-specific DNA-binding domain (BD) and a transcriptional-activation domain (AD) that recruits the transcriptional machinery. In this assay, hybrid proteins are generated that fuse a protein X to the BD domain and protein Y to the AD domain of a transcription factor. Interaction between X and Y reconstitutes the activity of the transcription factor and leads to expression of reporter genes (Figure 1.7). In the typical practice of this method, a protein of interest is fused to the DNA-binding domain and is screened against a library of activation-domain hybrids to select interacting partners

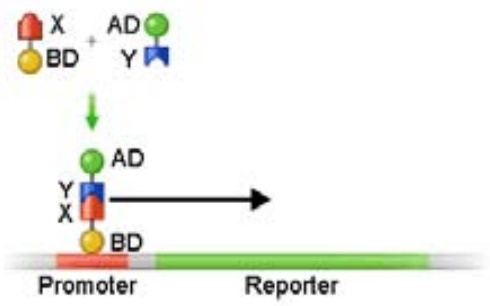


Figure 1.7 Yeast two-hybrid approach. The bait protein (X) that is fused to a DNA binding domain (BD) and the prey protein (Y) that is fused to the activation domain (AD) are co-expressed in the yeast nucleus; the interaction of X and Y leads to reporter gene expression (black arrow)

One of the key advantages of the two-hybrid assay is its sensitivity because it can detect interactions with dissociation constants around 10^{-7} M, in the range of most weak protein interactions found in the cell. Disadvantages of the yeast assay comprise the unavoidable occurrence of false negatives and false positives. False negatives include proteins such as membrane proteins and secretory proteins that are not usually amenable to a nuclear-based detection system. On the other hand, false positives seem to be

predominantly due to spurious transcription that does not derive from any interaction occurring between the hybrid proteins.

Since its creation, many improvements and variations of this technique have been reported. For example, it has been adapted to mammalian and bacterial cells[26, 27] or to specifically detect protein interactions that require a specific post-translational modification (called tethered catalysis method)[28].

1.1.1.2.2 Affinity purification and mass spectrometry (AP-MS)

The classic biochemical techniques for detecting protein interactions *in vitro* are immunoprecipitation and pullingdown assays, both based on affinity purification of a bait protein. These techniques were refined for proteomics applications by its coupling to mass spectrometry. Thus, this new approach combines the purification of protein complexes with identification of their individual components by mass spectrometry (Figure 1.8).

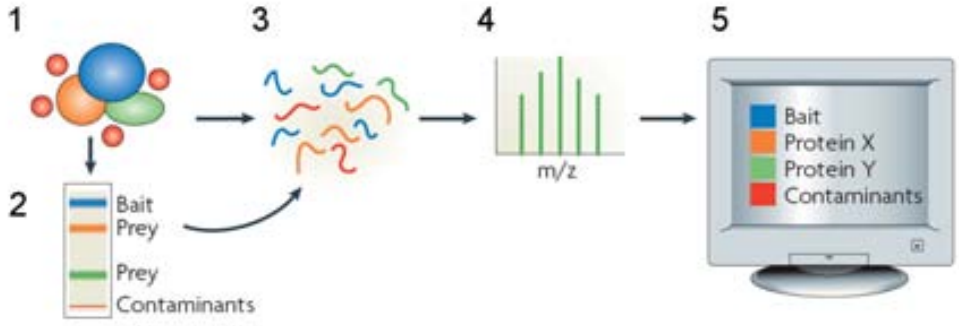


Figure 1.8 General overview of an affinity purification and mass spectrometry experiment. 1) The protein of interest (blue) is purified from a cell lysate together with its binding partners. 2) Proteins in the complex can be separated by SDS–PAGE or by some type of liquid chromatography 3) Proteins are subjected to proteolysis 4) Mass spectrometry (MS) analysis of peptides. 5) Database searching and statistical software are used to interpret the MS data to yield a list of proteins that were present in the initial sample[29].

Proteins of interest are simply expressed in-frame with an epitope tag (at either the N or C terminus), which is then used as an affinity handle to purify the tagged protein (the bait) along with its interacting partners (the prey). Although several different tags or tag combinations have been successfully used in many low-throughput studies[30], high-throughput studies have primarily employed the tandem affinity purification (TAP) tag system[31] (Figure 1.9).

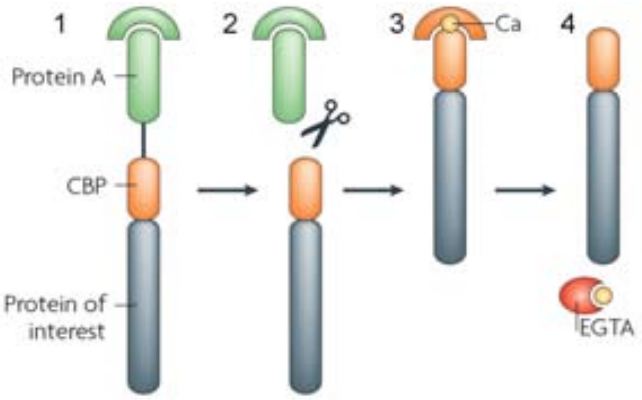


Figure 1.9 One of the TAP-based purification approach: 1) protein A is first immobilized on immunoglobulin G (IgG)–sepharose. 2) The tag is then cleaved with TEV (Tobacco etch virus protease) 3) The protein of interest, now fused only to the CBP (Calmoduline-binding peptide) is next immobilized on calmodulin–sepharose in the presence of calcium. 4) Calcium chelation (with EGTA) releases the recombinant protein, along with its interacting partners[29].

Afterwards, subunits in these complexes are usually identified by a combination of gel electrophoresis and mass spectroscopy. Although analysis of gel-purified proteins has been used most often so far, gel-free approaches allow for a more rapid and generic analysis and are increasingly used. In most cases, this involves peptide separation by reversed-phase liquid chromatography followed by two MS events. In the first scan, the mass/charge ratio (m/z) of the intact peptide is measured. Afterwards, the most abundant peptides are then specifically selected and subjected to fragmentation, yielding a tandem MS (MS/MS) spectrum. Two main strategies to ionize peptide ions, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) and their implementation on several types of tandem mass spectrometers have allowed for efficient sequencing of peptides derived from proteolytic digests of protein complexes.

One key advantage of this technique is that only the bait is genetically modified with the affinity tag, while the whole proteome is “fished” for prey. Furthermore, interactions between individual proteins and within entire protein complexes are being identified. However, transient interactions and complexes cannot be detected by TAP due to the stringent purification steps involved in the process[29].

1.1.1.2.3 Resonance energy transfer (RET)

Resonance energy transfer (RET) techniques are based on an energy transfer that occurs between a luminescent or fluorescent donor and a fluorescent acceptor[32]. These techniques enable the non-invasive monitoring of specific protein interactions.

The Fluorescence Energy Transfer (FRET) involves the energy transfer between fluorescent tags linked or genetically fused to two interacting proteins. It is a non-radiative process whereby energy from an excited donor fluorophore is transferred to an acceptor fluorophore within ~60 Å distance if their transitional dipoles are appropriately oriented[33].

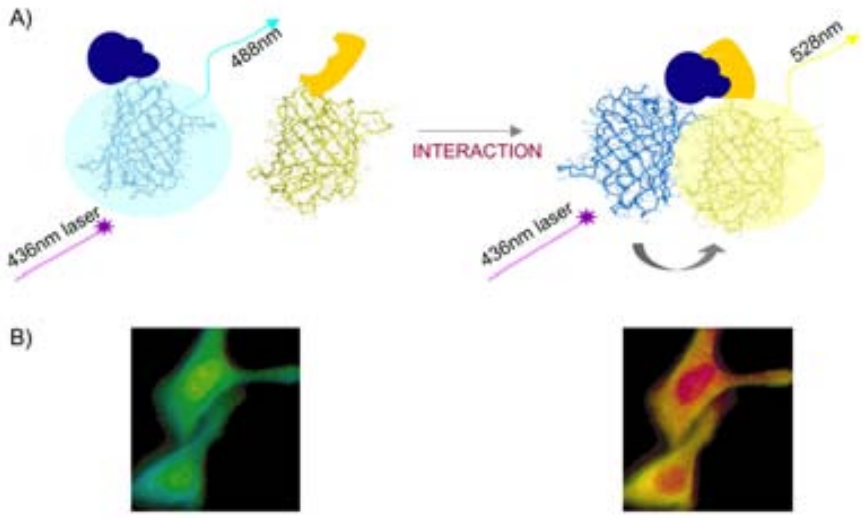


Figure 1.10 Using FRET to detect protein interactions. A) When there is no interaction, excitation of CYFP results in the emission of cyan light. As a result of the protein binding, CFP and YFP are brought into proximity and excitation energy is transferred, resulting in YFP fluorescence. B) FRET applied to measure intracellular calcium. Cyan fluorescence protein labeled calmodulin and yellow fluorescence protein labeled calmodulin binding peptide (M13-YFP) were co-expressed. Low Ca^{2+} levels led to little FRET and mostly blue emission (pseudocolor green) (left). High Ca^{2+} levels led to binding and FRET emission of YFP (pseudo color red) (right)[34].

After excitation of the first fluorophore, FRET is detected either by emission for the second fluorophore, or by alteration of the fluorescence lifetime of the donor. Two fluorophores that are commonly used are variants of the green fluorescent protein (GFP) like Cyan Fluorescent Protein (CFP) and Yellow Fluorescent Protein (YFP)[35]. A number of protein interactions have been visualized in cells by FRET microscopy[36-38].

Bioluminescence Resonance Energy Transfer (BRET)[39] is a naturally energy transfer phenomenon that occurs in some marine organisms (e.g. *Renilla reniformis*). In this process, Renilla luciferase protein (Rluc) that emits blue light in the presence of the substrate coelenterazine transfers energy to GFP or a variant (Figure 1.11). If there is no interaction between the two proteins of interest, Rluc and GFP will be too far apart for significant transfer and only the blue-light emitting from Rluc will be detected. The BRET signal is measured as the amount of green light emitted by GFP compared with the blue light emitted by Rluc.

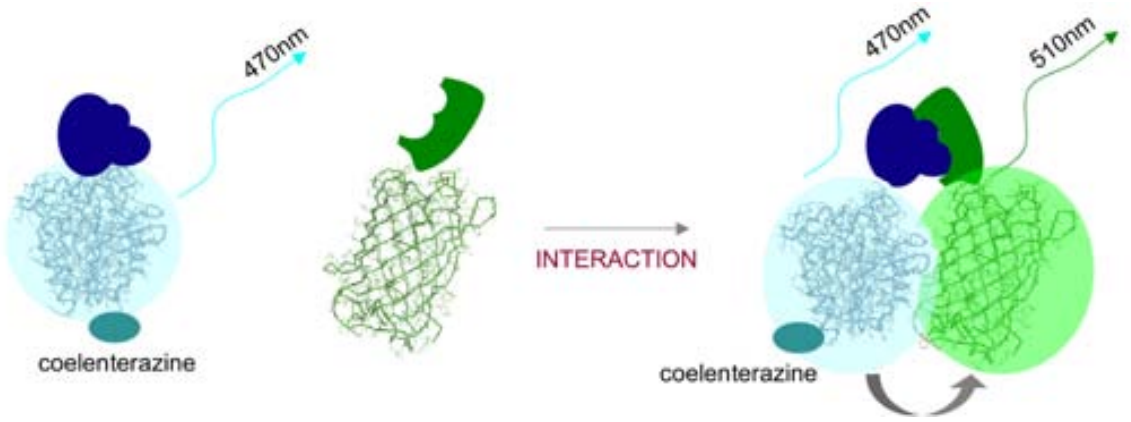


Figure 1.11 Bioluminescence resonance energy transfer (BRET) assay for monitoring protein–protein interactions. The protein–protein binding is evaluated by the fusion of one partner to Renilla luciferase (Rluc) and the other one to green fluorescent protein (GFP). On addition of the substrate coelenterazine, BRET emission from GFP is observed only when interaction between the two proteins occurs, which brings Rluc and GFP into close proximity.

The main difference between BRET and FRET is that BRET does not require the use of excitation illumination[40]. FRET based techniques are limited by the necessity for external illumination to initiate the fluorescence transfer. In this sense, FRET may be prone to complications due to simultaneous excitation of both donor and acceptor fluorophores. Specifically, even with monochromatic laser excitation, it is difficult to excite only the donor without exciting the acceptor fluorophore. Moreover, photobleaching of the fluorophores that could be a serious limitation of FRET is irrelevant to BRET. On the contrary, with FRET, dim signals can be amplified by simply increasing the intensity or duration of excitation, whereas in BRET the only option to improve low signal levels is to integrate the signal for a longer time.

Features of BRET and FRET offer some attractive advantages[39]. For instance, both techniques can be applied to determine whether the interaction changes with the time because of the non-invasive measurement. Both assays could be applied to monitor the dynamic processes of protein interactions *in vivo*, such as intracellular signaling. However, they have some limitations: their efficiency is dependent on proper orientation of the donor and acceptor dipoles.

1.1.1.2.4 Protein complementation assays (PCAs)

Among the most promising assays appeared recently are those based in protein fragment complementation, also called interaction trapping[41]. In these methods, the binding partners are fused to two rationally designed fragments of a reporter protein, which recovers its native structure and function upon binding of the interacting proteins.

To date enzymes like ubiquitin[43], dyhydrofolate reductase[42] or β-galactosidase[44] have been used as reporters. Enzyme based techniques require an incubation with the substrate, process that needs to be optimized with respect to concentration and reaction time depending of the proteins of interest.

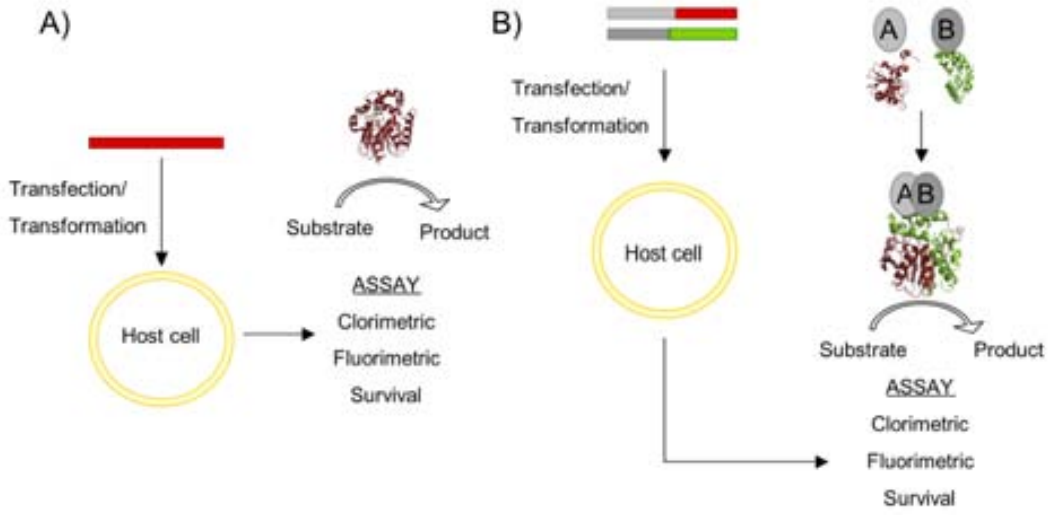


Figure 1.12 General Features of Protein complementation assays[42]. (A) An enzyme can be transformed/transfected into a host cell and its activity detected by an *in vivo* assay. (B) Interacting proteins (A, B) are fused to N- and C-terminal fragments of the gene encoding for the enzyme. Co-transformation/transfection of protein fusions (between proteins A/B and enzyme fragments) results in reconstitution of enzyme activity by binding between A and B. Reassembly of enzyme will not occur unless proteins A and B interact with each other.

For fast detection of interactions with low-background signal, PCAs with direct spectroscopic read-outs have been investigated. The green fluorescent protein (GFP) as well as several of its spectral variants[45] have been used. The use of fluorescent proteins in interaction trapping has important advantages: they have been shown to be functional in many cell types and cell compartments as tags for protein localization[46, 47] and their architecture assures that only those proteins interacting directly and no through one or more mediators will allow reconstitution of the fluorophore.

1.1.2 Bimolecular Fluorescence complementation (BIFC)

Bimolecular Fluorescence Complementation (BIFC) assays are PCA methods that use GFP and its spectral variants as reporters (Figure 1.13). As it has been mentioned before, BIFC has the advantage that the protein complex can be directly visualized in living cells without the need for staining with exogenous molecules.

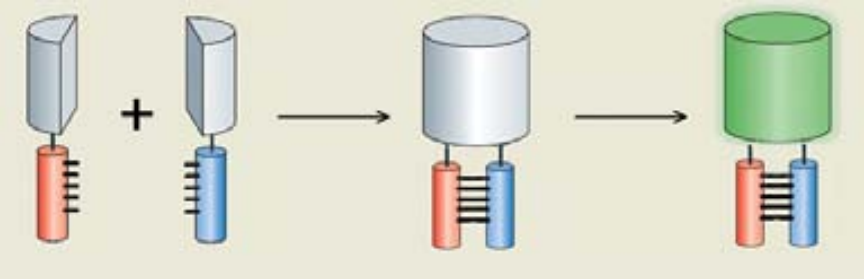


Figure 1.13 The BiFC approach is based on the formation of a bimolecular fluorescent complex when two non-fluorescent fragments of a fluorescent protein are brought together by an interaction between proteins A and B (red and blue cylinders) that are fused to the fragments. The interaction partners are indicated by coloured cylinders and the fluorescent protein fragments, in grey[48].

It has been demonstrated that the reassembly reaction of the fluorescent protein fragments requires their fusion to interacting proteins. In the Figure 1.14, a scheme shows the fundamentals of the BIFC mechanism whose dynamics have been studied previously[45]. The process starts with the interaction of the bait and prey proteins fused to the fluorescent fragments (complex I). Importantly, this binding occurs in competition with the alternative endogenous interaction partners present in the cell (complexes II). The interaction brings the two fluorescent fragments in proximity enabling their noncovalent reconnection and later folding. This process is slow and produces an intermediate (complex III), which undergoes slow maturation forming the native β-barrel structure of the fluorescent protein and its chromophore (complex IV).

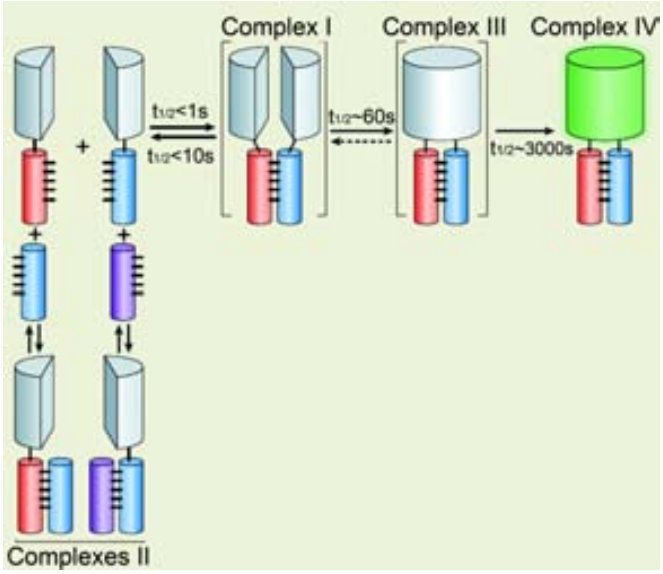


Figure 1.14 Pathway for bimolecular fluorescent complex formation. *In vitro* studies using purified proteins indicate that the initial associations between the fusion proteins (complex I) are mediated by the interaction partners. This interaction occurs in competition with alternative interaction partners, which could form mutually exclusive interactions (complexes II). The association between the fluorescent-protein fragments is slower and produces an intermediate (complex III), which undergoes slow maturation to produce the peptide fluorophore (complex IV).

A set of different fluorescent proteins and fragmentation sites can been used for BIFC (Table 1.1). Moreover, fragments from different fluorescent proteins can reassemble forming a protein with characteristic spectral properties. This fact enables the simultaneous visualization of multiple protein complexes in the same cell[49].

Table 1.1 Fluorescent proteins used in BIFC approach

Protein	Fluorescent protein fragment		Filters			
	Nomenclature	Dissection point	Excitation laser (nm)	Excitation filter (nm)	Dichromatic mirror (nm)	Suppression filter (nm)
Yellow Fluorescent Protein (YFP)	NYFP CYFP	(1-154)/(1-172) (155-238)/(173-238)	Argon (488)	500/20	515	535/30
Red Fluorescent Protein (mRFP1-Q66T)	mRFP1-Q66T-N mRFP1-Q66T-C	(1-168) (168-225)	He-Ne(594)	560/55	595	630/60
Green Fluorescent Protein (GFP)	NGFP CGFP	(1-157) (158-238)	Argon (488)	480/20	505	520/20
Cyan Fluorescent Protein (CFP)	NCFP CCFP	(1-157) (158-238)	Diode (440)	436/20	455	480/40
Venus Fluorescent Protein (VFP)	NVFP CVFP	(1-154)/(1-172) (155-238)/(173-238)	Argon (488)	495/35	515	545/40

BIFC can be also used to investigate the competition between mutually exclusive interaction partners as well as to compare the subcellular distributions of different complexes[50]. Nowadays, BIFC analysis has been used to study interactions among wide range of proteins in many cell types and organisms. In the following table, there are referenced some examples of the different applications of BIFC assays.

Table 1.2 Examples of protein interactions visualized using BIFC in different organisms

Proteins	Organism
Cytoskeleton proteins[51] Proteins involved in lipolysis[52]	Mammalian cells
Leucine zippers[53], SH3 domain[54]	Bacteria (E*scherichia.coli*)
Stomatin-like proteins (SPL) related to locomotion[55]	Nematode (*Caenorhabditis elegans*)
Proteins related to MicroRNA biogenesis[56] Membrane proteins[57]	Plants (*Arabidopsis, N.benthamiana*)
Signaling proteins[58], Transcription factors[59]	Yeast (*Saccharomyces cerevisiae*)

At the time this PhD thesis had been commenced, most BIFC studies had been focused on the detection of strong interactions (e.g. the homo and hetero-dimerization of

leucine zipper domains)[45]. Importantly, the majority of the relevant interactions in the cell are weak and, in many cases, transient. Given the great demand for methods that would allow the detection of such interactions, it was decided to study in detail the ability of BIFC to detect weak protein binding using SH3 domain interactions as a model.

1.1.3 The interactions of the Abl-SH3 domain as a test case

1.1.3.1 SH3 domains

SH3 domains are one of the most widely spread recognition modules in the proteome[60] and are involved in the recruitment of substrates and the regulation of kinase activity in tyrosine kinases by mediating specific but transient protein–protein interactions[61-65.] SH3 domains share common structural features: they are 50–70 amino acids long and consist of five β-strands arranged into two sheets packed at right angles.

Early studies indicated that most SH3 domains recognize proline-rich peptides with xPxxP (x = aliphatic amino acids or Pro) as a core conserved binding motif[66,67.] Due to the high content in prolines, these peptides adopt an extended, left-handed polyproline II helical conformation in these complexes[63, 68]. While the two proline dipeptides (xP) each occupy a well-conserved hydrophobic binding pocket on the surface of SH3 domains, it exists a third cleft (called the specificity pocket) recognized by a distal residue (Φ) N- or C-terminal to the xPxxP motif, i.e. ΦxxPxxP (class I ligands) or xPxxPxΦ (class II ligands)[60, 61]. However, several SH3 domains have been recently reported to bind sequences lacking the consensus PxxP motif. Therefore, the recognition repertoire of the SH3 domain family appears to be more diverse than originally thought[69-73].

1.1.3.2 SH3 domain of the Abl kinase

Abl tyrosine kinase is a protein that regulates cellular differentiation and apoptosis, being responsive to extra cellular signals (growth factors, cell adhesion and cytokines) and internal signals (DNA damage, oxidative stress)[74-76]. It also might affect cell growth positively or negatively depending on the cellular context[77] and become an oncogene if it is expressed in an unregulated manner. It contains an SH3 domain (Abl-SH3) that has been implicated in the negative regulation of the kinase activity by mediating protein-protein interactions[78].

1.1.3.3 Peptidic ligands of the SH3 domain

The SH3 domain of Abl tyrosine kinase selectively binds class I ligands and prefers a hydrophobic residue at position Φ. In fact, Met and Tyr are present at this position in the first two isolated Abl-SH3 binding proteins (3BP1 and 3BP2)[67]. Mutagenic analysis of both proteins identified short proline-rich sequences as binding sites for the Abl-SH3 domain. These regions share the consensus sequence XPXXPPPΨXP, where Ψ is a hydrophobic residue and X is any amino acid. In particular, prolines at positions 2, 7 and 10 of the ligand have been shown to be essential for the binding to this domain[66].

In order to create specific synthetic ligands for the Abl-SH3 domain, a group of decapeptides was designed[79] by increasing their tendency to adopt PPII conformation and creating new favorable interactions, relative to the original sequence in BP1. Among these, p41 peptide (APTMPPPLPP) interacted with SH3 domain more specifically with a K_d of 1.5 μM. In the Figure 1.15, it is shown the crystal structure of this complex at 1.6 Å resolution[80]. The p41 peptide-Abl SH3 domain interaction has been used in the present work as a model to characterize the feasibility of BIFC to detect *in vivo* weak peptide-protein interactions.

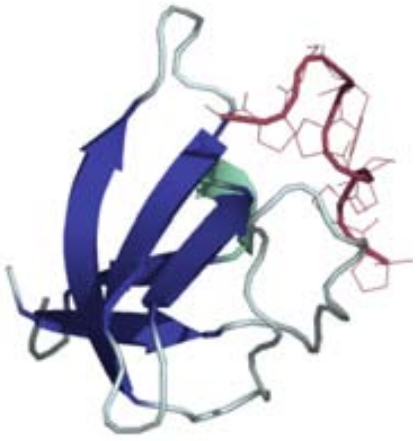


Figure 1.15 Crystal structure of the Abl-SH3 domain (in blue) interacting with the p41 peptide (in magenta)

1.1.3.4 Proteins interacting with the Abl-SH3 domain

The study of short proline-rich peptides is useful to dissect the different forces and factors contributing to binding, but they do not necessarily represent the *in vivo* binding affinity and specificity of complete proteins. To assess if BIFC could detect weak protein-protein interactions, we selected a polypeptide involved in the regulation of important cellular processes through their binding to the SH3 domain of Abl-tyrosine kinase: Breast cancer type 1 susceptibility protein.

Breast cancer type 1 susceptibility protein (BRCA1) is a tumour suppressor protein associated with breast and ovarian cancer[82]. The exact function of BRCA1 is not well defined but it is known that BRCA1 is implicated in a number of cellular processes[83] being a component of multiple repair pathways as a response to stress, specifically to agents that cause DNA damage[84, 85].

Figure 1.16 Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1[81].

It has been demonstrated that the interaction between Abl kinase and BRCA1 controls its kinase activity[86]. Also, it has been proposed that the binding could occur through the SH3 domain of Abl kinase and a tandem of two BCRT domains located at the C-terminus of BRCA1. The BRCT domain is an evolutionary conserved phospho-protein binding domain involved in cell cycle-control[87].

1.2 Objectives

→ Demonstrate the applicability of Bimolecular Fluorescence Complementation (BIFC) technique to detect weak protein interactions (μM range) *in vivo*.

→ Analyze the sensitivity of BIFC method to detect mutations that affect the interaction strength.

→ Study the capabilities of the coupling between BIFC and flow cytometry as a high-throughput technique to screen mutations that involve a specific interaction.

→ Test the ability of BIFC to discriminate the regions that are involved in a specific protein-protein interaction.

1.3 Experimental procedures

1.3.1 Plasmid construction

In the table 1.3 and Figure 1.17, there are summarized all the created constructions. All were produced using the strategy explained in the general Experimental Procedures section.

Both fluorescent protein fragments: C-terminal (CYFP) and N-terminal (NYFP) were amplified from YFP cDNA (Clontech). The Abl-SH3 domain was amplified from pET3d-Abl-SH3[79] and the DNA encoding p41 was created by direct annealing of two synthetic complementary DNA oligonucleotides. On the other hand, the C-terminal BRCT domains of BRCA1 (1638-1863) were amplified from pRSV-BRCA1[88].

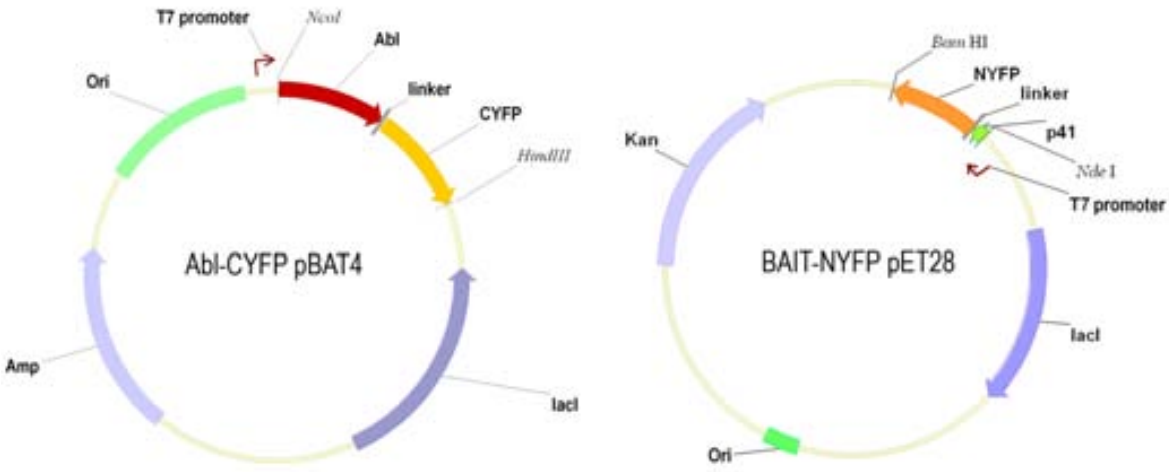


Figure 1.17 Scheme of the designed plasmids. They were designated Abl-CYFP pBAT4 and BAIT-NYFP pET28 (where BAIT accounts for the targets whose binding to the SH3 domain was assayed: peptide p41 and BRCA1).

Table 1.3 Design of the fusion protein constructs

Protein	Fluorescent protein fragment	Linker sequence	Plasmid	Restriction sites	Antibiotic resistance
c-Abl SH3 domain	CYFP (156-238)	SGGGSGGS	pBAT4	*NcoI, HindIII*	Ampicillin
p41 peptide	NYFP (1-155)	SGGGS	pET28a(+)	*NdeI, BamHI*	Kanamycin
BRCA1	NYFP (1-155)	SGGGSGGS	pET28a(+)	*NdeI, BamHI*	Kanamycin

1.3.2 Site-directed mutagenesis

Site-directed mutagenesis of the above-mentioned fusion proteins was performed using the *QuickChange site-directed mutagenesis kit* from Stratagene according to the procedure recommended by the manufacturer. It allows site-specific mutations in double-stranded plasmid. The basic procedure uses a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation (Figure 1.18). The oligonucleotide primers are extended during temperature cycling by a DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *Dpn* I. The *Dpn* I endonuclease is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to *Dpn* I digestion. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue supercompetent cells. Afterwards, all constructs have to be verified by DNA sequencing.

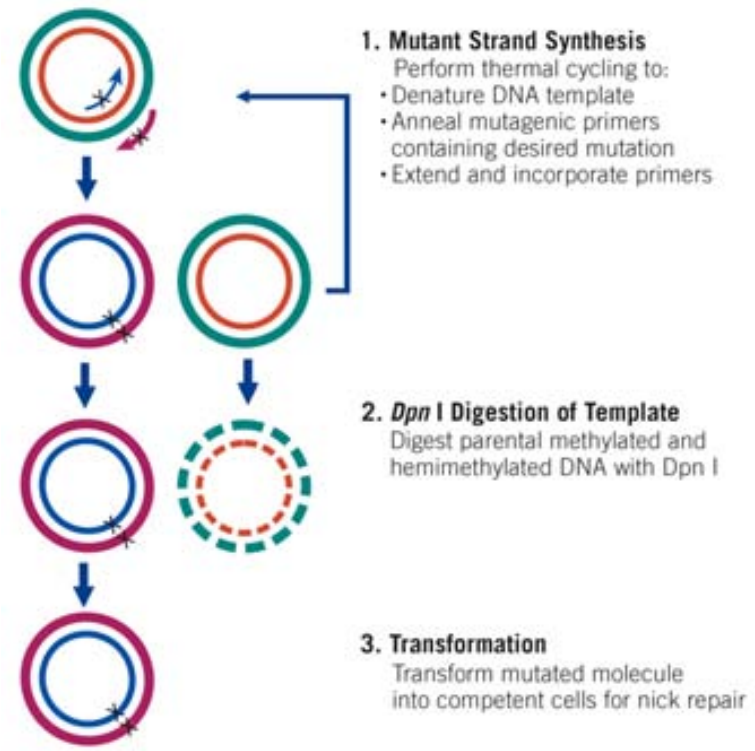


Figure 1.18 The QuikChange site-directed mutagenesis method. Mutant-strand synthesis is followed by DpnI digestion of the parental DNA template, and transformation of the resulting annealed double stranded nicked DNA molecules. After transformation, the XL-1 Blue *E. coli* cell repairs nicks in the plasmid.

1.3.2.1 Construction of a reduced library of the p41 peptide at position 4

Using mutated primers and site-directed mutagenesis a collection of p41-NYFP plasmids mutated at position 4 of p41 peptide was obtained. *E.coli* BL21(D3) cells, already transformed with a plasmid encoding Abl-SH3-CYFP were transformed with an equimolar mixture of the five plasmids. Fifty positives clones were picked from an LB/Kan/Amp plate, and protein expression and fluorescence quantification was carried as explained in the general Experimental Procedures section. The DNA of each colony class was sequenced to identify the residue at position 4 in the correspondent p41-NYFP fusion.

1.3.3 Modeling of the Complex of the Abl-SH3 Domain and p41 mutants

The coordinates of the X-ray of the Abl-SH3 domain interacting with the p41 peptide[80] were used to model the complex of Abl-SH3 with the p41 mutants. First, Tyr4 was mutated to Arg, Phe, Trp, or Gly using Swiss PdbViewer 3.7 and the complex submitted to Swiss Model (http://swissmodel.expasy.org) for energy minimization with the implemented GROMOS96 force field.

1.3.4 Determination of K_d by titration

Assuming a one-one complex between the SH3 domains and the peptide, it is possible to determine the K_d for the interaction of the different peptides, by monitoring the changes in the fluorescence emission, using the equation[79]:

$$F = F_f + \left(F_b - F_f\right)\left[pep_f\right] / \left(K_d + \left[pep_f\right]\right)$$

where F_f is the fluorescence of the free domain (Dom_f), F_b is the fluorescence of the complex (Comp), [pep_f] is the concentration of the free peptide in solution, and K_d is the dissociation constant in the equilibrium.

$$K_d = \left[Dom_f\right]\left[pep_f\right] / \left[Comp\right]$$

The concentration of free peptide can be calculated by subtracting the estimated concentration of the complex from the concentration of added peptide.

1.3.5 Urea denaturation of the reconstituted complex

200 µl of soluble cell fraction was dissolved into a final volume of 1 ml in 0.1 M Tris–HCl buffer (pH 7.5) containing selected concentrations (0-7 M) of urea. After incubation at room temperature during 24 hours, YFP fluorescence was monitored as described in the general Experimental Procedures section. The fitting was performed using the non-linear, least-squares algorithm provided with the software KaleidaGraph (Abelbeck Software).

1.3.6 Native electrophoresis and in-gel detection of YFP fluorescence

Cells co-expressing the different protein fusions were treated as described in the general Experimental Procedures section in order to obtain the soluble intracellular fractions that were loaded in the gel wells. Gel electrophoresis was performed at 4°C as previously described[89] without SDS or DTT in any buffer. The gel was exposed to UV light using a transilluminator to detect YFP fluorescence.

1.4 Results

1.4.1 A system to interrogate Abl-SH3 interactions

A crucial step in the BIFC approach is the design of the fusions of the bay and prey proteins to the corresponding fluorescent protein fragments. First, it has to be determined the fluorescent protein that will be used. The Enhanced Yellow Fluorescent Protein (YFP) was selected in this work. Nevertheless, other spectral variants as the CFP or GFP could be used instead.

Secondly, the topology of the protein fusions has to be decided. Taking into account the crystal structure of the complex between p41 peptide and Abl-SH3 domain[80], fusions were designed in order to avoid steric constraints in the reconstitution of the fluorescent protein: both proteins were fused to the N-terminus of the fluorescent protein fragments. On the other hand, if the bait and prey have dissimilar sizes as in this case, it is advisable to fuse the smaller one to the N-terminal fragment of the fluorescent protein and the bigger one, to the C-terminal moiety in order to avoid expression problems due to the protein fusion size. Therefore, Abl-SH3 domain was fused to the C fragment (CYFP) and p41 peptide, to the N fragment (NYFP). In the case of the binding of Abl-SH3 domain and BRCA1 protein, there was no previous knowledge about the complex structure. Thus, taking into account the previous design, BRCA1 was fused to the NYFP fragment (like p41 peptide).

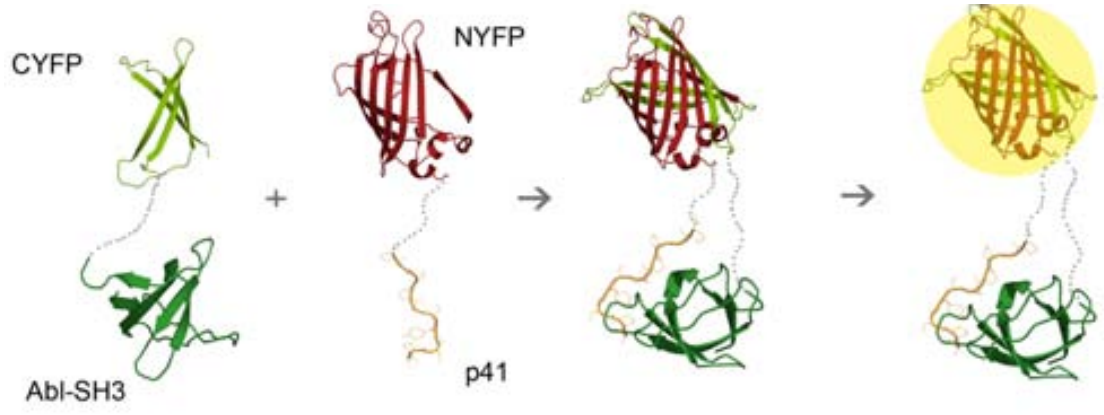


Figure 1.19 Models of the Abl-SH3-CYFP and p41-NYFP fusion proteins and their interaction. The schemes are based on the X-ray crystal structure of the complex between Abl-SH3 domain and p41 peptide[80] and the structure of GFP[90]. The structure of the linkers connecting the fused proteins is unknown.

The linker is another important part of the design. It could change in length and in composition allowing the interaction between the prey and bait proteins and the reassembly of the fluorescent protein. In our case, a polypeptide of glycines and serines was used in order to increase the flexibility and the solubility of the protein fusions, minimizing also the steric constraints that could occur during complex formation. The Abl-SH3 domain and BRCA1 proteins were separated from their respective YFP fragments by an octapeptide linker (SGGGSGGS). Meanwhile, the p41 peptide was separated by a shorter pentapeptide linker (SGGGS) to reduce the peptide mobility and the entropic penalty upon binding.

Through standard molecular biology techniques, a set of plasmids encoding for the different protein fusions was generated. Due to their different antibiotic resistance, they can be co-maintained and co-induced in *Escherichia coli*.

To characterize a given interaction, the Abl-SH3-CYFP and the correspondent BAIT-NYFP plasmid were co-transformed into *E.coli* BL21(DE3). Co-expression of the bait and prey fusion proteins was induced by addition of IPTG. After induction, cells were grown at 37°C or 18°C and the expression levels and solubility of protein fusions were analyzed by SDS-PAGE.

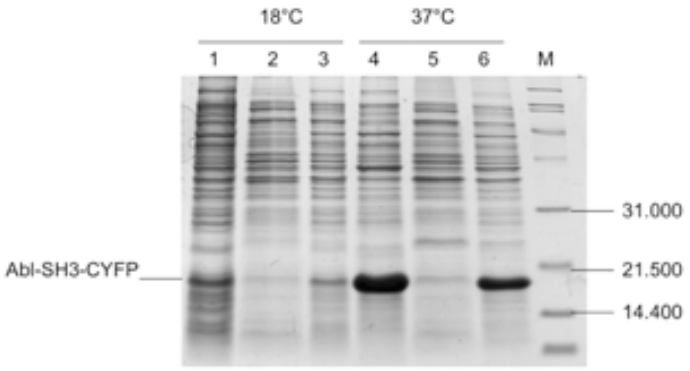


Figure 1.20 SDS-PAGE analysis of the inclusion bodies (IB) formation in the case of the protein fusion Abl-SH3-CYFP. The protein expression level was analyzed at different temperatures (18ºC and 37ºC) and in different cell fractions (1, 4, total; 2, 5 soluble fraction; 3, 6 insoluble fraction). (M) Molecular Weight Markers

In general, expression at 37°C resulted in higher fusion protein levels, but in most cases the protein failed to fold and was accumulated into inclusion bodies, preventing the establishment of specific interactions and consequently the detection of any fluorescence. The total protein expression level decreased at 18°C, but in all cases the fraction of soluble recombinant protein increased, sufficing to allow the reassembly of YFP fragments and detection of fluorescence emission in case that the interaction took place.

1.4.2 Exploring protein-peptide interactions: the Abl-SH3 and p41 system

1.4.2.1 Detecting Abl-SH3 domain and p41 peptide interaction

The interaction between Abl-SH3 domain and the poly-proline peptide p41 was used as a model to analyze if the establishment of weak protein interactions allowed the specific reassembly of YFP fragments *in vivo*. BIFC signal could be detected using a microscope or a fluorescence spectrophotometer.

First of all, *E.coli* cells were examined using fluorescent microscopy. Cells expressing either Abl-SH3-CYFP or p41-NYFP alone did not exhibit any fluorescence upon induction, showing that YFP fragments are not individually functional. However, when both protein fusions were co-expressed, the majority of the cells emitted high YFP fluorescence as examined by microscopy (Figure 1.21). To test if the specific interaction between the Abl-SH3 domain and the p41 peptide was driving the reconstitution of YFP or, on the contrary, it was just the result of the unspecific interaction between the two YFP fragments inside the cells, N-terminal YFP fragment was also co-expressed with Abl-SH3-CYFP. No fluorescence was detected in this case, emphasizing the requirement for the presence of both partners for YFP reconstitution

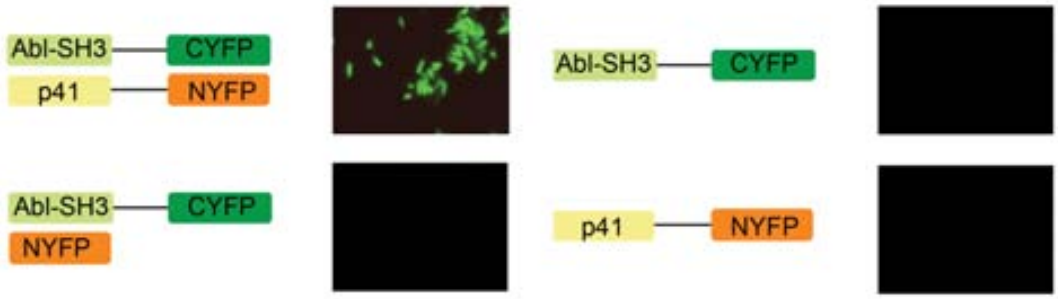


Figure 1.21 Visualization of the interaction between Abl-SH3 and p41 peptide by BIFC. Fluorescence images of *E. coli* induced cells expressing the protein fusions indicated beside.

Besides, the fluorescence of the soluble intracellular fraction could be measured by spectrophotometry. Analysis of the fluorescence emission spectra of the soluble fraction of cells co-expressing Abl-SH3-CYFP and p41-NYFP showed the typical maximum at 523 nm, as expected for native YFP and confirming the correct reassembly of the reporter (Figure 1.22). This fact indicated that the β-barrel structure and the chromophore were likely to be identical in the BIFC complex and in the intact fluorescent protein. Besides, the detected fluorescence levels correlated with the images obtained previously.

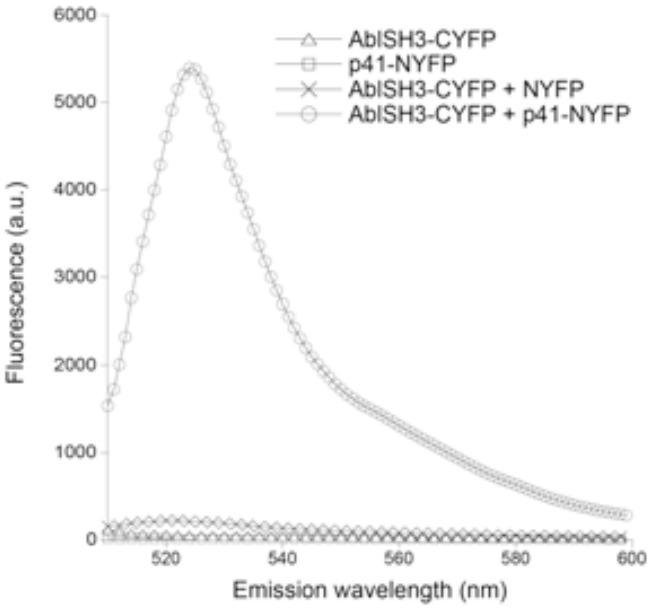


Figure 1.22 Fluorescence emission spectra of the soluble cell fraction from *E. Coli* cells co-expressing Abl-SH3-CYFP and p41-NYFP, Abl-CYFP and NYFP, Abl-SH3-CYFP and p41-NYFP.

1.4.2.2 BIFC sensitivity to mutations in the interaction surface

After demonstrating the applicability of BIFC approach to study weak protein interactions, we wanted to test its sensitivity to substitutions in the binding surfaces that modify binding capabilities. To this aim, a point mutation was introduced in the bait protein: Abl-SH3 domain. Specifically, Pro54 is together with Trp36 the most conserved residue in Abl-SH3 domains. It is located at the centre of the hydrophobic binding surface and plays an important role in the binding to poly-proline peptides. It has been shown that the mutation Pro54Leu reduces the binding of the Abl-SH3 domain to proline-rich ligands and deregulates Abl kinase activity *in vivo*[91]. Therefore, it was thought that this mutation could be a good choice to analyze whether such reduced binding affinity could be detected by the BIFC method. Accordingly, this mutation was introduced in the Abl-SH3-CYFP fusion protein (Abl-SH3P54L-CYFP).

The first step was to ensure by western blot an equal expression level of both protein fusions: Abl-SH3-CYP and Abl-SH3P54L-CYFP.

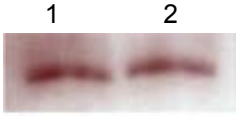


Figure 1.23 Western blot analysis of cells expressing: (1) Abl-SH3-CYFP and (2) Abl-SH3P54L-CYFP, detected with anti-GFP antibody.

Co-expression of Abl-SH3P54L-CYFP and p41-NYFP proteins resulted in cells emitting strongly reduced fluorescence regarding those expressing the wild-type Abl-SH3 (Figure 1.24A). Besides, this reduction could be better quantified by comparing the emission spectra of the respective soluble cell fractions (Figure 1.24B).

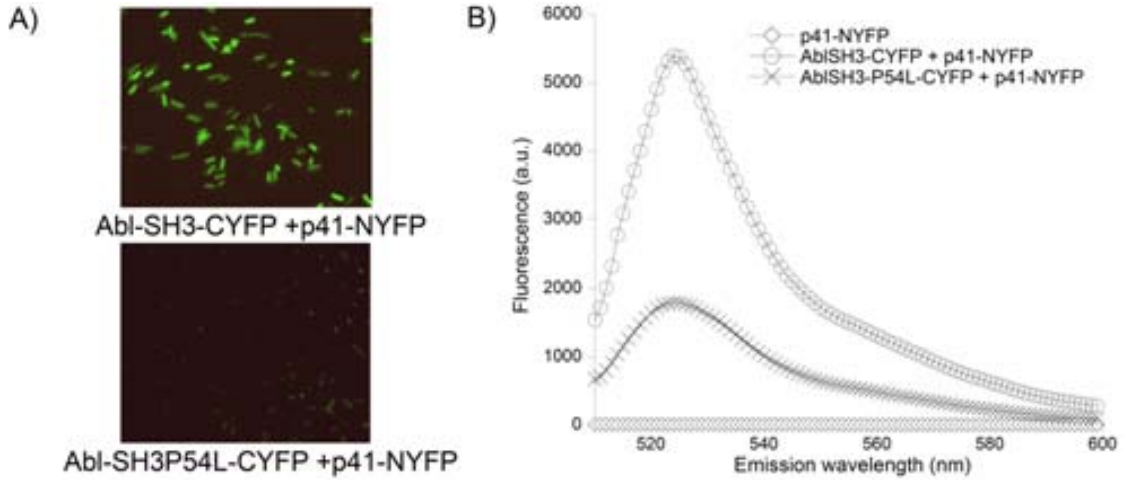


Figure 1.24 A) Fluorescence images of *E. coli* induced cells expressing the protein fusions indicated in each case. B) Fluorescence emission spectra of the soluble cell fraction from *E. coli* cells co-expressing Abl-SH3-CYFP or Abl- SH3P54L-CYFP and p41-NYFP.

1.4.2.3 YFP reconstitution traps the interaction between Abl-SH3 and p41

Although the interaction between the Abl-SH3 domain and p41 peptide was clearly required for YFP reassembly, its relevancy for maintaining regrouped YFP integrity was unknown. If the interaction between the bait and prey proteins was really essential, the presence of a proline-rich peptide that competes with the p41 sequence for the binding site of Abl-SH3 should result in a subsequent decrease of fluorescence emission. As described before, 3BP1 peptide is a natural proline-rich peptide (APTMPPPLPP), which binds to Abl-SH3 domain[80]. The interaction occurs through the same surface and in the same orientation than p41 peptide. Nevertheless, the addition of a >100 molar excess of 3BP1 peptide to the soluble fraction of cells co-expressing Abl-SH3-CYFP and p41-NYFP had negligible effect on fluorescence emission (Figure 1.25A) suggesting that the complex was already trapped by the YFP reconstitution.

To further confirm this point, the stability of the reassembled complex was analyzed in front of urea denaturation by monitoring the changes in YFP fluorescence emission. *In vitro*, the presence of 1.5 M urea sufficed to abrogate the interaction between a synthetic p41 peptide and the Abl-SH3 domain (Ventura, unpublished results).

In contrast, the Abl-SH3-CYFP/p41-NYFP complex exhibited little loss of fluorescence upon incubation in 2.0 M urea for several days, being the midpoint of urea denaturation around 2.8 M (Figure 1.25B).

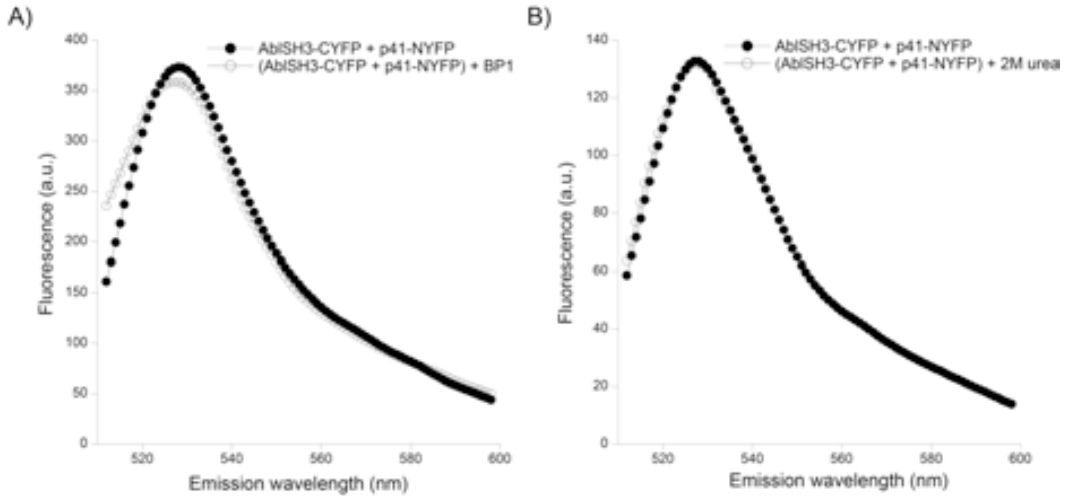


Figure 1.25 Stability of the reconstituted complex. Fluorescence spectra of the soluble cell fraction from *E.coli* cells co-expressing Abl-SH3-CYFP and p41-NYFP in the absence and presence of 2M urea (rigth) or a >100 molar excess of BP1, a p41 peptide competitor (left).

The urea denaturation profiles of both reconstituted Abl-SH3-CYFP/p41-NYFP and Abl-SH3P54L-CYFP/p41-NYFP complexes were compared. Similar denaturation curve and midpoint urea transitions were obtained (Figure 1.26). This fact suggested that the strength of the Abl-SH3/p41 interaction did not contribute significantly to the integrity of the complex, which was mainly maintained by the reassembly of fluorescent protein fragments.

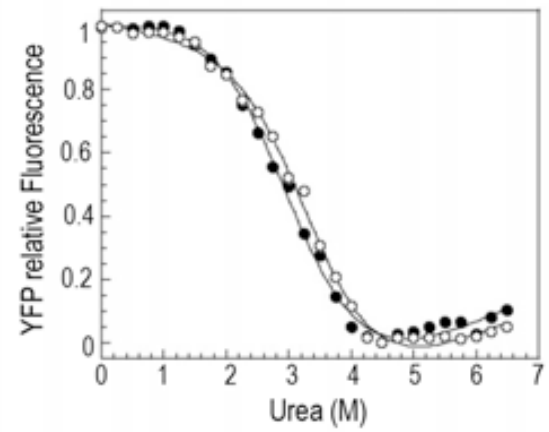


Figure 1.26 Denaturation of the soluble reconstituted complexes in the presence of urea and monitored by YFP fluorescence emission: Abl-SH3-CYFP/p41-NYFP (solid circles) and Abl-SH3P54L-CYFP/p41-NYFP (empty circles).

1.4.2.4 Screening of mutations that affect interaction strength

The BIFC technology could have the potential to become a robust but simple method to screen libraries of related ligands exhibiting different affinities and specificities for a specific target. If the fluorescence emission depended on the particular strength of the interaction between the bait and prey proteins, the system would probably allow parallel selection against several targets in the same experiment.

As a test case it was assayed whether differences in fluorescence emission upon binding to Abl-SH3-CYFP could be detected between the members of a small library of p41 peptide variants.

In the crystallographic structure of the complex, the N-terminus of p41 (residues 1 to 5) is bound into a valley between the RT and n-Src loops of Abl-SH3, a site that diverges in sequence in the family, being responsible for binding specificity[92]. The identity of these residues, and particularly of residue 4 (Tyr in p41 peptide), sharply modulates the binding affinity[79]. Based on this evidence, a set of plasmids encoding for five p41-NYFP variants differing in the position 4 of the p41 peptide was generated. The different peptides presented Trp, Phe, Tyr, Arg or Gly in the fourth position to evaluate if differences in size, entropy or polarity could affect the binding and fluorescence emission in the system.

The five plasmids were mixed in an equimolar ratio and transformed into cells already containing the plasmid encoding for Abl-SH3-CYFP. Fifty positives clones were selected and the co-expression was induced.

As expected, the clones could be classified into five classes according to their different fluorescence emission. This fact indicated that the mutated position in p41 peptide influenced the *in vivo* binding affinity and that the method is sensitive enough to detect such changes. The five classes were sequenced and the affinity order resulted to be: Tyr>Arg>Phe>Gly>Trp (Figure 1.27). With the exception of the Trp mutant, no significant differences in expression levels between protein fusions were observed, as analyzed by western blot (Figure 1.27). In addition, to gain structural insights into the affinities displayed by the different p41 variants, we modeled their complexes with Abl-SH3. According to the energies associated to the minimized models they ranked: Tyr>Arg>Trp>Phe>Gly (Figure 1.28)

The key contribution of the side chain of residue in position 4 to the interaction with the binding surface of the Abl-SH3 domain was clearly illustrated by the very low fluorescence emission displayed by the p41-Gly4 mutant.

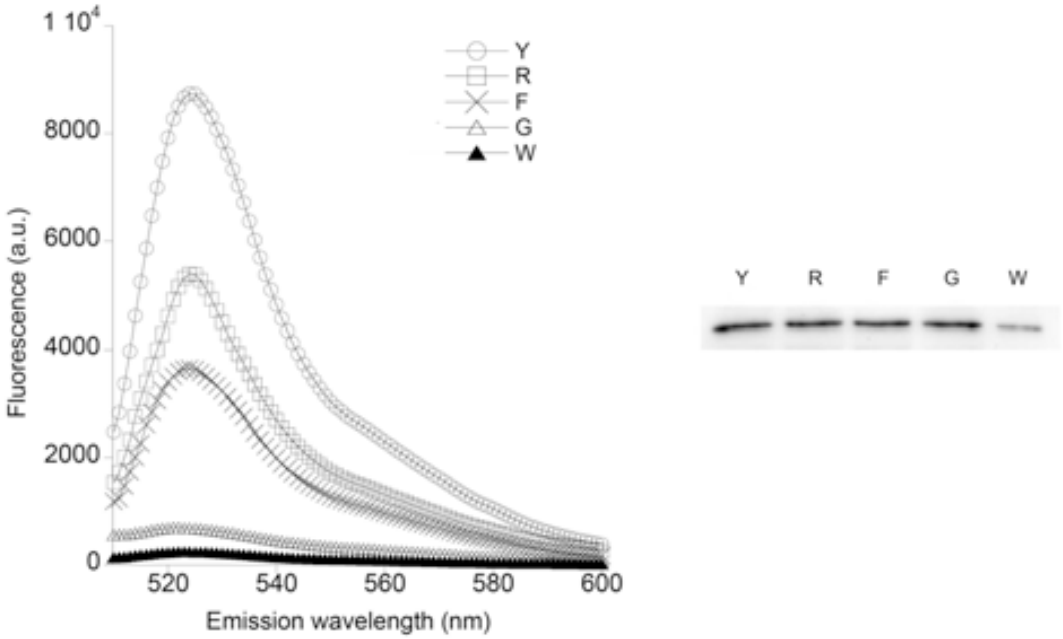


Figure 1.27 Analysis of a reduced p41 peptide library. Representative fluorescence emission spectra of the soluble cell fraction of *E. coli* from a set of 50 different clones co-expressing Abl-SH3-CYFP and a mixture of five mutants of p41-NYFP in the position 4 (left). Western blot analysis of cells expressing different p41-NYFP fusions detected with anti-GFP antibody. The residue in position 4 of p41 is indicated above (right).

In the native p41 peptide, wild-type Tyr, exhibited the highest fluorescence emission. The preference for this amino acid had been rationalized by the presence of a hydrogen bond between Tyr and the side chains of Ser12 and Asp14, lining the specificity pocket of the protein as seen in the crystal structure[80]. Also, the aromatic ring of Tyr allowed hydrophobic interactions between both surfaces[93, 94].

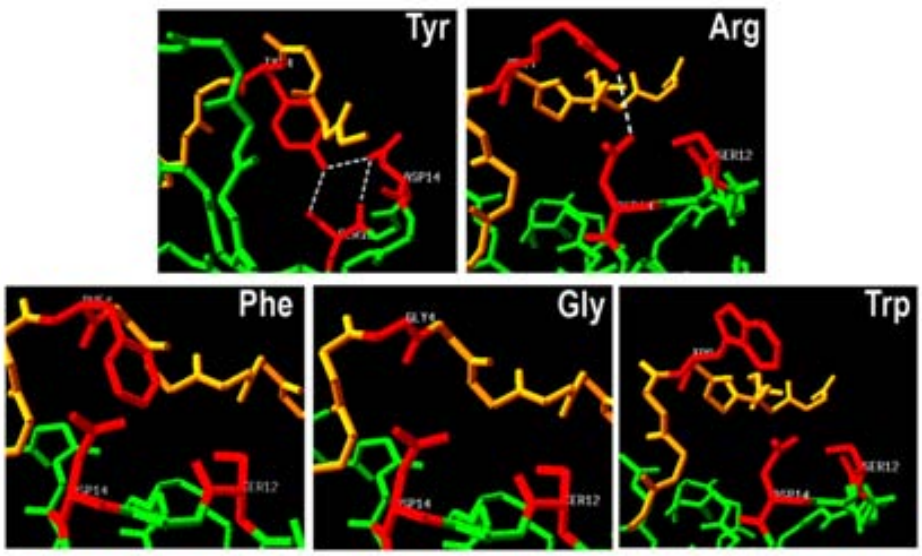


Figure 1.28 Modeling of the interaction between Abl-SH3 domain and p41 peptide with different mutations in the fourth amino acid: Tyr, Arg, Phe, Gly and Trp. The amino acids directly involved in the binding are depicted in red.

The reduced fluorescence of p41-Phe4 mutant relative to wild-type p41 confirmed the contribution of the hydrogen bonds established by Tyr4 to the binding energetics. Phe shared with Tyr the aromatic ring but lacks the hydroxyl group, being thus unable to establish hydrogen bonds. Moreover, a recent study of the binding energetics of proline-rich peptides to the Abl-SH3 domain confirmed the relevant role of hydrogen bonds in the interaction[95].

An interesting case is Arg4, which ranked second in our analysis. The preferred residue at the position 4 in most SH3 domains is Arg[66, 94], but Abl-SH3 displays a preference for hydrophobic residues. From the model, it appears that the failure of Arg to establish hydrophobic interactions could be compensated by the establishment of two short hydrogen bonds with the main chain carbonyl and the side chain hydroxyl of Ser12 in the Abl-SH3 domain. To test if the unexpected accommodation of a basic residue in the specificity pocket of Abl-SH3 could be reproduced and quantified *in vitro*, the interaction between a synthetic p41-Arg4 peptide and purified Abl-SH3 domain was titrated by monitoring the changes in Trp fluorescence. This technique had been applied before to determine the K_d of complexes between SH3 domains and designed peptides. Specifically, in the case of Abl-SH3 domain and peptide p41Y4R, a K_d of 20 ±1 μM was obtained.

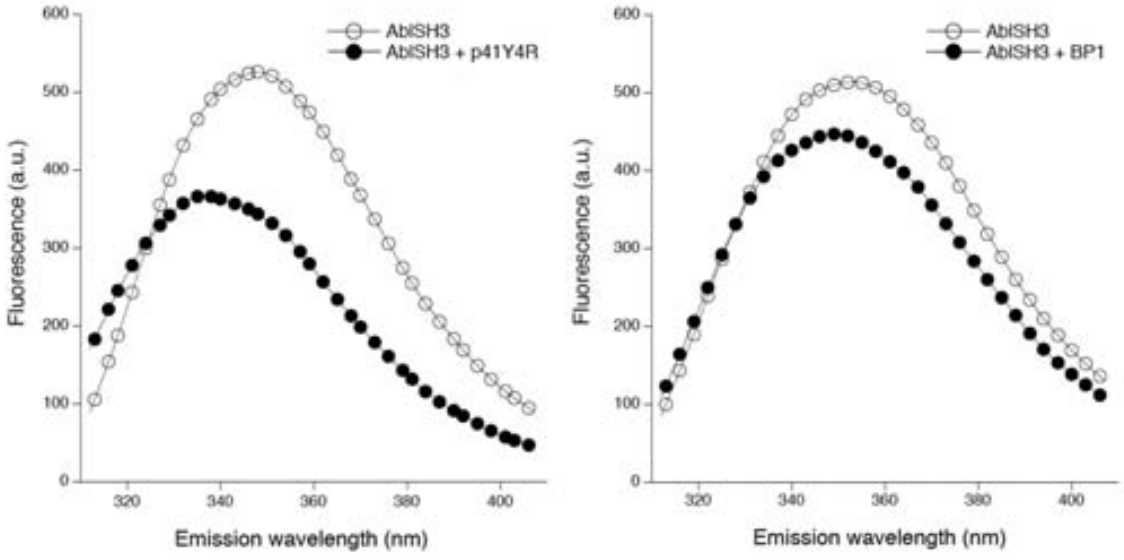


Figure 1.29 Fluorescence spectra of Abl-SH3 domain alone and complexed with p41Y4R (left) and BP1 peptide (right).

Although the affinity was ten times lower that the one reported for wild-type p41, it was still two times better than the one displayed by 3BP1 (43±2 μM in our hands) which

constituted a natural ligand of the Abl-SH3 domain[80]. This result stressed the ability of the method to easily detect weak interactions.

Surprisingly, the worst residue was Trp. This fact contrasts with the model ranking and with studies on related peptides that pointed Trp as a good residue according to its hydrophobicity and hydrogen bonding capability. It suggests that its low fluorescence could be a side-effect of the approach in this particular case. Western blot analysis indicates that this variant is expressed at lower levels than the rest of the fusions (Fig. 1.27). Also, the indole of Trp is the largest of the side chains of proteins, and it is likely that difficulties to accommodate this bulky amino acid in the binding surface could occur under some circumstances (as could be the fusion to the YFP fragment). Besides, undesired interactions of Trp with the N-terminus of the YFP, competing or impeding the interaction with the Abl-SH3 domain might occur. Any of these effects, or their combination could account for the observed low fluorescence emission of the Trp4-p41 variant.

1.4.2.5 Extending BIFC applicability

1.4.2.5.1 Native Electrophoresis

The stability of the reconstituted complex should permit to detect also weak interactions by native electrophoresis and subsequent imaging of in-gel YFP fluorescence. As shown in Figure 1.30, fluorescent bands could be detected in the gel only when Abl-SH3-CYFP and p41-NYFP were co-expressed. Also, the similar intensities of the lanes 3 and 4 demonstrated that it was not necessary to process the samples immediately as they could be frozen for subsequent analysis without significant loss of sensitivity.

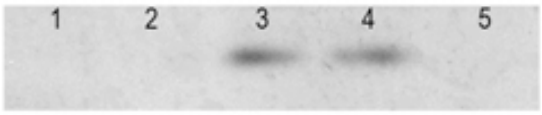


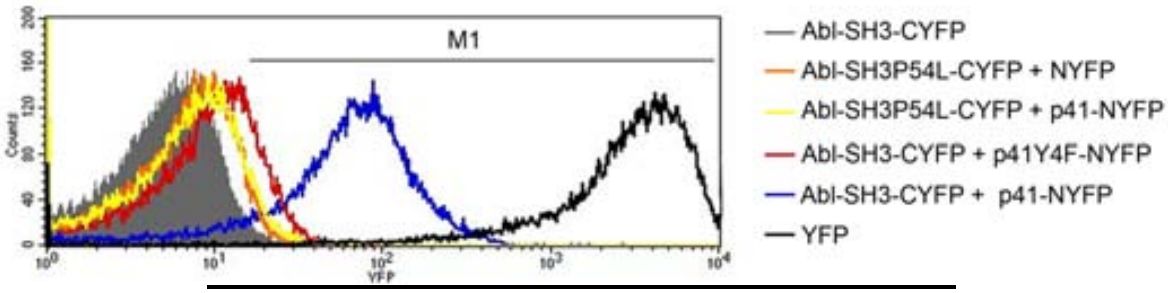
Figure 1.30 Native electrophoresis and in-gel UV imaging of YFP fluorescence. Native gel irradiated with UV light of the soluble intracellular fraction of *E. coli* cells expressing: (1) Abl-SH3-CYFP, (2) p41-NYFP, (3) Abl-SH3-CYFP and p41-NYFP, (4) Abl-SH3-CYFP and p41-NYFP (frozen previously) and (5) Abl-SH3-CYFP and NYFP.

1.4.2.5.2 Flow cytometry and cell sorting

Native electrophoresis provided an easy way to screen small libraries and subsequent individual recording of fluorescence emission permitted a qualitative evaluation of the interaction strength. Nevertheless, proteomic analyses require more high-

throughput approaches. In this sense, the coupling between BIFC and flow cytometry could have the potential of being a sensitive means of measuring weak protein interactions and simultaneously, it would allow dealing with large libraries providing a high-throughput sample rate and a qualitative evaluation of the interaction strength. It has to be taken into account that flow cytometry (FC) is a powerful method that allows the analysis of entire cell populations based on the characteristics of single cells flowing through an optical and/or electronic detection device. Modern flow cytometers are able to analyze thousands of cells every second in 'real time' and, when coupled to cell sorters, can actively separate and isolate cells with specific properties.

The interaction between Abl-SH3 and p41 peptide was used as a test case. First of all, in order to set the measure conditions of the flow cytometer, cells expressing native YFP and only the Abl-SH3-CYFP fragment alone were used as positive and negative controls, respectively. Also, cells co-expressing Abl-SH3-CYFP and NYFP fragment without the p41 peptide were analyzed to ensure that there was no significant spontaneous reconstitution of YFP from its fragments in the absence of binding. In this case, a homogenous population of cells with fluorescence emission just above of the negative control was detected (Figure 1.31).



Sample*	Mean Fluorescence
Abl-SH3-CYFP	6.10
Abl-SH3P54L-CYFP + p41Y4F-NYFP	7.64
Abl-SH3P54L-CYFP + p41-NYFP	8.12
Abl-SH3-CYFP + p41Y4F-NYFP	10.56
Abl-SH3-CYFP + p41-NYFP	79.23
YFP	80.90

*Sample name refers to the expressed protein fusions

Figure 1.31 Coupling BIFC to FC for the analysis and discrimination of transient binders**.** Cells expressing the different constructs were analyzed for yellow fluorescence emission. Frequency histograms of cells expressing: the prey fusion alone, Abl-CYFP (grey); native YFP (black); p41-NYFP and Abl-SH3-CYFP (blue); Abl-CYFP and p41Y4F-NYFP (red); Abl-SH3P54L-CYFP and p41-NYFP (yellow) and Abl-SH3P54L-CYFP and NYFP (orange). The mean value of each population is shown in the table below the histograms.

In contrast, cells co-expressing of Abl-SH3-CYFP and p41-NYFP were detected as a population with enhanced fluorescence below the positive control (cells expressing YFP). All these results could be quantified establishing a gate window of cells with a fluorescence mean above the negative control (gate window M1). Only 2% of the population of cells co-expressing Abl-SH3-CYFP and NYFP were found in M1; whereas 98% of cells co-producing Abl-SH3-CYFP and p41-NYFP belonged to this region. The mean fluorescence value of each population was also provided by the flow cytometer software.

To evaluate the ability of flow cytometry to distinguish between closely related binders with different affinity, cells expressing a mutation in one of the protein fusions (Abl-SH3-P54L-CYFP or p41Y4F-NYP) were also analyzed. In excellent agreement with the spectrophotometry data obtained in previous experiments[54], a homogeneous population of cells with strongly decreased fluorescence was detected relative to the one exhibited by cells containing the wild-type protein fusions. Analyzing the mean fluorescence, it could be concluded that this value could be taken as a measure of interaction strength because the emitted fluorescence was directly related to it.

Proteomic approaches to elucidate functional protein-protein interactions usually require the identification of a reduced number of good binding partners among a large population of non/bad-interacting polypeptides. To test if flow cytometry coupled to BIFC was suitable for such approach, a reduced number of cells (5 to 20%) expressing the wild-type Abl-SH3-CYFP and p41-NYFP fusions were mixed with a large excess of cells (80 to 95%) expressing Abl-SH3-CYFP and p41Y4G-NYFP (a low affinity binder). The analysis of the pool of cells rendered two populations with different mean. Strikingly, the number of cells exhibiting enhanced fluorescence corresponds precisely with the percentage of cells expressing the wild-type fusion in the mixture.

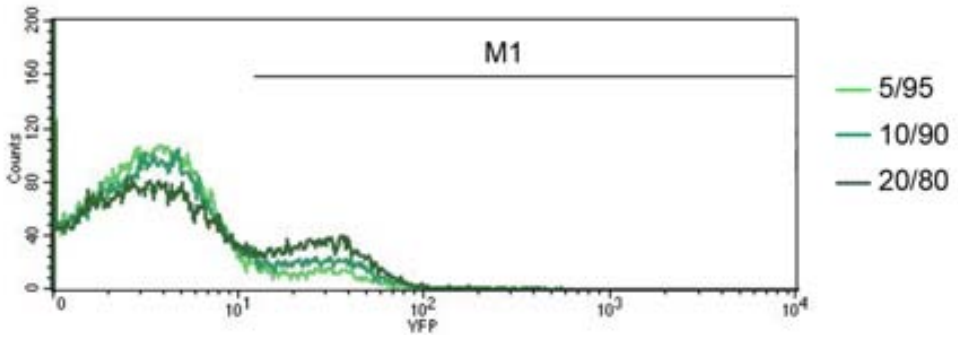


Figure 1.32 Frequency histograms of three populations of cells co-expressing the wild-type complex (Abl-SH3-CYFP and p41-NYFP) and the mutant complex (Abl-SH3-CYFP and p41-Y4G-NYFP) in different ratios shown in the legend.

With a fluorescence-activated cell sorting (FACS) device, cell sub-populations exhibiting increased mean fluorescence emission can be electronically deflected into separate collection tubes and sorted at high purity using slow sorting rates. We wanted to test if a FACS device could be useful to isolate cells with a higher mean fluorescence even though they were present in a low percentage (i.e. 3%).

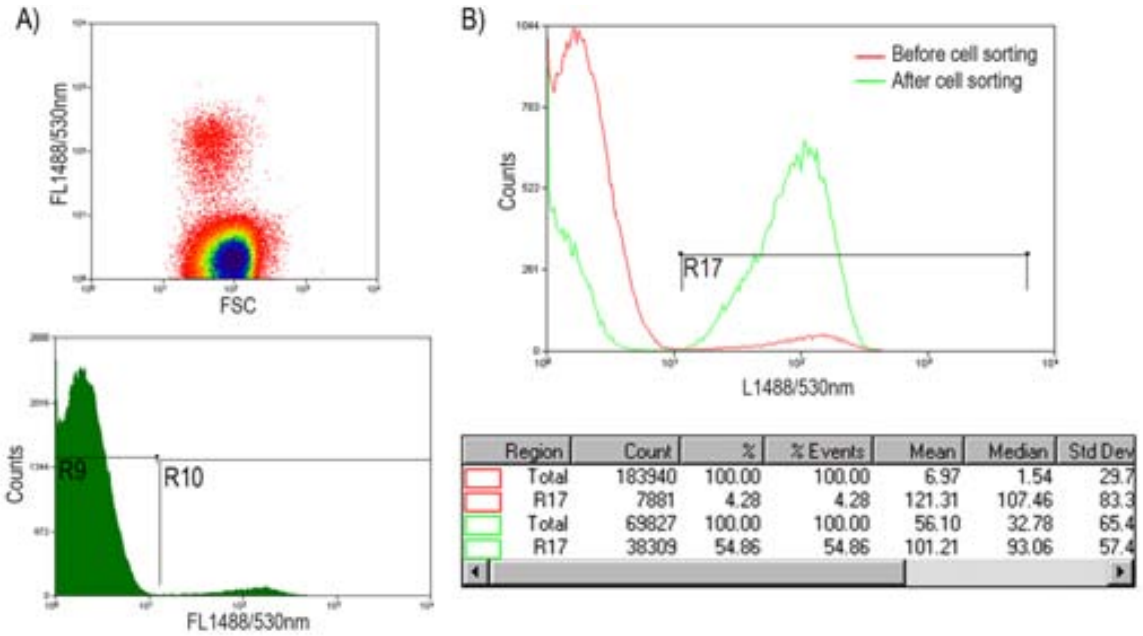


Figure 1.33 Cell sorting experiments**.** A) Frequency histogram of mixed population of cells co-expressing the wild-type complex (p41-NYFP and Abl-CYFP) and the mutant complex (Abl-CYFP and p41Y4F-NYFP) in the ratio: 97/3. The population of cells in R10 was the one expressing wild-type Abl-CYFP and p41-NYFP. B) Afterwards, sorted bacteria were reanalyzed to check the separation efficiency. Histograms obtained before (red) and after (green) the sorting are superposed. R17 region corresponds to bacteria expressing Abl-CYFP and p41-NYFP. The mean value of this subpopulation did not experiment any big change meanwhile the % of gated events increased.

One standard procedure in the cell sorting experiments is to reanalyze the sorted cells in order to ensure the success of the partition. In the Figure 1.33, it is shown the reanalysis of cells co-expressing Abl-SH3-CYFP and p41-NYFP that were separated from a population of cells co-expressing Abl-SH3-CYFP and p41Y4F-NYFP. These results suggested that the combined approach permitted to select good binders even if they were poorly represented in the global cell population. Overall, flow cytometry turned to be a fast, highly sensitive and discriminating technique for the evaluation of transient polypeptide interactions coupled to BIFC.

1.4.2.6 Exploring protein-protein interactions: the BRCA1 case

Poli-proline peptides are useful molecules to dissect the different forces contributing to SH3 binding, but their interaction does not necessarily reproduce the *in vivo* situation because, in the cell, interactions occur mostly between proteins. To assure the capabilities of BIFC to study weak physiologically relevant protein-protein interactions, the binding BRCA1 to the SH3 domain of Abl-tyrosine kinase was analyzed.

In human BRCA1, a PXXP motif (PQIP) is located just adjacent to the C-end of BRCT domains in a short and unstructured sequence stretch comprising the last 8 residues of BRCA1. This particular sequence had been proposed as the specific interaction site to Abl kinase because it coincided with the canonical SH3 domain binding sequence[86].

To confirm this extent, we fused the two BCRT domains of BRCA1 plus the C-terminal extension containing the PXXP motif (1638–1863) to the NYFP using a linker and co-expressed the fusion protein together with Abl-CYFP in *E. coli*. To our surprise, despite the fact that fusion proteins were expressed, cells fluorescence emission was very low, making imaging difficult. Nevertheless, the analysis of fluorescence emission spectrum in the fluorescence spectrophotometer allowed detecting a signal well above that of the controls, with the typical maximum at 523 nm, as expected for native YFP, confirming the correct reassembly of the reporter and thus, the interaction between proteins (Figure 1.34).

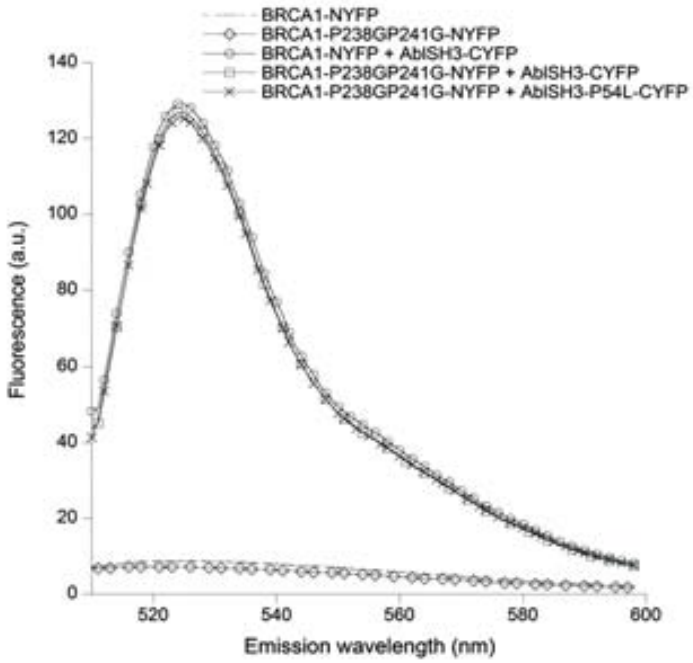


Figure 1.34 Fluorescence emission spectra of the soluble cell fraction from *E. coli* cells co-expressing Abl-SH3-CYFP and BRCA1-NYFP and different mutant forms of both fusions.

To assess whether BRCA1 actually interacts with Abl-SH3 domain *via* its PIQP motif, the two prolines were mutated to glycine. Cells expressing this mutated BRCA1NYFP version together with Abl-SH3-CYFP exhibited essentially the same fluorescence emission and spectrum as those having the wild-type sequence, indicating that the binding was independent of the existence of a PXXP motif. This was confirmed by the fact that the mutation P54L in the proline-rich peptide-binding site of the Abl-SH3 domain had no effect on the strength of the interaction, as proved by co-expression of BRCA1-NYFP and Abl-SH3P54L-CYFP and analysis of the resulting fluorescence emission spectrum (Figure 1.34). In conclusion, although the interaction between the two assayed domains seems to occur, it is rather weak and clearly not canonical. In agreement with the data, it has been shown that a region of about 300 residues at the N-side of the two BRCT domains significantly enhances the binding of these domains to Abl kinase[85].

1.5 Discussion

In this work, the use of BIFC method for the direct visualization of weak intracellular protein interactions has been described. The assay is sensitive enough to enable the detection of interactions between proteins that are poorly expressed in bacteria. It has to be taken into account that the visualization of the interactions directly in living cells eliminates potential artifacts associated with cell lysis or fixation. In addition, the interpretation of the fluorescent data is easy and there is no need for any complex data processing.

The BIFC method had been mainly applied to the study of strong protein-protein interactions. However, as it has been mentioned previously, the majority of interactions that play an important role in the cell are weak. In this work, it is shown that BIFC can easily detect SH3 domain interactions with K_d around 20 μM. Therefore, the method works for both strong and weak interactions.

The obtained results strongly suggest that the formation of the complex is mediated by specific contacts between the partners fused to YFP. Under the conditions of the assay, the binding appears to occur through the same protein regions involved in the native interaction and it is not driven by unspecific interactions between the fused YFP fragments, both requirements for the application of BIFC as protein-protein interaction detection method.

The specificity of the reassembly is exclusively due to the interaction between prey and bait proteins and it can be explained taking into account the high insolubility of the protein fusions and the irreversibility of the reassembly process. It has been demonstrated that the fragmentation of the fluorescent protein leads to two nearly insoluble proteins, even if they are linked to highly soluble proteins (e.g. leucine zippers)[53]. However, cells co-expressing both complementary protein fusions exhibit fluorescence by accumulating soluble and reassembled fluorescent protein complex in the cytoplasm. Probably, upon translation, a large fraction of the protein fusions aggregates due to improper folding while the rest is correctly folded and remains soluble. These soluble partners interact specifically and nucleate the reassembly of the fluorescent protein through an essentially irreversible process. Therefore, the protein fusions that had interacted become trapped, remaining soluble and pulling more of the insoluble fusions into solution by Le Chatelier's principle.

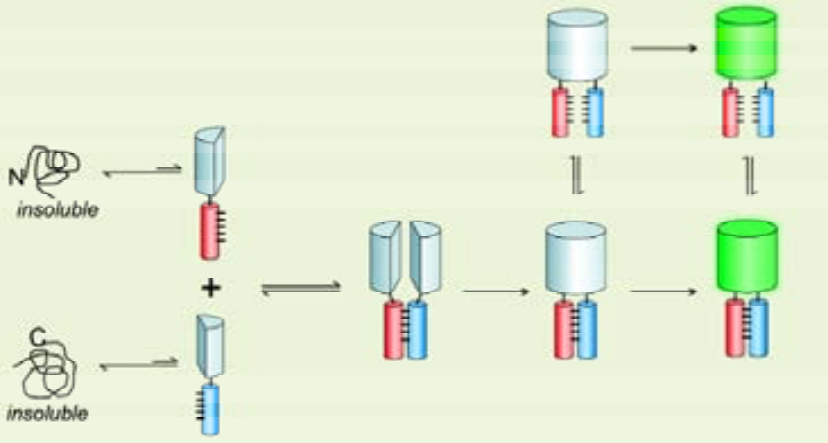


Figure 1.35 Mechanism of the Fluorescent Protein reassembly. Most of the protein fusions to the dissected fluorescent protein are insoluble. Interaction between the fluorescent protein fragments only occurs if they are fused to interacting proteins, which must nucleate the reassembly of the reaction, presumably from the small fraction of soluble fusion. Reassembly is essentially irreversible, which effectively pulls more of the fusions into solution. The protein-protein interaction is not necessary to maintain the reassembled complex.

The irreversibility of the reassembly implies that the method could have a rather general application on proteomics, since it acts as "trap" that could catch and immobilize transient interactions. The stability of the reassembled YFP would also explain the high sensitivity of the method because, once an interaction is trapped, the complex probably will become permanently able to emit fluorescence, as it is deduced from its stability against urea denaturation.

Evidences are collected that the fluorescence emission depends on the strength of the binding for weak interactions, as previously shown for strong ones. Thus, the approach has the potentiality to become a proteomic tool for screening libraries in order to find high affinity ligands, as illustrated here for position 4 of the p41 peptide. Nevertheless, the unexpected ranking of Trp in the screening suggests that, as in many other proteomic methods, one should be especially careful with false negatives that could arise from experimental difficulties.

On the other hand, the assay could be applied to map the regions involved in a given interaction without need of previous structural knowledge of the binding mechanism, as shown here for the interaction of Abl-SH3 domain with BRCA1. Our results suggested that the last 223 residues of the BRCA1 domain (including the BRCT tandem and the PIQP motif) play a rather moderate role in the constitutive binding to the SH3 domain in the Abl kinase allowing to explain why the human PXXP motif is not conserved in the BRCA1 proteins of other mammal species.

Finally, it has been demonstrated that BIFC could be coupled to flow cytometry and cell sorting. The combined approach results to be fast, highly sensitive and selective. This fact opens a new avenue for the proteomic analysis of intracellular weak protein-protein interactions. It also enables an easy way to identify targets via plasmid isolation of the positive cDNAS in the library, providing a general method for the identification and validation of target proteins involved in a given cellular process.

It is interesting to consider the advantages and disadvantages of BIFC when it is compared with well-established methodologies for detection of protein interactions. Like BIFC, FRET uses fluorescence emission as a reporter signal. Also, it has been successfully coupled to FC and it is useful for detecting and locating sites of protein interactions within cells. In contrast, complex dynamics can only be analyzed by FRET owing to the reversibility of the interaction between reporter proteins. However, the relatively small FRET dynamic range provided by fluorescent proteins limits its sensitivity. Moreover, it can be difficult to detect the interaction signal due to the background fluorescence resulting from direct acceptor excitation. This problem is avoided in BIFC where fluorescence does not occur in the absence of the interacting partners.

The coupling of TAP and mass spectrometry has been widely used to study protein-interaction networks. In this approach, interactions take place in the cellular environment but the need for cellular disruption and stringent purification steps might perturb native complexes and prevent the detection of weak or transient interactions.

BIFC shares with phage display technology its high-throughput sample rate as well as the ability to distinguish between protein binders with different specificity and/or affinity. However, display-based approaches are limited by their requirement for strong interactions and their *in vitro* context; these limitations do not apply to BIFC technology.

On the other hand, BIFC has in common with the Y2H assays the sensibility and the ability to detect weak interactions (with dissociation constants around micro molar range). However, Y2H requires the fusion proteins to be imported to the nucleus. Further, false positives caused by erroneous transcriptional activation of the reporter gene occur in Y2H.

An important point when considering a proteomic methodology is the occurrence of false positives and false negatives. In BIFC, false negatives could arise from steric constraints in fusion proteins or unspecific interaction of a binding partner with its fused YFP fragment that would prevent interaction with the target. Also, the inappropriate expression, folding or aggregation of the fusions could impede or decrease the

fluorescence emission. However these problems should not diminish its applicability in high-throughput protein-protein interaction studies because they are common to the Y2H or TAP approaches. Finally, one should expect a relative low ratio of false-positives because, YFP reconstitution and assembly, depends on partner association, even for weak interactions, as shown here.

For the implementation of BIFC into a standard proteomic high-throughput method, it will be required the construction of ordered arrays of strains expressing complete proteomes fused to YFP fragments, the use of selection methods based in fluorescence cell sorting as well as automation of target sequence identification. Nevertheless, overall the method has the potentiality to become a wide spread technology for the detection of binary protein-protein interactions and thus to contribute to the deciphering of intracellular biochemical pathways.

1.6 References

1. Piehler, J. New methodologies for measuring protein interactions in vivo and in vitro. *Curr Opin Struct Biol* **15**, 4-14 (2005).
2. Li, M. Applications of display technology in protein analysis. *Nat Biotechnol* **18**, 1251-1256 (2000).
3. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317 (1985).
4. Dunn, I.S. Assembly of functional bacteriophage lambda virions incorporating C-terminal peptide or protein fusions with the major tail protein. *J Mol Biol* **248**, 497-506 (1995).
5. Ren, Z.J. et al. Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4. *Protein Sci* **5**, 1833-1843 (1996).
6. Hoogenboom, H.R. et al. Antibody phage display technology and its applications. *Immunotechnology* **4**, 1-20 (1998).
7. Nedelkov, D. & Nelson, R.W. Surface plasmon resonance mass spectrometry: recent progress and outlooks. *Trends Biotechnol* **21**, 301-305 (2003).
8. Yan, Y. & Marriott, G. Analysis of protein interactions using fluorescence technologies. *Curr Opin Chem Biol* **7**, 635-640 (2003).
9. Lamken, P., Lata, S., Gavutis, M. & Piehler, J. Ligand-induced assembling of the type I interferon receptor on supported lipid bilayers. *J Mol Biol* **341**, 303-318 (2004).
10. Wilson, W.D. Tech.Sight. Analyzing biomolecular interactions. *Science* **295**, 2103-2105 (2002).
11. Fagerstam, L.G., Frostell-Karlsson, A., Karlsson, R., Persson, B. & Ronnberg, I. Biospecific interaction analysis using surface plasmon resonance detection applied to kinetic, binding site and concentration analysis. *J Chromatogr* **597**, 397-410 (1992).
12. MacBeath, G. & Schreiber, S.L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760-1763 (2000).
13. Haab, B.B. Advances in protein microarray technology for protein expression and interaction profiling. *Curr Opin Drug Discov Devel* **4**, 116-123 (2001).
14. Lueking, A. et al. Protein microarrays for gene expression and antibody screening. *Anal Biochem* **270**, 103-111 (1999).
15. Stoll, D. et al. Protein microarray technology. *Front Biosci* **7**, c13-32 (2002).
16. Joos, T.O. et al. A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* **21**, 2641-2650 (2000).
17. Zhu, H. et al. Global analysis of protein activities using proteome chips. *Science* **293**, 2101-2105 (2001).
18. Zhu, H. et al. Analysis of yeast protein kinases using protein chips. *Nat Genet* **26**, 283-289 (2000).
19. Templin, M.F. et al. Protein microarray technology. *Trends Biotechnol* **20**, 160-166 (2002).
20. Yang, Y., Wang, H. & Erie, D.A. Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. *Methods* **29**, 175-187 (2003).
21. Haustein, E. & Schwille, P. Single-molecule spectroscopic methods. *Curr Opin Struct Biol* **14**, 531-540 (2004).

22. Haustein, E. & Schwille, P. Ultrasensitive investigations of biological systems by fluorescence correlation spectroscopy. *Methods* **29**, 153-166 (2003).
23. Bacia, K., Majoul, I.V. & Schwille, P. Probing the endocytic pathway in live cells using dual-color fluorescence cross-correlation analysis. *Biophys J* **83**, 1184-1193 (2002).
24. Saito, K., Wada, I., Tamura, M. & Kinjo, M. Direct detection of caspase-3 activation in single live cells by cross-correlation analysis. *Biochem Biophys Res Commun* **324**, 849-854 (2004).
25. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).
26. Karimova, G., Dautin, N. & Ladant, D. Interaction network among Escherichia coli membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. *J Bacteriol* **187**, 2233-2243 (2005).
27. Vasavada, H.A., Ganguly, S., Germino, F.J., Wang, Z.X. & Weissman, S.M. A contingent replication assay for the detection of protein-protein interactions in animal cells. *Proc Natl Acad Sci U S A* **88**, 10686-10690 (1991).
28. Guo, D. et al. A tethered catalysis, two-hybrid system to identify protein-protein interactions requiring post-translational modifications. *Nat Biotechnol* **22**, 888-892 (2004).
29. Gingras, A.C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* (2007).
30. Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60**, 523-533 (2003).
31. Krogan, N.J. et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637-643 (2006).
32. Roda, A., Pasini, P., Mirasoli, M., Michelini, E. & Guardigli, M. Biotechnological applications of bioluminescence and chemiluminescence. *Trends Biotechnol* **22**, 295-303 (2004).
33. Wouters, F.S., Verveer, P.J. & Bastiaens, P.I. Imaging biochemistry inside cells. *Trends Cell Biol* **11**, 203-211 (2001).
34. Tsien, R.Y. & Miyawaki, A. Seeing the machinery of live cells. *Science* **280**, 1954-1955 (1998).
35. Tsien, R.Y. The green fluorescent protein. *Annu Rev Biochem* **67**, 509-544 (1998).
36. Day, R.N. Visualization of Pit-1 transcription factor interactions in the living cell nucleus by fluorescence resonance energy transfer microscopy. *Mol Endocrinol* **12**, 1410-1419 (1998).
37. Mahajan, N.P. et al. Bcl-2 and Bax interactions in mitochondria probed with green fluorescent protein and fluorescence resonance energy transfer. *Nat Biotechnol* **16**, 547-552 (1998).
38. Siegel, R.M. et al. Fas preassociation required for apoptosis signaling and dominant inhibition by pathogenic mutations. *Science* **288**, 2354-2357 (2000).
39. Xu, Y., Kanauchi, A., von Arnim, A.G., Piston, D.W. & Johnson, C.H. Bioluminescence resonance energy transfer: monitoring protein-protein interactions in living cells. *Methods Enzymol* **360**, 289-301 (2003).
40. Boute, N., Jockers, R. & Issad, T. The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol Sci* **23**, 351-354 (2002).

41. Michnick, S.W., Remy, I., Campbell-Valois, F.X., Vallee-Belisle, A. & Pelletier, J.N. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol* **328**, 208-230 (2000).
42. Pelletier, J.N., Campbell-Valois, F.X. & Michnick, S.W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci U S A* **95**, 12141-12146 (1998).
43. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc Natl Acad Sci U S A* **91**, 10340-10344 (1994).
44. Rossi, F., Charlton, C.A. & Blau, H.M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci U S A* **94**, 8405-8410 (1997).
45. Hu, C.D., Chinenov, Y. & Kerppola, T.K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**, 789-798 (2002).
46. Ding, D.Q. et al. Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library. *Genes Cells* **5**, 169-190 (2000).
47. Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R. & Wiemann, S. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* **1**, 287-292 (2000).
48. Kerppola, T.K. Visualization of molecular interactions by fluorescence complementation. *Nat Rev Mol Cell Biol* **7**, 449-456 (2006).
49. Hu, C.D. & Kerppola, T.K. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotechnol* **21**, 539-545 (2003).
50. Grinberg, A.V., Hu, C.D. & Kerppola, T.K. Visualization of Myc/Max/Mad family dimers and the competition for dimerization in living cells. *Mol Cell Biol* **24**, 4294-4308 (2004).
51. Anderie, I. & Schmid, A. In vivo visualization of actin dynamics and actin interactions by BiFC. *Cell Biol Int* **31**, 1131-1135 (2007).
52. Granneman, J.G. et al. Analysis of lipolytic protein trafficking and interactions in adipocytes. *J Biol Chem* **282**, 5726-5735 (2007).
53. Ghosh, I., Hamilton, A.D. & Regan, L. Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* **122**, 5658-5659 (2000).
54. Morell, M., Espargaro, A., Aviles, F.X. & Ventura, S. Detection of transient protein-protein interactions by bimolecular fluorescence complementation: the Abl-SH3 case. *Proteomics* **7**, 1023-1036 (2007).
55. Chen, B., Liu, Q., Ge, Q., Xie, J. & Wang, Z.W. UNC-1 Regulates Gap Junctions Important to Locomotion in C. elegans. *Curr Biol* (2007).
56. Fang, Y. & Spector, D.L. Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol* **17**, 818-823 (2007).
57. Zamyatnin, A.A., Jr. et al. Assessment of the integral membrane protein topology in living cells. *Plant J* **46**, 145-154 (2006).
58. Cole, K.C., McLaughlin, H.W. & Johnson, D.I. Use of bimolecular fluorescence complementation to study in vivo interactions between Cdc42p and Rdi1p of Saccharomyces cerevisiae. *Eukaryot Cell* **6**, 378-387 (2007).

59. Sung, M.K. & Huh, W.K. Bimolecular fluorescence complementation analysis system for in vivo detection of protein-protein interaction in Saccharomyces cerevisiae. *Yeast* (2007).
60. Mayer, B.J. SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253-1263 (2001).
61. Feng, S., Chen, J.K., Yu, H., Simon, J.A. & Schreiber, S.L. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* **266**, 1241-1247 (1994).
62. Kuriyan, J. & Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu Rev Biophys Biomol Struct* **26**, 259-288 (1997).
63. Lim, W.A., Richards, F.M. & Fox, R.O. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **372**, 375-379 (1994).
64. Macias, M.J. et al. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**, 646-649 (1996).
65. Mahoney, N.M., Rozwarski, D.A., Fedorov, E., Fedorov, A.A. & Almo, S.C. Profilin binds proline-rich ligands in two distinct amide backbone orientations. *Nat Struct Biol* **6**, 666-671 (1999).
66. Ren, R., Mayer, B.J., Cicchetti, P. & Baltimore, D. Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**, 1157-1161 (1993).
67. Cicchetti, P., Mayer, B.J., Thiel, G. & Baltimore, D. Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science* **257**, 803-806 (1992).
68. Yu, H. et al. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933-945 (1994).
69. Kang, H. et al. SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55. *Embo J* **19**, 2889-2899 (2000).
70. Kato, M., Miyazawa, K. & Kitamura, N. A deubiquitinating enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP. *J Biol Chem* **275**, 37481-37487 (2000).
71. Mongiovi, A.M. et al. A novel peptide-SH3 interaction. *Embo J* **18**, 5300-5309 (1999).
72. Fazi, B. et al. Unusual binding properties of the SH3 domain of the yeast actin-binding protein Abp1: structural and functional analysis. *J Biol Chem* **277**, 5290-5298 (2002).
73. Kami, K., Takeya, R., Sumimoto, H. & Kohda, D. Diverse recognition of non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and Pex13p. *Embo J* **21**, 4268-4276 (2002).
74. Woodring, P.J., Hunter, T. & Wang, J.Y. Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. *J Cell Sci* **116**, 2613-2626 (2003).
75. Wang, J.Y. Regulation of cell death by the Abl tyrosine kinase. *Oncogene* **19**, 5643-5650 (2000).
76. Puri, P.L. et al. A myogenic differentiation checkpoint activated by genotoxic stress. *Nat Genet* **32**, 585-593 (2002).
77. Renshaw, M.W., Kipreos, E.T., Albrecht, M.R. & Wang, J.Y. Oncogenic v-Abl tyrosine kinase can inhibit or stimulate growth, depending on the cell context. *Embo J* **11**, 3941-3951 (1992).
78. Brasher, B.B. & Van Etten, R.A. c-Abl has high intrinsic tyrosine kinase activity that is stimulated by mutation of the Src homology 3 domain and by

autophosphorylation at two distinct regulatory tyrosines. *J Biol Chem* **275**, 35631-35637 (2000).
79. Pisabarro, M.T. & Serrano, L. Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry* **35**, 10634-10640 (1996).
80. Pisabarro, M.T., Serrano, L. & Wilmanns, M. Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J Mol Biol* **281**, 513-521 (1998).
81. Williams, R.S., Green, R. & Glover, J.N. Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nat Struct Biol* **8**, 838-842 (2001).
82. Miki, Y. et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994).
83. Zheng, L. et al. Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1. *Mol Cell* **6**, 757-768 (2000).
84. Cortez, D., Wang, Y., Qin, J. & Elledge, S.J. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* **286**, 1162-1166 (1999).
85. Lee, J.S., Collins, K.M., Brown, A.L., Lee, C.H. & Chung, J.H. hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature* **404**, 201-204 (2000).
86. Foray, N. et al. Constitutive association of BRCA1 and c-Abl and its ATM-dependent disruption after irradiation. *Mol Cell Biol* **22**, 4020-4032 (2002).
87. Yu, X., Chini, C.C., He, M., Mer, G. & Chen, J. The BRCT domain is a phospho-protein binding domain. *Science* **302**, 639-642 (2003).
88. Foray, N. et al. Gamma-rays-induced death of human cells carrying mutations of BRCA1 or BRCA2. *Oncogene* **18**, 7334-7342 (1999).
89. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
90. Ormo, M. et al. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392-1395 (1996).
91. Van Etten, R.A., Debnath, J., Zhou, H. & Casasnovas, J.M. Introduction of a loss-of-function point mutation from the SH3 region of the Caenorhabditis elegans sem-5 gene activates the transforming ability of c-abl in vivo and abolishes binding of proline-rich ligands in vitro. *Oncogene* **10**, 1977-1988 (1995).
92. Cesareni, G., Panni, S., Nardelli, G. & Castagnoli, L. Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett* **513**, 38-44 (2002).
93. Rickles, R.J. et al. Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *Embo J* **13**, 5598-5604 (1994).
94. Weng, Z. et al. Structure-function analysis of SH3 domains: SH3 binding specificity altered by single amino acid substitutions. *Mol Cell Biol* **15**, 5627-5634 (1995).
95. Palencia, A., Cobos, E.S., Mateo, P.L., Martinez, J.C. & Luque, I. Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design. *J Mol Biol* **336**, 527-537 (2004).

CHAPTER 2

Monitoring the interference of protein-protein interactions by Bimolecular Fluorescence Complementation (BIFC)

2.1 Introduction

The dynamic processes of living organisms (DNA replication, gene regulation, transcription and splicing of mRNA, protein synthesis and subsequent secretion or many pathways associated with cell signalling) are mediated by protein-protein interactions[1]. Besides, dysfunctions in interactions can be responsible for the development of pathological processes, for example Alzheimer's and prion diseases[2, 3]. On the other hand, also interactions between virus-encoded components or between viral proteins and cellular factors occur during the replication and assembly of human viruses in host cells. Therefore, many of protein interactions could ultimately become potential drug targets.

The development of drug discovery strategies based on selective disruption of protein assemblies presents important advantages. First of all, the specificity inherent to protein interactions requires its interference being also highly selective. An inhibitor designed to bind an enzyme active site might have limited therapeutic applicability due to the structural similarities between the human and the pathogen catalytic mechanisms, whereas the greater structural variability of protein-protein interactions interfaces may provide an opportunity for the selective targeting of pathogen proteins. Secondly, single mutations in the reduced number of residues that conform active sites often lead to drug resistance. On the contrary, the large surfaces involved in protein-protein interactions difficult the appearance of resistance phenomena. Furthermore, at least in some cases, the alteration of the binding equilibrium due to the presence of an inhibitor could be sufficient to produce a significant biological effect without the need to completely inhibit the target protein-protein interaction.

Overall, the disruption of specific protein interactions could be a good strategy for target-focused therapeutical treatment.

2.1.1 Methods to detect antagonists of protein-protein interactions

Different methods have been developed in order to detect and evaluate specific antagonists of protein interactions. Most of them work *in vitro* and accordingly, there is an increasing necessity for new systems able to detect the interference of physiologically relevant protein interactions inside a cellular background.

2.1.1.1 In vitro assays

2.1.1.1.1 Capillary Electrophoresis coupled to laser-induced fluorescence detection

This technique requires incubation of the sample with fluorescence labelled compounds that act as affinity probes. Once these probes have interacted with the different components of the sample, the complexes are separated by capillary electrophoresis prior to a fluorescence detector. For the first time, it was applied to select and quantify the inhibitory effect of fluorescently labeled phosphopeptides over three different SH2 domains[4]. These domains were mixed together with the fluorescent peptides and thereafter separated by capillary electrophoresis (CE) coupled to laser-induced fluorescence detection (LIF). Peaks corresponding to the protein-peptide complex and free peptide were detected. Besides, inhibitors added to the mixture could be detected by their effect on the amount of complex detected. Because this method allowed separation of different SH2 protein complexes, a single inhibitor could be screened against multiple SH2 domain proteins to probe for inhibition and selectivity for a multiplexed assay.

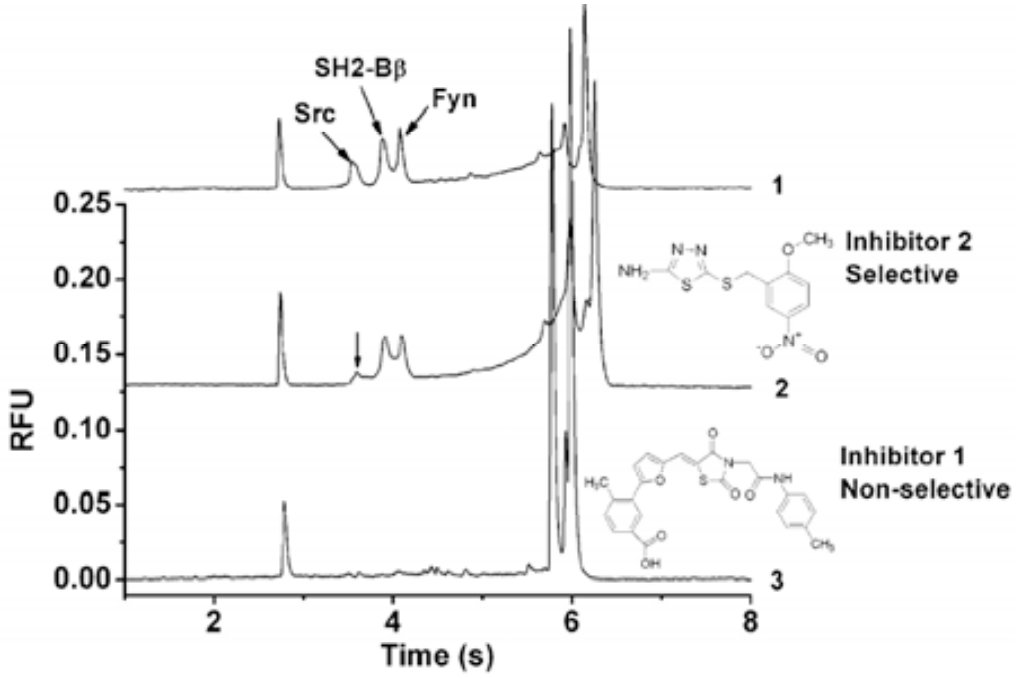


Figure 2.1 Selective and non-selective inhibition of SH2 domains (SRC, SH2-Bβ, and Fyn)**.** The analyzed sample contained different concentrations of the three SH2 domains, a fluorescently labeled peptide as affinity probe and an internal standard. For inhibition, inhibitor 1 or 2 was added to the sample. The inhibitor 2 was selective towards Src domain and caused a decrease in Src signal; whereas inhibitor 1 was non-selective causing the disappearance of all three signals.

CE has several properties that make it advantageous for this application: multinanalyte capability and no limitation to purified proteins or detection of individual complexes. Also no immobilization is required as the separation occurs in aqueous solution. Besides, it can be applied to a high-throughput screening.

2.1.1.1.2 Indirect immunolabeling and fluorescence cross-correlation spectroscopy

This assay is based in the antibody recognition of the interacting proteins and its subsequent visualization by fluorophore tagging. Primary antibodies directed against both proteins of interest and cognate secondary reagents labeled with spectrally different fluorophores are added in one step. The primary antibodies that in turn are recognized by the secondary reagents bind the correspondent proteins. When the interacting proteins form the complex, the secondary reagents are close enough to emit a fluorescent signal detected by fluorescence cross-correlation spectroscopy (FCCS)[5].

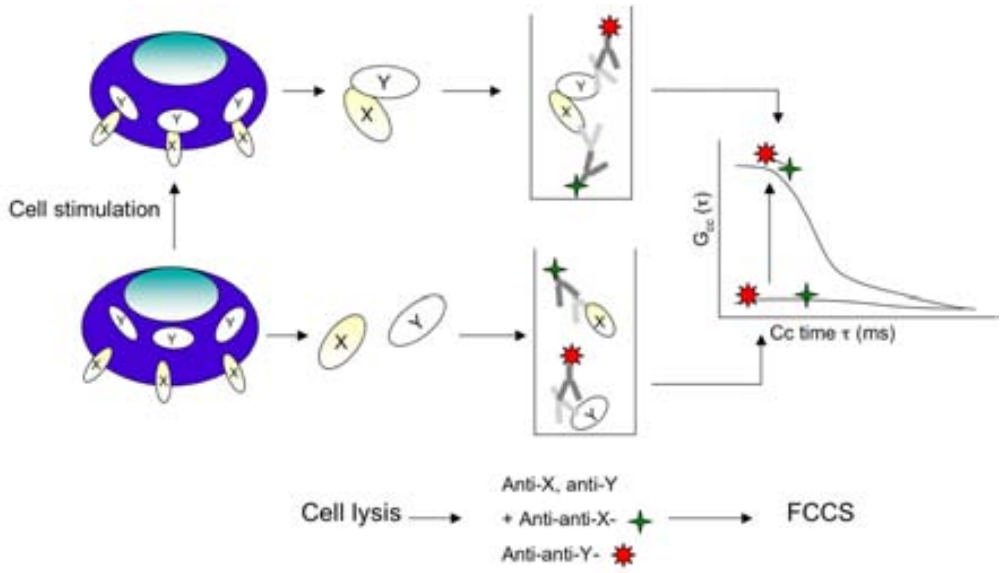


Figure 2.2 Overview of protein complex detection by indirect immunolabeling and FCCS**.** Primary and fluorophore-labeled secondary antibodies against the potentially interacting proteins X and Y are added to the lysate. After incubation, the FCCS-measurement reveals the presence of doubly labeled particles and hence the interaction of X and Y.

The procedure can be applied to few microliters of crude cell lysate but the requirement of specific primary antibodies for both the prey and bait proteins strongly restricts the general applicability of the method.

2.1.1.1.3 Luminescence Resonance Energy Transfer

Luminescence Resonance Energy Transfer (LRET) is a modification of FRET (explained in the previous chapter) using a lanthanide-based donor fluorophore (like Eu and Tb) and compounds like Cy5 (an europium chelate) as an acceptor[6]. Any acceptor dye with spectroscopic properties matching the emission of the lanthanide can be used in LRET. Comparing with FRET, LRET has prolonged fluorescent lifetimes. This fact causes a highly signal-to-noise ratio and higher sensitivity.

This methodology was applied for the first time to screen for antimicrobial drugs against crucial protein-protein interactions related to RNA synthesis. Specifically, the binding of sigma factors to the core of the RNA polymerase was studied because it is an essential process for the specific initiation of transcription in eubacteria and thus for cell growth. The interaction surface is highly conserved among eubacteria but differ significantly from eukaryotic RNA polymerases. Therefore, sigma factor binding is a promising target for drug discovery. In this experiment, the sigma factor σ70 was linked to a Eu chelating compound and a region of the β' subunit of the RNA polymerase was labeled with the compound IC5 (an acceptor dye equivalent to Cy5)[7]. And the inhibition of sigma binding was measured by the loss of LRET through a decrease in IC5 emission. The assay was applied to the detection of natural inhibitors in one hundred extracts of marine sponges. Particularly, one inhibitor was discovered with an IC_{50} of 1 μM.

2.1.1.2 In vivo assays

It has to be taken into account that *in vitro* selected candidates are not necessarily efficiently transported into cells, stable or selective enough to function in the complex context of the entire host proteome. In-cell selection approaches allow the simultaneous optimization of permeability, stability, affinity and specificity of the different tested compounds. Therefore, they promise to provide more potent and selective leads for novel chemotherapeutic compounds.

2.1.1.2.1 Reverse two hybrid approach

The majority of *in vivo* methods to detect interaction antagonists are based on the reverse two-hybrid approach. This assay is an upside-down version of the yeast two-hybrid: the interaction of the bait and prey proteins may be toxic or lethal for the yeast cells because of the toxicity of the reporter gene (e.g. URA3)[8].

In these conditions, the dissociation of the interaction confers a selective growth advantage that can be conveniently used to identify dissociating molecules.

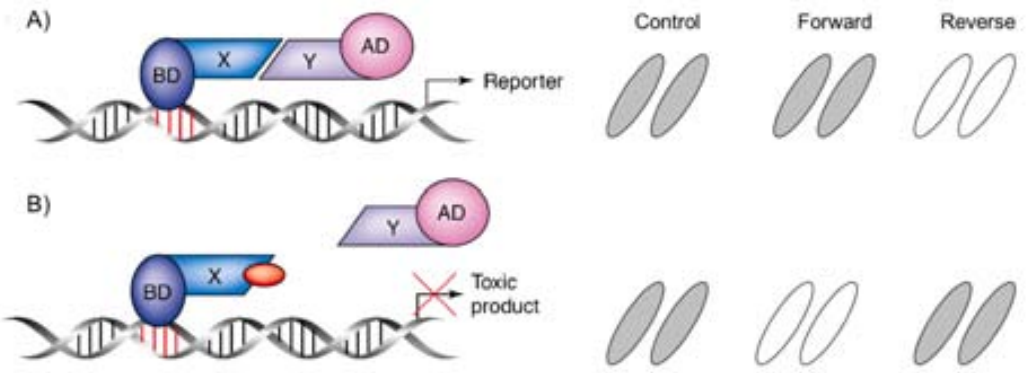


Figure 2.3 Two-hybrid systems. The grey and white patches on the right-hand side represent growing and non-growing yeast cells, respectively; under normal conditions (control), the yeast cells grow whether or not a two-hybrid interaction takes place. (A) In the forward two-hybrid selections, potential interactions are identified by the transcriptional activation of a reporter gene required for growth, which confers a selective advantage. (B) In the reverse two-hybrid selections, the interaction activates the expression of a 'toxic gene' and thus the prevention of the interaction provides a selective advantage.

A bacterial version of this method has also been developed based in the bacteriophage regulatory circuit[9]. Interestingly enough, this reverse two-hybrid in bacterial background can be co-compartimentalized in host cells with genetically encoded small-molecule libraries, which allows coupling of inhibition to DNA decoding.

2.1.1.2.2 Reverse mammalian protein-protein interaction trap

A reverse two hybrid inspired approach has been developed for mammalian cells: Reverse MAPPIT (Mammalian Protein-Protein Interaction Trap)[10]. It is based in the type I cytokine signalling pathway (Figure 2.4) and the readout is based in light measurement instead of cell survival. In this particular case, the interaction occurs in the cytoplasm whereas the reporter gene is activated in the nucleus. This is an advantage because nuclear translocation of the fusion proteins is not necessary preventing erroneous transcriptional activation of the reporter gene.

In the forward MAPPIT, the interaction between bait and prey protein results in the recruitment of a gp130 fragment containing STAT3 recruitment sites, thereby complementing the signaling-deficient chimeric receptor-bait. Activation of STAT3 is monitored by using the STAT3-responsive rat Pap reporter gene fused to luciferase. In contrast, in reverse MAPPIT, the recruitment of a prey protein fused to an inhibitory domain results in suppression of signaling via a functional chimeric receptor-bait.

Interference with the bait-prey interaction by a competitor protein or compound results in restoration of signalling, which is monitored by reporter gene induction

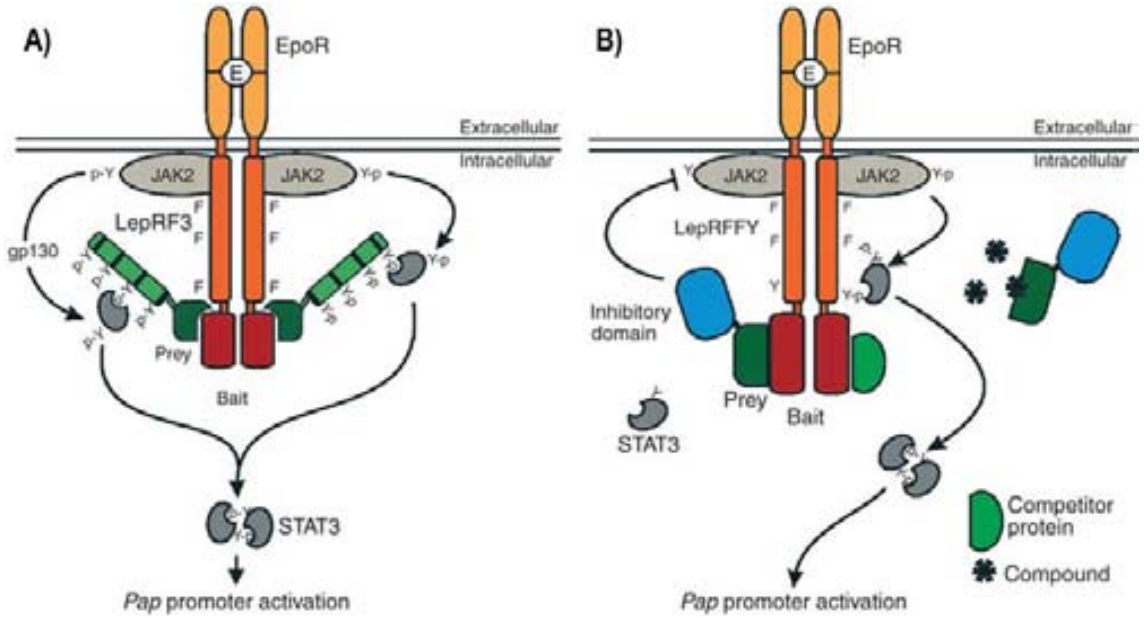


Figure 2.4. Principles of forward and reverse MAPPIT. (A) Forward MAPPIT. (B) Reverse MAPPIT (E, erythropoietin; EpoR, erythropoietin receptor; LepRF3 and LepRFFY, leptin receptor variants lacking or containing a functional STAT3 recruitment site, respectively; F, phenylalanine; pY, phosphotyrosine; JAK2, Janus kinase 2)[11].

2.1.2 BIFC applied to study the interference of protein interactions

In the previous chapter, the application of BIFC approach to the detection and study of weak protein interactions was demonstrated[12]. This sensitivity has been exploited recently to measure spatial and temporal changes in protein complexes in response to drugs that activate or inhibit particular pathways in living human cells[13, 14]. Sometimes a drug is effective against a certain disease or pathology without any information about the underlying mechanism by which the drug produces its effect. In other words, many times observations of new and useful properties of drugs are usually made by serendipity. The method is based in the principle that a cascade effect could cause that drugs that had an activity on one component in the pathway could also alter the amount or the localization of a downstream BIFC labelled complex.

The developed strategy consisted of six steps (Figure 2.5). First, various protein interactions were chosen to act as reporters of different relevant pathways (step 1) and their fusions to BIFC fluorescent protein fragments were created. Changes in protein complexes were measured in populations of cells grown on microtiter plates using microscopy (steps

2–4). The signal was extracted using different algorithms (steps 4-5), which were designed to quantify an increase, a decrease or a change of the fluorescence signal localization. Finally, the effect of each drug on each assay was tabulated and subjected to hierarchical clustering of drugs and assays (step 6).

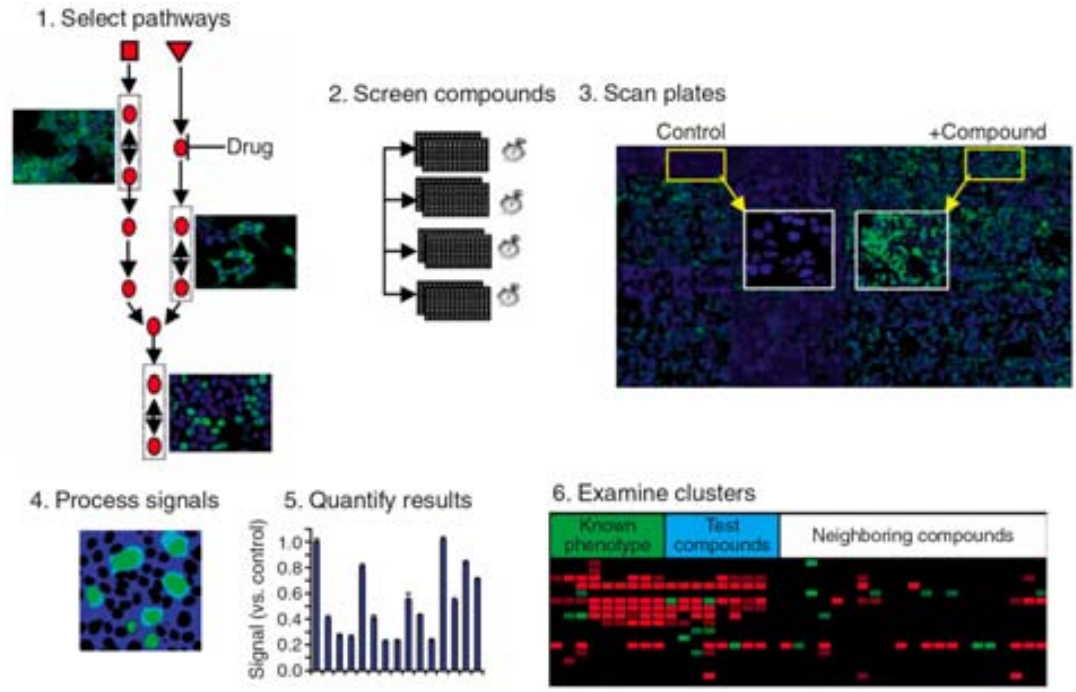


Figure 2.5 Strategy for pharmacological profiling of compounds using BIFC. (1) Pathways of interest were selected and BIFC assays were created on specific protein interactions. Assays measured dynamics of specific pathway by quantifying changes in protein complexes that were elicited in response to drugs. (2) Cells expressing BIFCs arrayed in 96-well plates were treated with compounds (3) Multiple images were captured. Pixel intensities from BIFC signals were extracted from one or several cell compartments (4) and tabulated for individual compound treatments (5). Data for each compound versus BIFC response at different times were represented as an array.

With this technique, the mechanism of action of four novel drugs was discovered. It has to be taken into account that the understanding of the drug action would enable optimization of a chemical structure to enhance desirable attributes and avoid undesirable ones. Clear differences in activity between closely related compounds in a structural class have been detected with this study, suggesting that minor chemical modifications might result in differences in compound activity in living cells. Therefore, these strategies may help to clarify the understanding of drug action and enhance the productivity of drug-discovery research.

The need for a native-like binding suggested that, in addition to monitor indirect drug effects on downstream complexes, BIFC could evolve into a tool for screening inhibitors of specific intracellular protein interactions. If the formation of the complex is

competed against other partners, fluorescent protein reassembly will be different and thus, also changes in fluorescence signal could be detected (Figure 2.6). Moreover, we have shown that the intensity of the fluorescence signal in BIFC depends on the protein interaction strength. Thus, in principle, BIFC might discriminate between modulators with different potency.

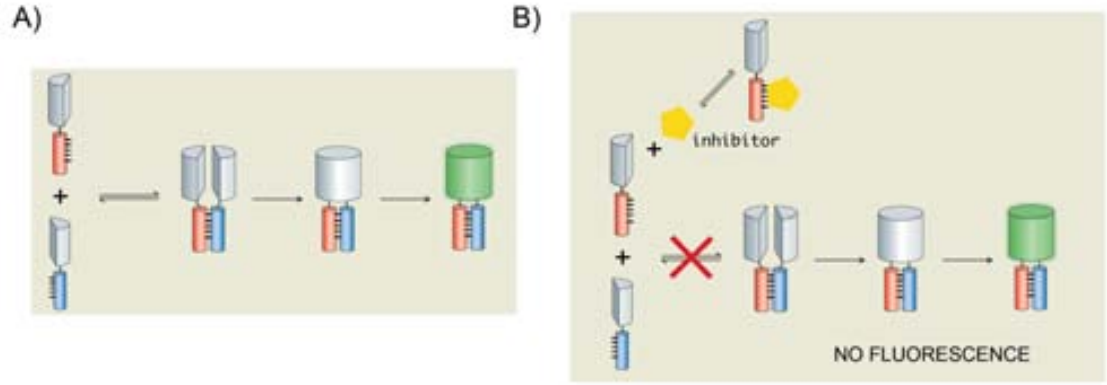


Figure 2.6 Schematic representation of BIFC and its application for detection of protein interaction inhibitors. A) BIFC can be applied as a method to detect protein-protein interactions. When the protein complex is formed between the bait and prey proteins, the reassembly of the YFP occurs and fluorescence is detected. B) When an inhibitor is present in solution, it binds with one of the interaction partners hindering the interaction between bait and prey proteins and thus the YFP reassembly. Consequently, no fluorescence is detected.

To test the above mentioned hypothesis, in the present work the BIFC method was applied to monitor the inhibition of the binding between *E.coli* Hsp70 chaperone and a short hydrophobic peptidic substrate by pyrrhocoricin derived antibacterial peptides.

2.1.3 Inhibition of DnaK chaperone activity by pyrrhocoricin

2.1.3.1 Pyrrhocorricin and its derivates

Resistance to antibiotics is developing at an alarming rate and this trend affects most antimicrobial drug families[15]. Therefore, this continuing spectre of bacterial resistance has driven a sustained search for new agents that possess activity against conventional antibacterial drug-resistant strains. One way to reach this goal would be the discovery and clinical development of an agent that acts over a new target, which has not yet experienced selective pressure in the clinical setting[16]. In addition, this target should be essential to the bacterial growth and survival, and sufficiently different from similar macromolecules in the human host. Among the most promising compounds that fulfill these requirements the small, proline-rich peptides originally isolated from insects stand

out. Apidaecin, drosocin, and pyrrhocoricin were suggested to kill bacteria by acting stereospecifically on a bacterial protein[17-19].

Pyrrhocoricin is a peptide originally isolated from the European sap-sucking *Pyrrhocoris apterus*[17]. It is non-toxic to eukaryotic cells and healthy mice and has good *in vitro* activity against model bacterial strains. Moreover, if it is administered intravenously *in vivo*, it can protect mice from systemic *E.coli* challenge[20, 21].

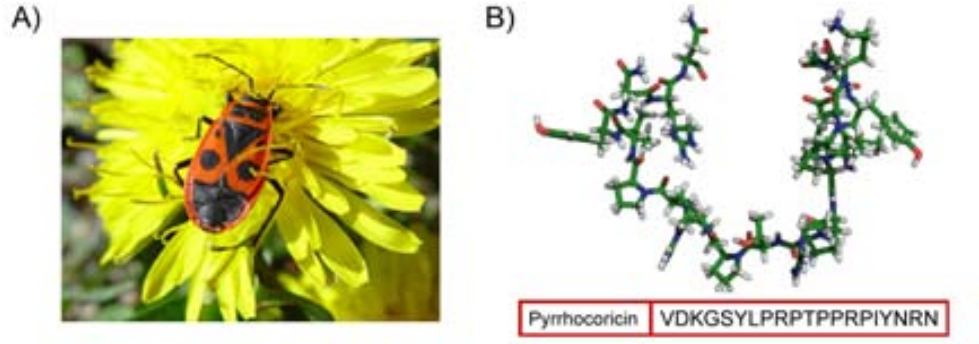


Figure 2.7 A) Picture of the sap-sucking bug *Pyrrhocoris apterus*. B) NMR structure of pyrrhocoricin peptide with its sequence displayed below

It has been suggested that the peptide kills sensitive species by binding to 70kDa bacterial heat shock protein (Hsp) DnaK and inhibiting its chaperone activity after penetration in *E.coli* cells[22, 23].

2.1.3.2 DnaK Chaperone

DnaK chaperone is a member of the highly conserved 70-kDa heat shock protein family (hsp70) with both constitutive and stress-induced functions. Hsp70 chaperones assist a large variety of protein folding processes in the cell by transient association with short peptide segments of proteins. These segments are not usually solvent exposed and are accessible to the chaperone only in non-native conformations. All cellular processes generating such conformers generate DnaK substrates including the *novo* protein synthesis[24], protein translocation[25], assembly and disassembly of protein complexes[26] and protein misfolding, especially under stress conditions[27].

DnaK consists of two domains: a 45kDa N-terminal ATPase domain and 25 kDa C-terminal substrate-binding domain. The ATPase domain contains the nucleotide-binding site in a channel between two lobes. The substrate-binding domain is composed of two subdomains, each with a characteristic structure and most likely functional properties. The substrate binding cavity is formed by a β-sandwich of two times four antiparallel β-strands

and four upward-protruding connecting loops[28] (Figure 2.8). It is followed by an α-helical subdomain that consists of five antiparallel helices. Helix B closes the substrate binding cavity without interacting directly with the bound substrate. The distal part of the helix B together with the helices C, D and E builds up a hydrophobic helical core that constitutes a lid-like structure.

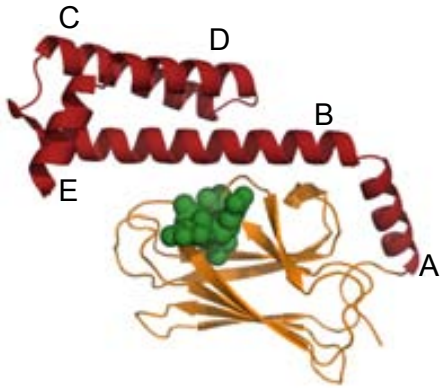


Figure 2.8 Crystal structure of the *E.coli* DnaK chaperone binding domain with the peptide NRLLLTG[28] The peptide is represented as green spheres, the red domain corresponds to the multihelical lid and the orange one, to the peptide binding cavity .

Several types of interactions contribute to substrate binding[28, 29]. Hydrogen bonds formed primarily between the backbones of two pocket-forming loops and the peptide backbone mediate recognition of the extended peptide conformation. Van der Waals interactions of side chains lining the substrate binding cavity with peptide side chains mediate DnaK's preference for hydrophobic residues. Hydrophobic contacts are possible to approx. five consecutive residues of the substrate. And the binding motif of the DnaK chaperone has been disclosed by screening of cellulose-bound peptides[30]. It is characterized by a core of four or five consecutive amino acid residues enriched in hydrophobic residues, especially Leu, and flanking regions enriched in basic residues. Negatively residues are disfavoured.

The substrate binding and release cycle is depending on ATP binding and hydrolysis. In the ADP-bound state, substrate binds with high affinity; but upon exchange for ATP, substrate affinity is substantially decreased with faster on/off-rates.

Some peptides that bind DnaK *in vitro* have been engineered from natural ligands, like the 22-residue pre-peptide of mitochondrial aspartate aminotransferase (PMAP)[31]. The peptide p5' (CALLLSAARR) corresponds to the central binding site of the pre-peptide with an additional non-natural amino-terminal cysteine residue and a fluorescent group

covalently attached. It interacts with DnaK with a K_d of 0.1 or 41.7 μM depending on the absence or presence of ATP[32].

The strength of this interaction is in the same range as the one between Abl-SH3 domain and the proline-rich peptide p41 (K_d = 5 μM) that was characterized using BIFC in the previous chapter. Therefore, the study of the *in vivo* binding of the chaperone DnaK to the peptide ALLLSAARR (p5*) might confirm the ability of BIFC to detect weak specific interactions.

DnaK is the bacterial homologous of human Hsp70 that has been shown to play an important role in the aggregation/disaggregation of disease-linked polypeptides[33]. To test if this capability is shared by the prokaryotic chaperone, we investigated the DnaK recognition of the Alzheimer's related peptide Aβ42 using BIFC.

2.1.3.3 Inhibition of DnaK chaperone activity by pyrrhocoricin

The binding interface of the interaction between DnaK chaperone and pyrrhocoricin peptide has been studied in depth in order to develop new pyrrhocoricin derivates displaying higher inhibitory activity. It is known that both termini N and C of pyrrhocoricin are need to kill bacteria because the isolated halves alone or their equimolar mixture are completely inactive[21]. Specifically, it has been determined that the N-terminus (1-10) serves as the pharmacophore and hence as interaction domain with the bacterial target protein, whereas the C terminal half aids the delivery of the peptide inside the cells. These results are in agreement with an alanine scanning realized in order to determine the crucial residues for pyrrhocoricin activity[34]. In this study, it was concluded that the essential region for the interaction is located between Asp2 and Pro10[34].

The binding site of the pyrrhocoricin peptide in the chaperone is still unknown. Nowadays, there are two theories regarding the possible interaction interface. The first one supports that pyrrhocoricin binds to the substrate binding site of the chaperone in a competitive inhibition mechanism[35]. It is based in the fact that pyrrhocoricin contains a classic DnaK-binding site. Also DnaK chaperone binds polications as pyrrhocoricin. Furthermore, this theory would explain the stereospecificity of pyrrhocoricin inhibition because DnaK does not bind peptides composed of all D-amino acids.

The second theory asserts that pyrrhocoricin binding site is located in the neighbourhood of the hinge between α-D and α-E helices affecting the chaperone ability to refold misfolded proteins[36]. First of all, the specific binding of pyrrhocoricin to a DnaK fragment containing α-D and α-E helices has been demonstrated[36]. In addition,

pyrrhocoricin inhibits *in vitro* the ATPase activity of DnaK which is allosterically modulated by the C-terminal lid domain, particularly by the α-D and α-E helices[37]. A model derived using a flexible docking approach suggests that the binding of pyrrhocoricin to this region would prevent the frequent opening and closing of the lid over the substrate binding site[36].

Figure 2.9 Characteristic structure of pyrrhocoricin and the C-terminal region of the *E.coli* DnaK as it was generated by the flexible docking process[36].The green domain corresponds to the multihelical lid, the yellow domain corresponds to the peptide binding cavity and the red structure corresponds to pyrrhocoricin

The model is in agreement with another study on the binding site of drosocin (a related proline-rich antibacterial peptide)[38]. *In silico* phylogenetic studies of DnaK in different bacterial species with distinct drosocin susceptibilities suggest that the susceptible bacteria share a high homologous sequence in the lid region of DnaK.

Even without an exact knowledge about the binding mechanism to the chaperone, the inhibition of DnaK chaperone by pyrrhocoricin would be a good test case to investigate the BIFC applicability to screen interaction modulators, using as a model the binding of the peptide ALLLSAARR (p5*) to the chaperone moiety.

2.2 Objectives

→ Application of BIFC approach to detect the interference of protein interactions. Specifically, it will be studied the pyrrhocoricin inhibition of the interaction between the chaperone DnaK and a peptidic substrate (p5*).

→ Determine the BIFC capabilities to screen compounds with difference inhibitory capacity.

→ Analyze the coupling between BIFC and flow cytometry to screen inhibitors of protein-protein interactions.

2.3 Experimental procedures

2.3.1 Construction of the protein fusions

The table 2.1 summarizes the protein fusions designed for the experiments. In order to obtain the different protein fusions, it was used the strategy described in the general Experimental Procedures section. And, in the Figure 2.10, it is shown a scheme of the constructed plasmids.

Table 2.1 Design of the fusion proteins

Protein	Fluorescent protein fragment	Linker sequence	Plasmid	Restriction sites	Antibiotic resistance
DnaK	CYFP (156-238)	SGGGSGGS	pBAT4	*NcoI, HindIII*	Ampicillin
Peptide p5*	NYFP (1-155)	SGGGSGGS	pET28a(+)	*NdeI, BamHI*	Kanamycin
Peptide Aβ42	NYFP (1-155)	SGGGSGGS	pET28a(+)	*NdeI, BamHI*	Kanamycin

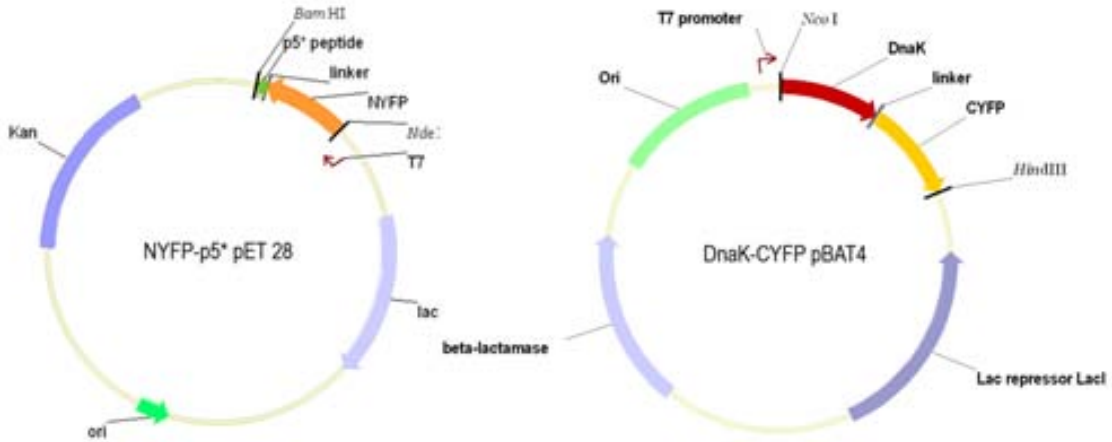


Figure 2.10 Scheme of the used plasmids. Plasmid pET28 encoding the protein fusion NYFP-p5*. This design was also followed in the case of NYFP-peptide Aβ42. Plasmid pBAT4 encoding the protein fusion DnaK-CYFP.

The DnaK chaperone (residues 385-638) was amplified from genomic DNA of *E.coli* and the CYFP fragment (156-238), from enhanced EYFP cDNA (Clontech). On the other hand, the DNA encoding for p5* was created by direct annealing of two synthetic complementary DNA oligonucleotides. And Aβ42 was amplified directly from a vector previously described[39]. The constructions Abl-SH3-CYFP and p41-NYFP were obtained as it has been detailed in the first chapter.

2.3.2 Peptide synthesis.

Peptides were synthesized on solid-phase using standard Fmoc/tBu-chemistry and *in situ* activation with 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) in the presence of N,N'-di-iso-propylethylamin (DIPEA). All peptides were cleaved with TFA and purified by RP-HPLC using a linear acetonitrile gradient in the presence of 0.1% TFA. The purity of the peptides was measured by analytical RP-HPLC using a Jupiter C18-column (Phenomenex Inc.) and their sequence was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS; 4700 proteomic analyzer; Applied Biosystems).

2.3.3 Culture media and growth conditions.

Competent *E.coli* BL21(D3) cells were transformed using heat shock protocol with compatible plasmids encoding the fusion proteins DnaK-CYFP and NYFP-p5* and plated on LB plates with ampicillin (50 μg/ml) and kanamycin (35 μg/ml). They were grown overnight at 37ºC.

Next day, an overnight culture was set up picking a single colony from the LB agar plate into 1ml liquid LB containing both antibiotics. The culture was incubated at 37ºC overnight at 250 r.p.m. In the morning, 1ml of the culture was centrifuged for 5 min at 1,300x*g* and the supernatant was removed. Then, the cell pellet was resuspended in fresh LB. The culture is afterwards diluted 1:100 in LB containing the appropriate antibiotics. This culture is incubated at 37ºC and 250 r.p.m. until the A_{600}=0.6. Then, the pyrrhocoricin analogs are added to a final concentration of 13 μM. In the control positive cells, the same volume of distilled and sterile water was added. The cells were grown at 30ºC for one hour. After that, the temperature was decreased at 18ºC and 30 minutes later, protein expression was induced by adding IPTG (isopropyl-β-D thiogalactopyranoside) to a final concentration of 1mM. The cultures were incubated at 18ºC and 250 r.p.m. during 16 hours.

It has to be taken into account that the peptide concentration in the culture medium was below the lowest antimicrobial dose in LB medium (30 μM) and no significant changes in cell density related to peptide antibacterial activity were observed.

The incubation of the pyrrhocoricin as well as the induction of the protein expression could be performed in a 96-well plate (non treated black microwell from Nunc). This strategy was followed to monitor the BIFC signal in concentration- and time-

dependent manner. A culture of cells co-transformed with plasmids encoding both protein fusions were grown following the steps described above. Once A_{600nm}= 0.6, it was fractioned in a 96-well plate. Specifically, in each well, 200 µl of the culture were added. Afterwards, the required volume of pyrrhocoricin was also added and the plate was incubated at 30ºC for one hour. After that, the temperature was decreased at 18ºC and 30 minutes later, protein expression was induced by IPTG (1mM final concentration). The cultures were incubated at 18ºC and 250 r.p.m. In order to measure the absorbance of the culture, the plates were read using 1420 VICTOR^{3}™ from Perkin Elmer.

To study the irreversibility of the complex, protein expression was induced when the A_{600} = 0.6 and cells were grown at 18ºC for 48 hours until a stable fluorescence signal was attained. Then, pyrrhocoricin derivates were added to a final concentration of 13µM and the fluorescence signal was measured after 5 hours of incubation (at 18ºC and 250 r.p.m).

2.4 Results

2.4.1 Design strategy

As it has been mentioned in the previous chapter, a crucial step in any BIFC assay is the design of the fusions between the interacting proteins and the respective EYFP fragments; design considerations include which fragment is fused to which partner and the primary sequence orientation of the fusions. In our case, it was taken into account that the fusion of DnaK with the smaller C-terminal fragment of the fluorescent protein (CYFP) and p5* peptide, with the larger N-terminal moiety (NYFP) generated fusions of limited size. In addition, false negatives due to wrong spatial disposition or steric impediments between EYFP fragments had to be avoided. Therefore, based upon the crystal structure of a complex between the substrate-binding domain of DnaK and a peptidic substrate[28], the fusion of the chaperone to the N terminus of CYFP (DnaK-CYFP) and the peptide to the C terminus of NYFP (NYFP-p5*) was considered the best option.

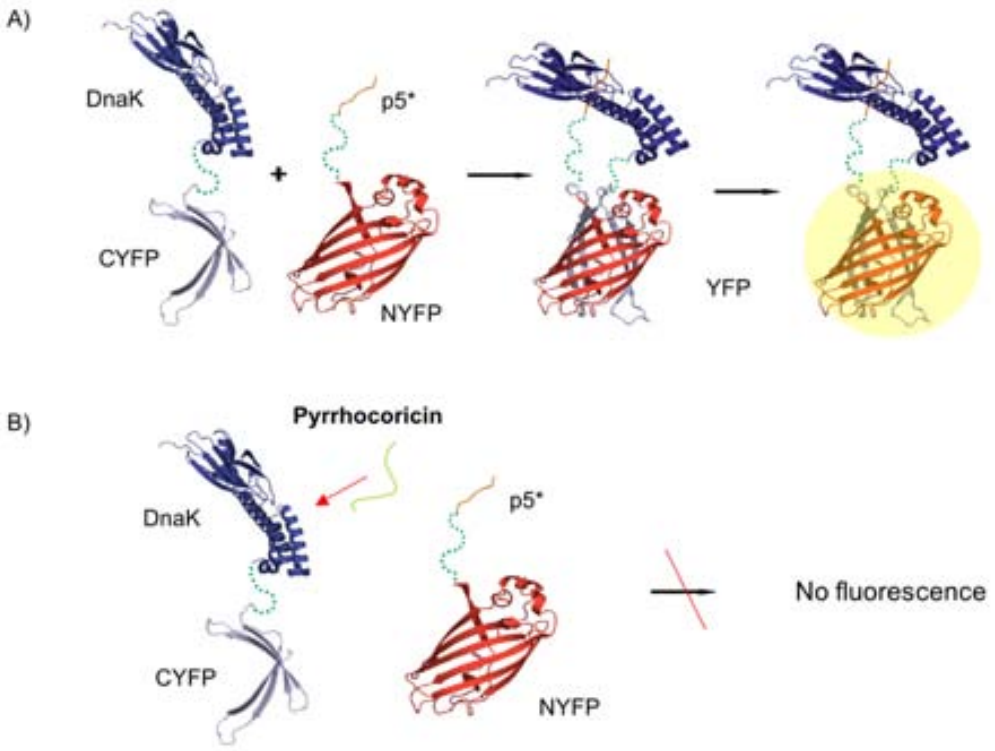


Figure 2.11 Schematic representation of BIFC and its application for detection of interaction inhibitors. Models of DnaK chaperone and p5* peptide fusion proteins are based on the X-ray structure of the complex between DnaK and the peptide NRLLLTG[28] and the crystal structure of GFP[40]. The structure of the linkers connecting the fused proteins is unknown.

2.4.2 Detection of DnaK chaperone interactions by BIFC

The first step was to check the capability of BIFC method to detect DnaK chaperone interactions *in vivo*. As it was performed in the previous chapter, BIFC was applied to two different DnaK interactions. The first one was with a designed ligand (p5* peptide); and the second one involved a natural ligand: the amyloid peptide Aβ42 in order to confirm this binding *in vivo*.

2.4.2.1 Interaction of DnaK with a synthetic ligand: p5 peptide*

The first step was the detection of the interaction between the DnaK chaperone and the p5* peptide by BIFC approach. In the BIFC approach, it is important to design controls to ensure that the fluorescence signal is caused by the selective binding of the prey and bait proteins and not mediated by unspecific contacts. With this purpose in mind, we double-checked the binding specificity by monitoring the interaction of DnaK with an hydrophobic peptide (p41 peptide) that does not fit in its active site according to the crystal structure of the chaperone as well as the interaction of p5* with the Abl-SH3 domain, a protein that recognizes hydrophobic peptides but requires a specific proline-rich motif for binding (absent in p5*)[12, 41]. DnaK-CYFP was expressed simultaneously with the complementary protein fluorescent fragment (NYFP) fused to p41 peptide. The same experiment was also performed in the case of NYFP-p5* to test its binding to the Abl-SH3 domain fused to the CYFP fragment. Besides, protein fusions alone were also expressed in order to check that each fluorescent fragment alone did not display any fluorescence under the experimental conditions. Whereas the cells expressing the swapped protein fusions (Abl-SH3-CYFP+NYFP-p5* or DnaK-CYFP+p41-NYFP) did not exhibit any fluorescence under the microscope, *E.coli* cells co-transformed with the plasmids encoding for NYFP-p5* and DnaK-CYFP presented fluorescence signal (Figure 2.12).

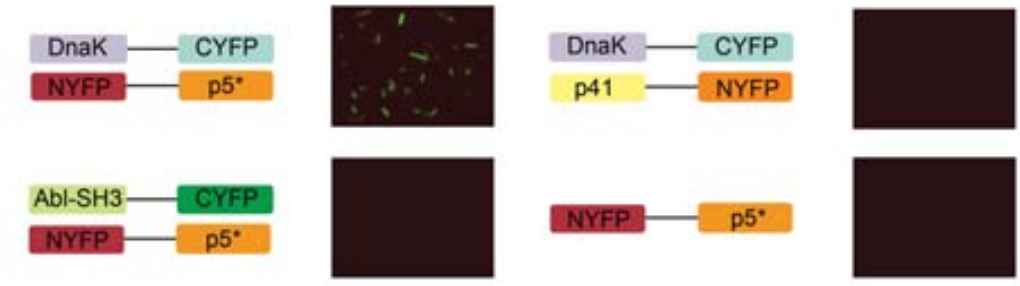


Figure 2.12 Visualization of DnaK chaperone-p5*peptide interaction using BIFC. Fluorescence images of *E.coli* induced cells expressing the protein fusions indicated beside.

This fact emphasized the requirement for the specific interaction between the chaperone and p5* peptide in order to obtain a positive readout.

In addition, the fluorescence signal was quantified using fluorescence spectrophotometry (Figure 2.13). The obtained results agreed with the microscopy images: only the cells expressing DnaK-CYFP and NYFP-p5* showed a fluorescence signal. Again the typical maximum at 527 nm coincided with the native EYFP emission wavelength, confirming the correct reassembly of the reporter.

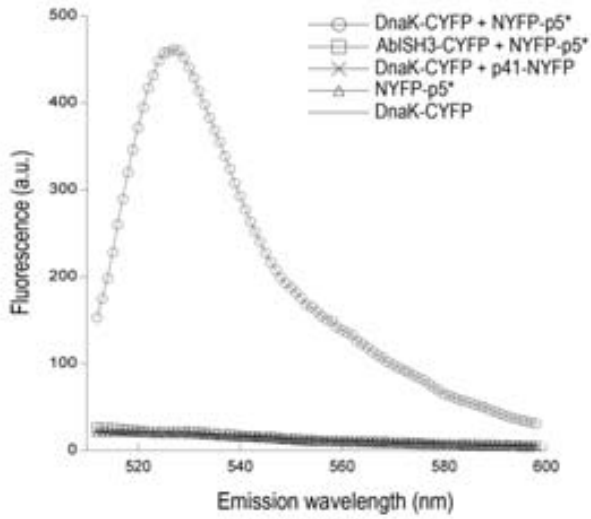


Figure 2.13 Fluorescence emission spectra of the *E.coli* cells expressing the fusions indicated in the legend.

Also, it has to be stressed that in both experiments cells transformed with either DnaK-CYFP or NYFP-p5* alone did not exhibit any fluorescence demonstrating again that EYFP fragments are not individually functional.

2.4.2.2 Interaction of DnaK with a disease-linked ligand: amyloid peptide Aβ42

Aβ42 is an amyloidogenic peptide that plays a central role in Alzheimer's disease. The deposition of insoluble protein fibrils is associated with a number of neurodegenerative diseases[42]. Specifically, in Alzheimer's disease, the primary component of these amyloid fibrils is the amyloid-β peptide (Aβ)[43, 44]. Aβ is generated *in vivo* by sequential proteolytic cleavage of the amyloid precursor protein (APP) by two proteolytic enzymes (γ and β secretases)[45]. The most abundant cleaved forms found in amyloid plaque are a 40-mer and a 42-mer segments (Aβ40 and Aβ42). Although Aβ40 is produced in greater abundance, the slightly longer Aβ42 is more amyloidogenic[44, 46]. In a previous

report, it was demonstrated that DnaK hinders the aggregation induced by Alzheimer's β amyolid fibrils[47]. Besides, in a proteomic study, Hsp70 and other chaperones were identified as highly expressed proteins in normal pancreatic mouse islet[48]. Moreover, another studies suggest a protective role of chaperones Hsp70 in front of Aβ aggregation[49-51]. All these results underscored the importance of the protein quality control network in these metabolically active cells and gave indications of a possible interaction between the Aβ42 peptide and DnaK.

Using the fusion between DnaK chaperone and C-terminal fragment of YFP as a bait, the peptide Aβ42 was fused to the C-terminus of the NYFP fragment with the same topology as p5* peptide. *E.coli* cells were co-transformed with both fusions and microscopy images as well as fluorescence spectra were obtained (Figure 2.14).

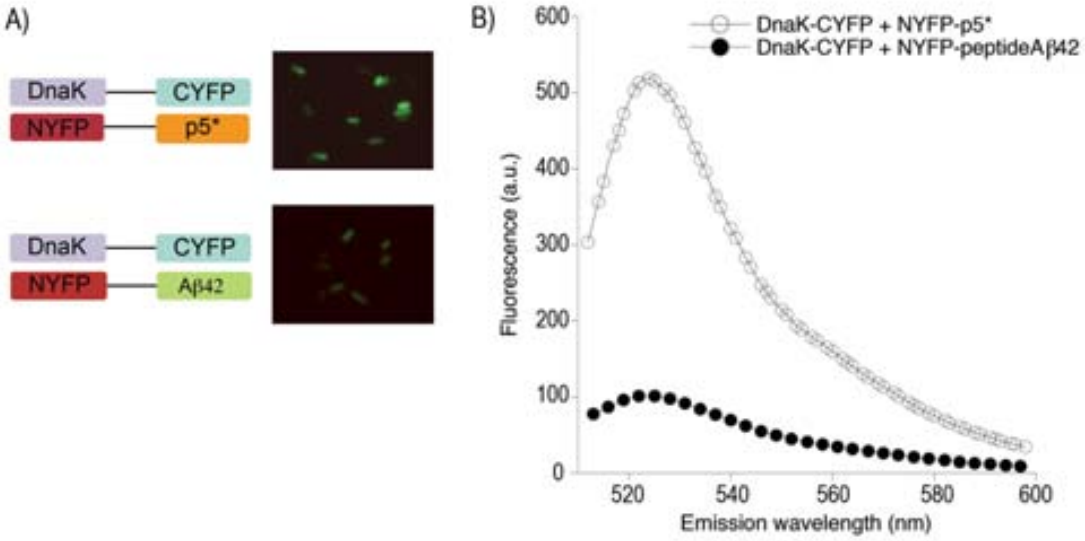


Figure 2.14 Visualization of DnaK chaperone/peptide Aβ42 interaction using BIFC. A) Fluorescence images of *E.coli* induced cells expressing the protein fusions indicated beside B) Fluorescence emission spectra of the *E.coli* cells expressing the fusions indicated in the legend

The signals detected with the microscope and the spectrophotometer were weaker than the ones acquired in the case of the interaction with the p5* peptide. However, the specific binding of Aβ42 peptide and DnaK could be detected. Therefore, the interaction between an amyloidogenic peptide and the chaperone DnaK could be confirmed *in vivo* using BIFC technique.

2.4.3 Interference of L-pyrrhocoricin with the DnaK-p5* interaction

Once the specificity of DnaK-p5* interaction was demonstrated, the next step was to explore the BIFC ability to detect molecules that interfered with the interaction between

DnaK and p5* peptide by blocking the chaperone. The experimental procedure was the same followed in the previous section, but in this case, an incubation step with the inhibitor was be performed before the addition of IPTG and thus previous to the recombinant protein expression.

According to the *in vitro* results obtained in the first chapter, EYFP reassembly traps the interaction between binding partners, in such a way that either the addition of competitors or the disruption of the interaction by chemical agents has little effect on fluorescence emission. On one hand, this fact is of great advantage to detect transient interactions. On the other hand, it would avoid the detection of interaction antagonists, because, once the interaction has occurred and EYFP has reassembled, the presence of compounds able to compete favourably with one of the binding partners would have little effect on the fluorescence emission. This limitation can be circumvented, in principle, by adding the modulator before protein expression, in such a way that it can penetrate into the cell and have a chance of impeding the reporter protein reassembly. This could be helpful in the screening of biologically relevant antagonists: a pre-incubation step should preclude the selection of high affinity inhibitors that were unable to penetrate efficiently or be stable within the cellular environment.

Both efficient entry and widespread distribution of pyrrhocoricin in bacterial cells have been demonstrated previously using fluorescein-labeled peptides[34]. This permits the direct addition of the peptide to the culture before the induction of DnaK-CYFP/NYFP-p5* recombinant expression.

As a negative control, a pyrrhocoricin analogue made all of D-amino acids (D-pyrrhocoricin) was used. Due to the stereospecificity of the interaction, this peptide neither binds DnaK *in vitro* nor has antibacterial activity, despite its efficient penetration into cells[36]. As reported in the previous section, cells expressing both partners (DnaK-CYFP and NYFP-p5*) in the absence of any forms of pyrrhocoricin exhibited fluorescence (Figure 2.15A). The presence of D-pyrrhocoricin in the media did not change significantly cell fluorescence emission. Instead, when the active L-pyrrhocoricin was present, a significant reduction in fluorescence emission was observed by microscopy or by recording the fluorescence spectra of the cell cultures (Figure 2.15).

However, these differences in fluorescence emission could be due to pyrrhocoricin effect over the cell growth or to different expression levels in any of the two partners: DnaK-CYFP or NYFP-p5*. The pyrrhocoricin concentration used in these assays was beforehand proved to inhibit DnaK activity without any effect upon the cells[36]. In this

sense, by measuring the culture optical density, no significant effect upon cell growth was observed in cells exposed to D- or L-pyrrhocoricin. Besides, the levels of recombinant DnaK-CYFP and p41-NYFP did not change significantly in the presence of the peptide as it was shown by western blot (Figure 2.15B).

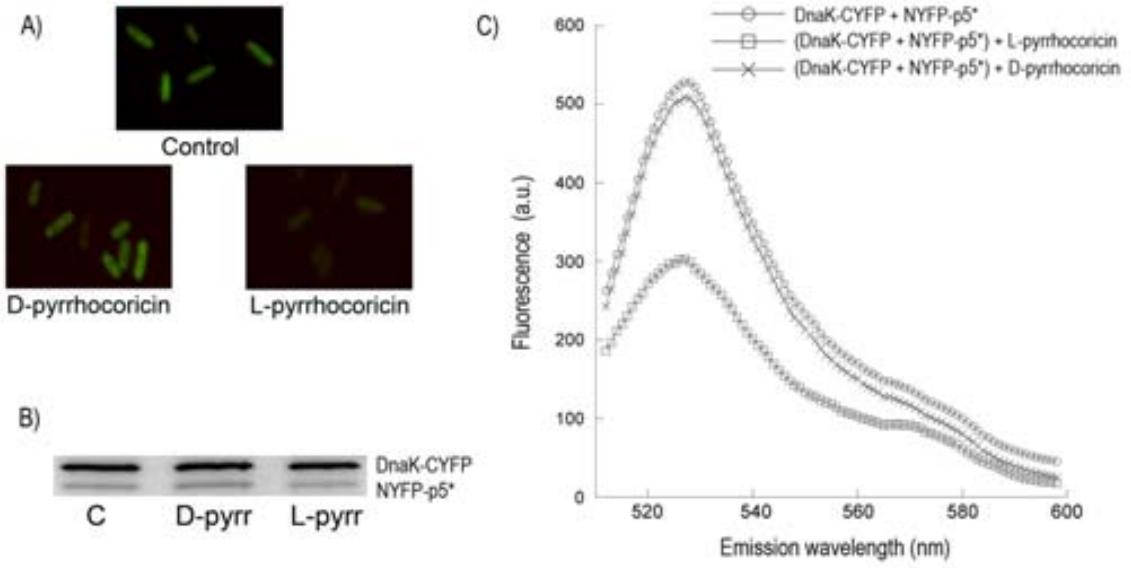


Figure 2.15 Detection of the disruption of DnaK chaperone/p5* peptide interaction using BIFC. (A) Fluorescence images of *E.coli* induced cells transformed with the plasmids encoding for DnaK-CYFP and NYFP-p5* and incubated with the peptides indicated in each case. (B) Western blot analysis of cells expressing DnaK-CYFP and NYFP-p5* and incubated with different pyrrhocoricin derivates. (C) Fluorescence emission spectra of the *E.coli* cells co-expressing DnaK-CYFP and NYFP-p5* and incubated with the peptides indicated in the legend.

To gain insights into the dynamics of the process, the BIFC signal was monitored in a concentration- and time-dependent manner. Thus, different concentrations of pyrrhocoricin were added before the induction of protein expression, and the increase in fluorescence was monitored (Figure 2.16). The fluorescence emission increased progressively after induction due to the continuous protein expression and the requirement of fluorophore maturation. From the data, it could also be deduced that the BIFC signal was dependent on the concentration of L-pyrrhocoricin. No inhibitory effect was detectable below 0.5 mM. At higher peptide concentrations, a steady decrease in fluorescence proportional to the L-pyrrhocoricin concentration could be detected. This fact strongly supports that the detected changes in fluorescence reflect the specific action of the inhibitor on the peptidechaperone interaction.

Overall, the previously described in vitro L-pyrrhocoricin binding to DnaK has been further verified *in vivo*. The data suggest that, after induction of DnaK expression, the intracellular L-pyrrhocoricin interacts with the nascent recombinant chaperone, competing

with p5* for complex formation. The subsequent EYFP reassembly is also affected. Meanwhile, the all-D version does not compete with the p5* peptide for the binding to DnaK and does not alter EYFP reassembly. Therefore, it appears that the BIFC method is sensitive to the presence and concentration of protein interaction inhibitors. Moreover, the interpretation of the results is easy and the data could provide relevant functional insights. In this case, it allows confirmation that, in vivo, DnaK is a specific target for pyrrhocoricin, and it correlates the antimicrobial activities of the L- and D- versions of this peptide with their differential affinities for the chaperone

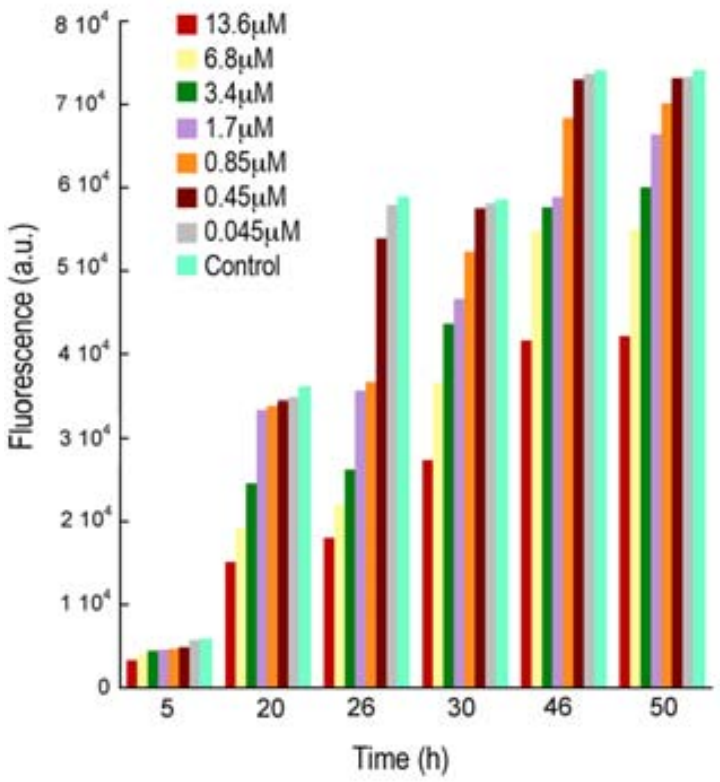


Figure 2.16 Irreversibility of BIFC complex. Fluorescence emission signal of cultures that, after co-expressing DnaK-CYFP and NYFP-p5* and attaining a stable emission signal, were post-incubated with L-pyrrhocoricin (0,045-13 μM) and its derivates (13 μM) for 5h.

2.4.4 Correlation between the activity of pyrrhocoricin derivates and their ability to interfere with *in vivo* DnaK interactions

To be an effective method for the screening and identification of protein interaction inhibitors, the BIFC assay should be able to distinguish compounds exhibiting different inhibitory potency. A complete alanine-scan of pyrrhocoricin was previously performed to identify the specific residues involved in its antibacterial activity. It was demonstrated that the active domain of pyrrhocoricin does not span the entire molecule and that the active site is positioned at the N terminus. Specifically, mutations in the region Asp2-Pro10

caused a strong decrease in antibacterial activity, whereas the rest of amino acids could be replaced without any consequence on the antimicrobial action[34]. This decay in antibacterial activity by side-chain truncation was supposed, but still not proved, to be directly linked to a lower *in vivo* DnaK binding capability.

Table 2.2 Amino acid sequences of the peptides used in this study

Name	Sequence
L-pyrrhocoricin	VDKGSYLPRPTPPRPIYNRN
D-pyrrhocoricin	VDKGSYLPRPTPPRPIYNRN
pyrrL7A	VDKGSYAPRPTPPRPIYNRN
pyrrT11A	VDKGSYLPRPAPPRPIYNRN
Pip-pyrr-MeArg dimer	(PipDKGSYLPRPTPPRPIYN[MeArg]N)$_2$Dab

To confirm this point, the ability of two alanine mutants to interfere with the DnaK-p5* interaction was analyzed using BIFC. A mutant of the binding site (Leu7Ala) and a mutant in an adjacent position to this region (Thr11Ala) were selected. Whereas mutation of Leu7 strongly decreased pyrrhocoricin functionality, the activity of the Thr11 mutant was close to that of the wild-type peptide. Using the above described protocol, the interference of the two mutated peptides with the DnaK-p5* interaction was compared. As confirmed by microscopy, cells grown in the presence of the less active Leu7Ala variant were more fluorescent than the ones incubated with Thr11Ala peptide (Figure 2.17A). Again the expression levels of DnaK-CYFP and NYFP-p5* were fairly similar in both conditions (Figure 2.17B) and no differences in cell growth could be detected.

When the fluorescent signal was quantitatively measured by spectroscopy (Figure 2.17C), significant differences could be detected between the two pyrrhocoricin analogs. The presence of Leu7Ala peptide did not decrease the signal at the fluorescence emission maximum relative to the signal in the absence of the peptide. Meanwhile, Thr11Ala caused an important decrease in the fluorescence emission maximum, comparable to that observed for wild-type L-pyrrhocoricin. This indicated that Leu7 (but not Thr11) was required not only for peptide antimicrobial activity but also for *in vivo* binding to DnaK. This result suggests that the pharmacophore corresponds with the binding site to DnaK, and it provides further evidence that this chaperone is the molecular target of pyrrhocoricin. Therefore, the antibacterial action of this family of insect-derived molecules depends on its specific effects on the protein quality control machinery.

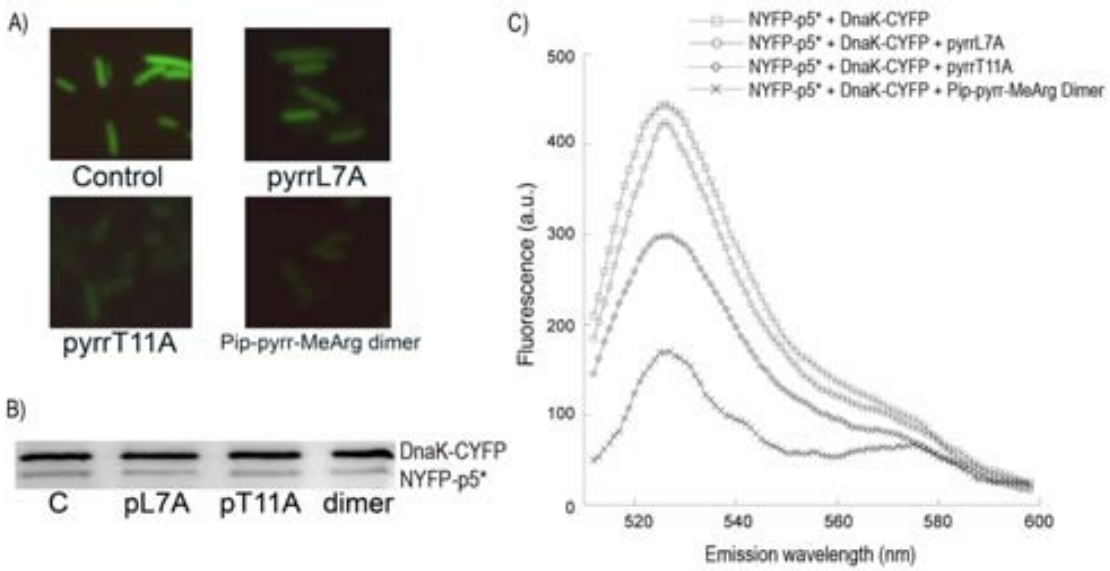


Figure 2.17 Detection of the disruption of DnaK chaperone/p5* peptide interaction using BIFC. (A) Fluorescence images of *E.coli* induced cells transformed with the plasmids encoding for DnaK-CYFP and NYFP-p5* and incubated with the peptides indicated in each case. (B) Western blot analysis of cells expressing DnaK-CYFP and NYFP-p5* and incubated with different pyrrhocoricin derivates. (C) Fluorescence emission spectra of the *E.coli* cells co-expressing DnaK-CYFP and NYFP-p5* and incubated with the peptides indicated in the legend.

Whereas antimicrobial peptides with activity directed primarily toward membranes kill bacteria very fast, members of the proline-rich peptide family need longer to exhibit their antibacterial activity. This fact suggests that pyrrhocoricin variants with higher resistance to the bacterial intracellular and/or secreted proteases would be better antibiotics. Accordingly, it was shown that a C-terminally tethered dimer of pyrrhocoricin, 4-amino-4-carboxy-piperidine acid-pyrrhocoricin-*N*-methyl-arginine (Pip-pyrr-MeArg) dimer displayed both higher stability and antibacterial activity because of the methylation of its peptidase sensitive amide bonds[52].

When cells co-expressing Dnak-CYFP and NYFP-p5* were pre-incubated with the dimeric analogue, the strongest reduction in fluorescence emission was observed (Figure 2.17). Again, no significant decrease in chaperone expression levels was detected. All these data indicate a higher *in vivo* capacity of Pip-pyrr-MeArg dimer for blocking DnaK than the wild-type peptide. The increased inhibition of DnaK by the dimeric form could be due to better penetration through the bacterial membrane, an improved protease resistance or the presence of two copies of the DnaK binding motif.

2.4.5 Irreversibility of the BIFC complex

In the experiments performed, we assumed that the complex between fluorescent protein fragments is mostly irreversible[12]. The certainty of this assumption is, however,

becoming controversial because the complex between the fluorescent fragments has been shown to be at least partially reversible in several recent cases[53-55]. To check this point, L-pyrrhocoricin (in different concentrations) and its variants were added after the complex formation, and the fluorescence was recorded after 5 hours incubation. As it is shown in the Figure 2.18, no decrease in the fluorescence signal could be detected in any of the studied peptides. Only in the case of Pip-pyrr-MeArg dimer, a nearly imperceptible decrease of the fluorescence could be detected (around 2%). This result indicates that, in our case, the reassembled complex was highly stable, precluding the addition of the peptide after protein expression.

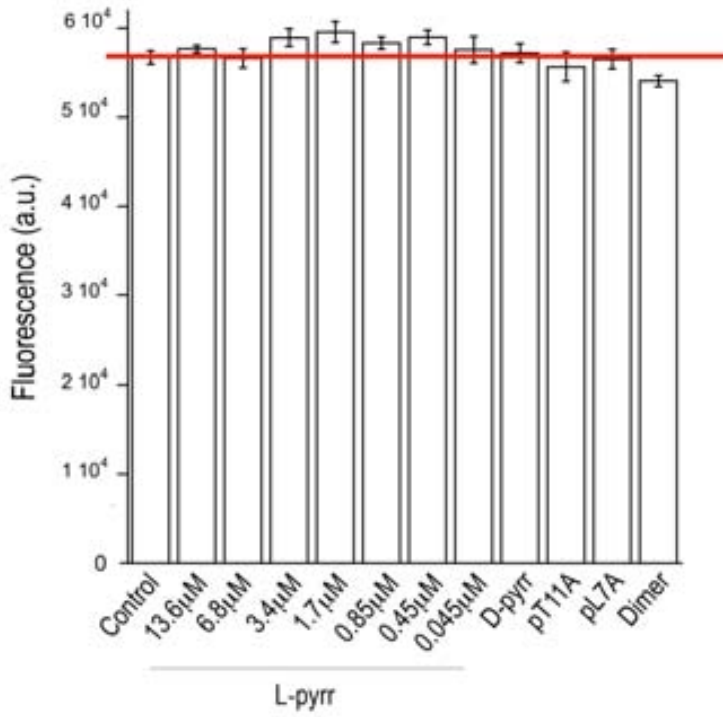
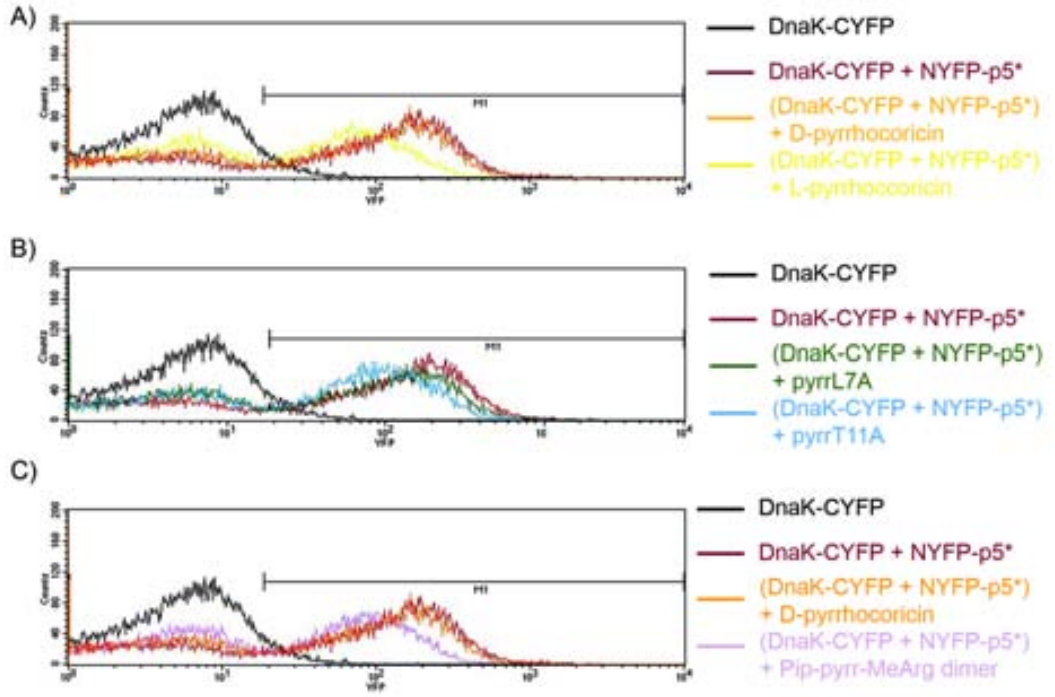


Figure 2.18 Fluorescence emission signal of cultures expressing DnaK-CYFP and NYFP-p5* that have been incubated with L-pyrrhocoricin (0,045-13μM) and its derivates (13μM) after the protein expression.

2.4.6 Application of flow cytometry to the screening of interaction inhibitors

The capabilities of BIFC for high-throughput screening of mutations that affect the strength of protein interactions was previously demonstrated by its coupling to flow cytometry[12, 56]. Therefore, we sought to test whether this approach could had the potential to be a sensitive method for detecting protein interaction inhibitors. The results of this analysis are shown in Figure 2.19.

Cells co expressing DnaK and p5* fusions to YFP fragments or only DnaK-CYFP were used as positive and negative controls. Two main populations with associated low and high mean YFP fluorescence were detected for cells expressing one or two fusions. Separate populations of cells could be defined based on comparison of the fluorescence emitted by negative and positive control cells (gate window M1).



Sample	Mean Fluorescence
DnaK-CYFP	8
(DnaK-CYFP + NYFP-p5*)	176.43
(DnaK-CYFP + NYFP-p5*) + L-pyrr	94.06
(DnaK-CYFP + NYFP-p5*) + D-pyrr	157.9
(DnaK-CYFP + NYFP-p5*) + pyrrL7A	149.34
(DnaK-CYFP + NYFP-p5*) + pyrrT11A	113.45
(DnaK-CYFP + NYFP-p5*) + dimer	102.71

Figure 2.19 Coupling of BIFC to flow cytometry. (A) Frequency histograms of four populations of cells co-expressing DnaK-CYFP (black), DnaK-CYFP and NYFP-p5* (red) and incubated with L-pyrrhocoricin (yellow) or D-pyrrhocoricin (orange). (B) Frequency histograms of cells expressing DnaK-CYFP (black), co-expressing DnaK-CYFP and NYFP-p5* (red) and incubated with pyrrL7A (green) or pyrrT11A (blue) (C) Frequency histograms of four populations of cells co-expressing DnaK-CYFP (black), DnaK-CYFP and NYFP-p5* (red) and incubated with D-pyrr (orange) or Pip-pyrr-MeArg dimer (purple). The mean value of each population is shown in the table below.

To evaluate the ability of flow cytometry to detect the presence of specific inhibitors of protein interactions, cells co-expressing DnaK-CYFP and NYFP-p5* and incubated in the presence of L- or D-pyrrhocoricin were analyzed. In excellent agreement

with the data presented in the previous section, incubation with the active L-variant resulted in the population of cells in M1 exhibiting a significantly decreased mean fluorescence relative to cells grown in free media (or in the presence of the inactive all-D peptide). Moreover, the assay was able to distinguish between inhibitors with point mutations inside or outside the pharmacophore. Cells incubated in the presence of pyrrT11A produced a population in M1 with lower mean fluorescence than that of the cells incubated with pyrrL7A. Incubation with the Pip-pyrr-MeArg dimer form also resulted in a significant decrease of the mean fluorescence of cells in M1 relative to the positive control. These results suggest that BIFC coupling to flow cytometry could be used, in principle, for the identification of protein interaction antagonists with optimal in vivo biological activity (penetrability, stability, specificity and/or affinity).

2.5 Discussion

First of all, the detection of two different interactions of DnaK chaperone provides yet another example of the ability of BIFC to detect specifically weak protein interactions (in the micromolar range).

Nevertheless, the most important point of this work is the application of BIFC assay to the detection of antagonists of protein interactions particularly in the case of the inhibition of DnaK activity by pyrrhocoricin. The data provide further evidence that DnaK chaperone is the molecular target of pyrrhocoricin and its derivates. Moreover, the results obtained with the pyrrhocoricin alanine derivates coincide with previous data: the main binding site of pyrrhocoricin is located between Asp2 and Pro10. Also, this finding could be correlated with the *in vitro* antibacterial activity of these peptides: L7A exhibits a higher IC_{50} than T11A. Therefore, the antibacterial action of this family of insect derived molecules depends on their specific effects over the protein quality control machinery, specifically over the DnaK chaperone activity.

One important aspect of the application of BIFC is that it allows the *in vivo* screening and detection of interactions modulators that display optimal intracellular activity. This fact is of great advantage because the compound is tested taking into account its cell permeability or its stability in front of bacterial proteases. Importantly, the coupling of BIFC to flow cytometry provides a useful tool for the screening of biologically relevant inhibitors of protein interactions.

In order to apply BIFC method to study the interference with protein interactions a key point is the irreversibility of the fluorescent protein reassembly. If the complex between both fragments cannot be undone, the inhibitor should be added before the protein expression. If the reassembly is dependent on the protein interaction, the inhibitor can be added after the appearance of fluorescence. All our previous results *in vitro* (explained in the first chapter) point to an irreversibility of the reassembly. Also this fact has been demonstrated in other interactions under the same conditions[12, 57, 58]. However, it is still a controversial issue because in some experiments, rapid changes in the fluorescence signal have been observed *in vitro* and *in vivo*[53-55]. Nevertheless, in these cases, a decrease in the fluorescence signal was recorded without any direct demonstration of dissociation of the fusion proteins. Besides, during *in vivo* experiments, it is possible that its formation is reversible, but another possibilities such as protein degradation cannot be excluded. In this

sense, when BIFC was applied to study the interaction between a nuclear factor-κB and its modulator IκBα[14], it was demonstrated that the BIFC complex was sensitive to proteosomal degradation.

In another relevant work, the reassembly of *Renilla* luciferease to study protein kinase A activities has been used[59]. Due to the presence of subdomains in this protein, its reassembly is reversible allowing the study of protein dynamics. The authors state that fluorescent proteins cannot be applied to study the response to an inhibitor because its reassembly is irreversible. From our point of view, the use of *Renilla* luciferase (or another bioluminescent protein) will be advisable in the study of protein dynamics[59, 60]. However, one cannot disregard that BIFC based in fluorescent proteins provide a stable fluorescence signal that is especially useful when you are working with transient or weak protein interactions. And we have been able to circumvent the problem of irreversibility by preincubating the cells with the inhibitor. Therefore, the choice between the two proteins (fluorescent or bioluminescent) will depend on the objective of the study and the characteristics of the studied interaction.

2.6 References

1. Pawson, T. & Scott, J.D. Signaling through scaffold, anchoring, and adaptor proteins. *Science* **278**, 2075-2080 (1997).
2. Cohen, F.E. & Prusiner, S.B. Pathologic conformations of prion proteins. *Annu Rev Biochem* **67**, 793-819 (1998).
3. Selkoe, D.J. The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487-498 (1991).
4. Yang, P. et al. Multiplexed detection of protein-peptide interaction and inhibition using capillary electrophoresis. *Anal Chem* **79**, 1690-1695 (2007).
5. Stoevesandt, O. & Brock, R. One-step analysis of protein complexes in microliters of cell lysate using indirect immunolabeling & fluorescence cross-correlation spectroscopy. *Nat Protoc* **1**, 223-229 (2006).
6. Heyduk, T. Luminescence resonance energy transfer analysis of RNA polymerase complexes. *Methods* **25**, 44-53 (2001).
7. Bergendahl, V., Heyduk, T. & Burgess, R.R. Luminescence resonance energy transfer-based high-throughput screening assay for inhibitors of essential protein-protein interactions in bacterial RNA polymerase. *Appl Environ Microbiol* **69**, 1492-1498 (2003).
8. Tucker, C.L. High-throughput cell-based assays in yeast. *Drug Discov Today* **7**, S125-130 (2002).
9. Horswill, A.R., Savinov, S.N. & Benkovic, S.J. A systematic method for identifying small-molecule modulators of protein-protein interactions. *Proc Natl Acad Sci U S A* **101**, 15591-15596 (2004).
10. Lemmens, I., Lievens, S., Eyckerman, S. & Tavernier, J. Reverse MAPPIT detects disruptors of protein-protein interactions in human cells. *Nat Protoc* **1**, 92-97 (2006).
11. Eyckerman, S. et al. Reverse MAPPIT: screening for protein-protein interaction modifiers in mammalian cells. *Nat Methods* **2**, 427-433 (2005).
12. Morell, M., Espargaro, A., Aviles, F.X. & Ventura, S. Detection of transient protein-protein interactions by bimolecular fluorescence complementation: the Abl-SH3 case. *Proteomics* **7**, 1023-1036 (2007).
13. MacDonald, M.L. et al. Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol* **2**, 329-337 (2006).
14. Yu, H. et al. Measuring drug action in the cellular context using protein-fragment complementation assays. *Assay Drug Dev Technol* **1**, 811-822 (2003).
15. McCormick, J.B. Epidemiology of emerging/re-emerging antimicrobial-resistant bacterial pathogens. *Curr Opin Microbiol* **1**, 125-129 (1998).
16. Goldman, R.C. & Gange, D. Inhibition of transglycosylation involved in bacterial peptidoglycan synthesis. *Curr Med Chem* **7**, 801-820 (2000).
17. Bulet, P., Urge, L., Ohresser, S., Hetru, C. & Otvos, L., Jr. Enlarged scale chemical synthesis and range of activity of drosocin, an O-glycosylated antibacterial peptide of Drosophila. *Eur J Biochem* **238**, 64-69 (1996).
18. Casteels, P. & Tempst, P. Apidaecin-type peptide antibiotics function through a non-poreforming mechanism involving stereospecificity. *Biochem Biophys Res Commun* **199**, 339-345 (1994).
19. Hoffmann, R., Bulet, P., Urge, L. & Otvos, L., Jr. Range of activity and metabolic stability of synthetic antibacterial glycopeptides from insects. *Biochim Biophys Acta* **1426**, 459-467 (1999).

20. Otvos, L., Jr. Antibacterial peptides isolated from insects. *J Pept Sci* **6**, 497-511 (2000).
21. Otvos, L., Jr. et al. Insect peptides with improved protease-resistance protect mice against bacterial infection. *Protein Sci* **9**, 742-749 (2000).
22. Hartl, F.U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571-579 (1996).
23. Houry, W.A. Chaperone-assisted protein folding in the cell cytoplasm. *Curr Protein Pept Sci* **2**, 227-244 (2001).
24. Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. & Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**, 693-696 (1999).
25. Wild, J., Rossmeissl, P., Walter, W.A. & Gross, C.A. Involvement of the DnaK-DnaJ-GrpE chaperone team in protein secretion in Escherichia coli. *J Bacteriol* **178**, 3608-3613 (1996).
26. Alfano, C. & McMacken, R. Heat shock protein-mediated disassembly of nucleoprotein structures is required for the initiation of bacteriophage lambda DNA replication. *J Biol Chem* **264**, 10709-10718 (1989).
27. Mogk, A. et al. Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. *Embo J* **18**, 6934-6949 (1999).
28. Zhu, X. et al. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606-1614 (1996).
29. Rudiger, S., Buchberger, A. & Bukau, B. Interaction of Hsp70 chaperones with substrates. *Nat Struct Biol* **4**, 342-349 (1997).
30. Rudiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *Embo J* **16**, 1501-1507 (1997).
31. Schmid, D., Baici, A., Gehring, H. & Christen, P. Kinetics of molecular chaperone action. *Science* **263**, 971-973 (1994).
32. Pierpaoli, E.V., Gisler, S.M. & Christen, P. Sequence-specific rates of interaction of target peptides with the molecular chaperones DnaK and DnaJ. *Biochemistry* **37**, 16741-16748 (1998).
33. Zheng, Z. & Yenari, M.A. The application of HSP70 as a target for gene therapy. *Front Biocience* **11**, 699-707 (2006).
34. Kragol, G. et al. Identification of crucial residues for the antibacterial activity of the proline-rich peptide, pyrrhocoricin. *Eur J Biochem* **269**, 4226-4237 (2002).
35. Chesnokova, L.S., Slepenkov, S.V. & Witt, S.N. The insect antimicrobial peptide, L-pyrrhocoricin, binds to and stimulates the ATPase activity of both wild-type and lidless DnaK. *FEBS Lett* **565**, 65-69 (2004).
36. Kragol, G. et al. The antibacterial peptide pyrrhocoricin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. *Biochemistry* **40**, 3016-3026 (2001).
37. Slepenkov, S.V., Patchen, B., Peterson, K.M. & Witt, S.N. Importance of the D and E helices of the molecular chaperone DnaK for ATP binding and substrate release. *Biochemistry* **42**, 5867-5876 (2003).
38. Bikker, F.J. et al. Evaluation of the antibacterial spectrum of drosocin analogues. *Chem Biol Drug Des* **68**, 148-153 (2006).
39. de Groot, N.S. & Ventura, S. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J Biotechnol* **125**, 110-113 (2006).
40. Ormo, M. et al. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392-1395 (1996).

41. Pisabarro, M.T., Serrano, L. & Wilmanns, M. Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J Mol Biol* **281**, 513-521 (1998).
42. Koo, E.H., Lansbury, P.T., Jr. & Kelly, J.W. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc Natl Acad Sci U S A* **96**, 9989-9990 (1999).
43. Selkoe, D.J. The cell biology of beta-amyloid precursor protein and presenilin in Alzheimer's disease. *Trends Cell Biol* **8**, 447-453 (1998).
44. Selkoe, D.J. Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* **81**, 741-766 (2001).
45. Wang, R., Sweeney, D., Gandy, S.E. & Sisodia, S.S. The profile of soluble amyloid beta protein in cultured cell media. Detection and quantification of amyloid beta protein and variants by immunoprecipitation-mass spectrometry. *J Biol Chem* **271**, 31894-31902 (1996).
46. Roher, A.E. et al. beta-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease. *Proc Natl Acad Sci U S A* **90**, 10836-10840 (1993).
47. Konno, T. Amyloid-induced aggregation and precipitation of soluble proteins: an electrostatic contribution of the Alzheimer's beta(25-35) amyloid fibril. *Biochemistry* **40**, 2148-2154 (2001).
48. Nicolls, M.R. et al. Proteomics as a tool for discovery: proteins implicated in Alzheimer's disease are highly expressed in normal pancreatic islets. *J Proteome Res* **2**, 199-205 (2003).
49. Evans, C.G., Wisen, S. & Gestwicki, J.E. Heat shock proteins 70 and 90 inhibit early stages of amyloid beta-(1-42) aggregation in vitro. *J Biol Chem* **281**, 33182-33191 (2006).
50. Fonte, V. et al. Interaction of intracellular beta amyloid peptide with chaperone proteins. *Proc Natl Acad Sci U S A* **99**, 9439-9444 (2002).
51. Magrane, J., Smith, R.C., Walsh, K. & Querfurth, H.W. Heat shock protein 70 participates in the neuroprotective response to intracellularly expressed beta-amyloid in neurons. *J Neurosci* **24**, 1700-1706 (2004).
52. Cudic, M. et al. Development of novel antibacterial peptides that kill resistant isolates. *Peptides* **23**, 2071-2083 (2002).
53. Cole, K.C., McLaughlin, H.W. & Johnson, D.I. Use of bimolecular fluorescence complementation to study in vivo interactions between Cdc42p and Rdi1p of Saccharomyces cerevisiae. *Eukaryot Cell* **6**, 378-387 (2007).
54. Demidov, V.V. et al. Fast complementation of split fluorescent protein triggered by DNA hybridization. *Proc Natl Acad Sci U S A* **103**, 2052-2056 (2006).
55. Guo, Y., Rebecchi, M. & Scarlata, S. Phospholipase Cbeta2 binds to and inhibits phospholipase Cdelta1. *J Biol Chem* **280**, 1438-1447 (2005).
56. Morell, M., Espargaro, A., Aviles, F.X. & Ventura, S. Study and selection of in vivo protein interactions by coupling bimolecular fluorescence complementation and flow cytometry. *Nat Protoc* **3**, 22-33 (2008).
57. Ghosh Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *Journal of American Chemical Society* **122**, 5658-5659 (2000).
58. Magliery, T.J. et al. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc* **127**, 146-157 (2005).

59. Stefan, E. et al. Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc Natl Acad Sci U S A* **104**, 16916-16921 (2007).
60. Luker, K.E. et al. Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc Natl Acad Sci U S A* **101**, 12288-12293 (2004).