APPLICATION OF PI-DECONVOLUTION TO THE SCREENING OF PROTEIN LIGAND COMBINATORIAL LIBRARIES USING THE YEAST-TWO-HYBRID ASSAY

A Thesis Submitted to the College of
Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of Masters of Science
In the Department of Biochemistry
University of Saskatchewan
Saskatoon

By

Alberto Aparicio

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work; or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5

ABSTRACT

Reagents that bind proteins are applicable in biology for detection of molecules, perturbation of signaling pathways and development of small-molecule pharmaceuticals. Protein ligands interact with proteins, inhibiting or altering their function. They are isolated from combinatorial libraries to interact with a specific target, using selection techniques such as phage display or yeast-two-hybrid assay. For the latter, one inconvenience is the detection of false positives, which can be solved by screening pools containing the samples to be tested, instead of individual samples. Samples are distributed in the pools following a pooling design. The PI-deconvolution pooling design was developed to screen cDNA libraries using the yeast-two-hybrid assay, which are smaller in size than protein ligand combinatorial libraries. Modifications to the PI-deconvolution screening technique were developed to adapt it to the screening of protein ligand combinatorial libraries using the yeast-two-hybrid assay. Every spot of the array containing the combinatorial library was randomly pooled. However, the yeast-two-hybrid assay loses sensitivity when strains are pooled. As PI-deconvolution requires detecting every interaction, we determined the optimal amount of library members that can be pooled in a spot, and the optimal number of replicates to ensure the detection of an interaction.

The yeast-two-hybrid assay was used to perform a screening of a combinatorial library with seven domains of BCR-ABL, which were pooled according to PI-deconvolution. BCR-ABL is a chimeric protein with unregulated kinase activity that is responsible for chronic myelogenous leukemia. The scaffold used in the combinatorial library was an engineered intein that forms lariat peptides. After a screening of this library was performed, positive interactions were detected in 775 spots of the arrays that contained 1432 positive hits. Only 53 spots were deconvoluted. The coding sequences of the lariat peptides were determined for 23 lariat peptides interacted with the GEF domain of BCR, and for ABL, two with the FABD domain, one with the SH1 domain, and one with the SH3 domain. Finally, a β-galactosidase assay was performed to assess the affinity of the lariat peptides for their target.

The isolated lariat peptides are potential inhibitors of BCR-ABL that can have therapeutic potential. This study will improve other screenings of combinatorial libraries with the yeast-two-hybrid assay.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Dr. Geyer for his guidance in this project. I would like to specially thank Kris Barreto for introducing me to the lab, showing me some hints and being supportive through every obstacle I encountered. I also would like to thank the members of the Geyer Lab who were involved in one way or another in my project, Wendy Bernhard, Erika Beaton-Brown, Teodora Zlateva, Carolina Gonzales and Elodie Pastural. The summer and project students, Lindsay Sikorski, Vincent Chan See and Nick Baniack were very helpful in getting the project done. Special thanks to my parents and friends in Saskatoon for all of their help and support during graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

$^{+}$ADE	Plus adenine
AD	Activation domain
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
cDNA	Complimentary DNA
CFU	Colony forming units
CML	Chronic Myelogenous leukemia
CSM	Complete supplement mixture
d	Deoxy prefix
DBD	DNA-binding domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
ds	Double-stranded prefix
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FCC	Frozen competent cell
gal	Galactose
glu	Glucose
−HIS	Minus histidine
HSC	Hematopoietic stem cell
IPTG	Isopropyl-thio-β-D-galactoside
Kan	Kanamycin
−LEU	Minus leucine
LB	Luria Bertani
NLS	Nuclear Localization Signal
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

S. cerevisiae	Saccharomyces cerevisiae
SNP	Single nucleotide polymorphism
STM	Human Stefin A
TF	Transcription factor
−TRP	Minus Tryptophan
X-Gal	5-bromo-4chloro-indoyl-β-D-galactopyranoside
YPDA	Yeast peptone dextrose adenine

1. INTRODUCTION

Antibodies are the most widely used protein ligands for biotechnology applications, including immunoprecipitation, immunohistochemistry, bioseparation, diagnostics, protein chips, ELISA, and therapeutics. However, because of their large size and the requirement of disulfide bonds to preserve their structure they are not suitable for intracellular applications (Nygren and Skerra, 2004). Artificially developed molecules can be developed as alternatives to antibodies, after advances have been made to generate functional diversity and screening methods.

Protein ligands are proteins that physically interact with molecules. They are isolated from protein libraries, for which variants that bind the desired targets are selected using common approaches that detect protein-protein interactions, such as phage display and yeast-two-hybrid assay. Screening a protein library using the yeast-two-hybrid assay presents limitations, since some of the interactions detected with this assay are false. This makes it necessary to reconfirm the detected interactions, through a process that involves a second yeast-two-hybrid assay with plasmids isolated separately (Geyer and Brent, 2000). This is a time-consuming and labour-intensive task, which would be preferably avoided.

Areas of biology that commonly screen libraries employ *pooling* of samples. It suggests testing mixtures of samples or members from a library, called *pools,* instead of analyzing individual samples. The distribution of library members in the pools is dictated by a *pooling design*. Each pooling design offers different benefits, such as reduction of the number of tests, test automation, and detection of false positives in the yeast-two-hybrid assay (Thierry-Mieg, 2006).

The *PI-deconvolution* pooling design was recently developed for the screening of cDNA libraries using the yeast-two-hybrid assay (Jin *et al.*, 2006). In PI-deconvolution, the entire library or the targets that will be screened are pooled into pairs. For every pair, half of the targets are placed in one pool (named positive), and the remaining half in another pool (named negative). The distribution of the library members or targets across the pairs is shuffled for every pair. Overall, a specific code, formed by a positive hit present in the positive or negative pool of every pair, can be assigned to every target or library member that was pooled and tested. When tests are performed for every pool versus an array of protein ligands, positive

interactions are deconvoluted to a specific code (a string of plus or minuses, every pair has a '+' or '−' outcome), which corresponds to a specific bait. As PI-deconvolution tests for a yeast-two-hybrid interaction several times, and confirms that an interaction in a pool is not detected if a target is not present, it avoids the detection of false negatives and false positives.

In this study, PI-deconvolution was employed to screen combinatorial protein ligand libraries using the yeast-two-hybrid assay. Optimal conditions to screen a combinatorial protein ligand library were determined. Screening of a protein ligand combinatorial library with targets pooled according to PI-deconvolution was performed with seven domains of BCR-ABL and a negative control (LexA DNA-binding domain). The following domains were screened: Coiled-Coil (CC), a peptide containing the tyrosine Y177, Guanidine Exchange Factor (GEF), SH3, SH2, Tyrosine kinase (SH1) and F-Actin binding domain (FABD). We isolated and sequenced protein ligands (that interacted with the GEF domain of BCR and the SH3, SH1 and FABD domains of ABL). Finally, we developed a criterion to analyze the output from the screening, referred to as *deconvolution score*. We expect that future screenings of combinatorial protein ligand libraries using the yeast-two-hybrid will benefit from the PI-deconvolution pooling design and will benefit at different scales, particularly if more targets are screened and robotic equipment is employed.

2. REVIEW OF THE LITERATURE

2.1. ISOLATION OF PROTEIN LIGANDS

The extent to which we can understand and control biological systems depends on our ability to manipulate the components of the system under study. Proteins are responsible for the majority of events that take place in a cell and are the preferred targets for perturbing biological systems. One way to alter protein function is to use chemicals that interact specifically with proteins. Antibodies produced by the mammalian immune system are an exceptional source of proteins that interact with other molecules. They can be isolated to interact against virtually any kind of molecule. Antibodies are used in a wide range of applications including therapy, studying protein function, diagnostics, bioseparation, and biodetection of molecules. However, they cannot be used for intracellular applications and are more adequate for extracellular applications. Artificial protein ligands that overcome these limitations can be developed. Protein ligands that interact with specific targets are selected by *in vitro* technologies from large protein libraries (10^6 - 10^{10} different molecules). Advances in molecular biology have boosted research in this field, mainly by improving the generation of protein libraries and methodologies that detect protein-protein interactions.

2.1.1. Alternative protein ligands to antibodies

Antibodies have been so far the most widely used protein ligands for biotechnology applications, such as immunoprecipitations, immunohistochemistry, bioseparation, diagnostics, protein chips, ELISAs, and even therapeutic purposes. The region that contains the antigen-binding site (Fab) is composed of four domains (V_H, C_{H1}, V_L, C_L) that interact with each other (Worn and Plückthun, 2001), while the immunological effector function resides in the constant region of an antibody (Padlan, 1994). Although antibodies have evolved during millennia to act as the main component of the adaptive immune system, their use is mainly restricted to extracellular applications (Nygren and Skerra, 2004). As they form disulfide bonds between the light and heavy chains, the stability of antibodies within the cell is affected by the reducing

intracellular environment, which prevents the formation of these disulfide bonds, and affects protein conformational folding (Worn and Pluckthun, 2001). Modifications to the original antibody molecule that intend to make it functional in the intracellular environment take advantage of the modularity of the antigen-binding site. The antigen binding ability of an antibody is conserved even when only the V_H and V_L domains are present. These two domains can either be associated non-covalently (Fv fragment) (Skerra and Plückthun, 1988), connected by a peptide linker (scFv) (Bird *et al.*., 1988), a disulfide bond (dsFv) (Glockshuber *et al.*, 1990; Brinkmann *et al.*, 1993), or both (sc-dsFv) (Young *et al.*, 1995). These modifications to the original antibody molecule have been shown to bind targets intracellularly and have been used as therapeutic agents and for validation of potential drug targets (Lobato and Rabbits, 2004).

Another alternative is to develop ligand molecules that are not based on the antibody molecule. Early attempts to develop ligands based on linear peptides did not prove to be the best choice. Linear peptides that interact with proteins do not show high affinity and cannot be isolated for the majority of targets screened (Ladner, 1995). Advances made in molecular biology and combinatorial chemistry to generate functional diversity and screening methods have boosted the development of novel protein ligands with promising applications.

2.1.2. Scaffold engineering

Scaffold engineering can be understood as grafting or integrating an affinity function into the structural framework of a stably folded protein with suitable properties (Nygren and Skerra, 2004). As for protein ligands, the scaffold is the support for one or several short amino-acid regions (7-20aa average) that bind a specific molecule, mainly as a protruding finger, or a flat or crevice-like surface (Hosse *et al.*, 2006). The peptide displayed in the scaffold is constrained by either embedding the peptide in the surface loop of the scaffold, or by using disulfide bonds to generate peptide loops on the surface of a scaffold. Both approaches limit flexibility of the peptides. Constrained peptides generally bind tighter due to their limited flexibility, which reduces the entropy penalty upon complex formation (Ladner, 1995). For example, peptides constrained in the scaffold Thioredoxin A of *E. coli* can bind between 100- and 10,000- fold better than the linear version of the same peptide (Geyer and Brent, 2000).

Different properties are desired in a protein scaffold. The ideal scaffold protein should be small (*i.e.* for chemical synthesis), highly soluble, thermally stable, able to fold independently, expressed in large quantities, highly specific, and bind its target with high affinity (Nygren and Skerra, 2004; Hose *et al.*, 2006). Different scaffolds are better for isolating protein ligands that interact with specific targets. Some scaffolds are better at binding small molecules, whereas others are better at binding large protein surfaces. The ideal scenario would be to develop a *universal library* of protein ligands, based on a scaffold that could bind any kind of molecular topology, from small molecules to linear peptides to protein surfaces. Different protein ligands that could bind an entire genome could be generated, providing an invaluable set of tools to understand the molecular mechanisms occurring in the cell. Nonetheless, we are far from reaching this point. Applications based on antibodies are still positioned at the top of the majority of life sciences applications, whereas many of the novel scaffold-type ligands have only been investigated in terms of principal applicability in different contexts (Nygren and Skerra, 2004).

Two factors are important in the identification of a protein ligand that interacts with an specific target (Colas, 2000). The generation of protein libraries and the selection of proteins that can bind desired targets. The diversity of a protein library is generated by synthetic means. Most protein ligand libraries are based on a synthetic nucleic acid template, such as a plasmid, or mRNA; this depends on the selection technique that will be employed. The most common approaches to generate diversity in a library include mutagenesis by PCR, chemical synthesis of oligonucletides, and exon shuffling. Chemical synthesis of peptides has also been attempted, although to a lesser extent. For DNA-based libraries, it is easy to identify the sequence of the isolated protein ligands and is possible to amplify them by PCR or growing the plasmid in bacteria. This is a significant advantage over libraries based on peptides or small molecules. Selection methodologies that detect protein-protein interactions include phage display, the yeast-two-hybrid system, or mRNA display, among others.

2.1.3. Methodologies for the selection of protein ligands from combinatorial libraries

Protein ligands are isolated from combinatorial libraries, constituted by several scaffold proteins, each one having a different sequence in its binding region (variable region). The most popular *in vitro* methodologies to isolate protein ligands include phage-display, yeast display, ribosome, and mRNA display. In phage display, protein ligands are isolated by a strategy that involves the coating of the target protein on a plastic surface, followed by exposure to a library of combinatorial protein ligands expressed on the surface of a bacteriophage. Subsequently a washing step is performed, which removes proteins that do not bind the target, while the scaffold proteins that bound the target proteins are retained. This step can be performed several times, accompanied by mutagenesis, resulting in the selection of protein ligands with higher affinities (Sidhu *et al.*, 2000). One advantage of phage display is that large libraries can be screened (up to 10^{10} members) (Harrison *et al.*, 1996); this holds true for other *in vitro* selection techniques as well. On the other hand, the yeast-two-hybrid system can only survey smaller library sizes, in the order of 10^6 to 10^7 members, because the transformation efficiency is lower in yeast. Nonetheless, if selected protein ligands isolated by *in vitro* methodologies will be used intracellularly, there is no guarantee that the isolated protein ligands will be functional. The selection procedure takes place in an oxidizing environment, unlike the reducing one found in the cytoplasm and the nucleus (Geyer and Brent, 2000). They can be toxic or acquire a different conformation once expressed inside cells. However, phage display is a preferred method if the target is a secreted protein or a membrane receptor (Deshayes *et al.*, 2002).

2.1.3.1. Selection of protein ligands from combinatorial libraries using the yeast-two-hybrid assay

The yeast-two-hybrid assay is an *in vivo* assay that allows the detection of protein-protein interactions *in vivo*. It is based on the separable nature of transcription factors (Causier, 2004). In a transcription factor, the DNA-binding domain (DBD) acts to localize the protein to specific DNA sequences within the genome, whereas the activation domain (AD) contacts the transcription machinery to activate gene transcription A transcription factor can be functional even when the two domains are not present in the same protein, they only need to be physically

linked. Fields and Song (1989) noted that it was possible to reconstitute a transcription factor by bringing the DBD and AD of a transcription factor together after fusing each to one of a pair of physically interacting proteins (**Figure 2.1**). Yeast was the organism of choice because it is eukaryotic, well studied and relatively easy to manipulate.

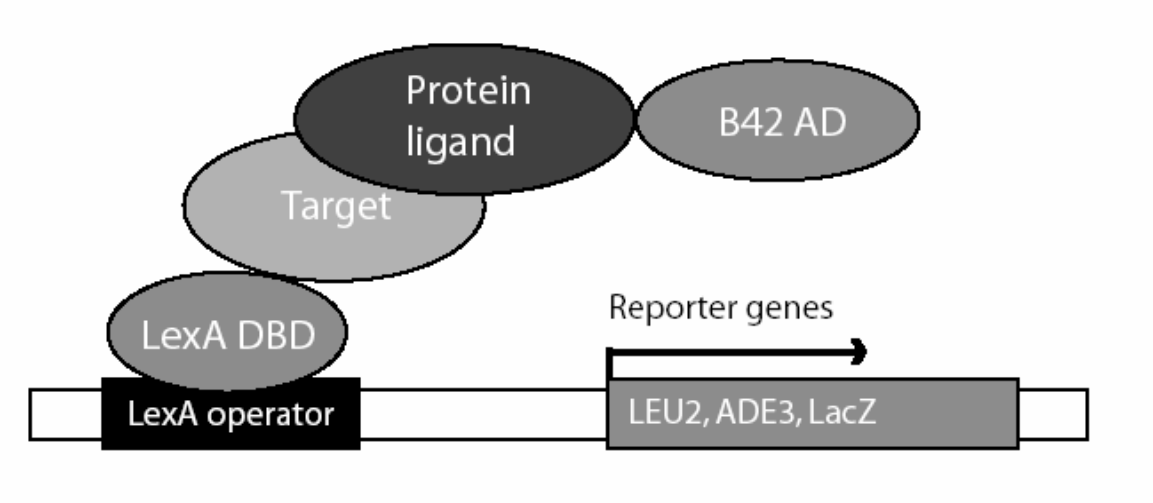


Figure 2.1. Overview of the yeast-two-hybrid assay.

Inside a yeast cell nucleus are expressed the bait, a target protein fused to a LexA DNA-binding domain and the prey, a protein ligand fused to the B42 activation domain. If the bait and prey interact, a transcription factor is reconstituted with the LexA DNA-binding domain and the B42 activation domain, inducing the expression of reporter genes such as LEU2 and ADE3 that will allow the yeast cells to grow in synthetic media that lacks leucine and adenine, as well as LacZ, which causes the colonies to turn blue in the presence of X-Gal.

The yeast-two-hybrid assay has been used to isolate protein ligands from combinatorial libraries against different targets. Its main advantage is its ability to detect protein-protein interactions within eukaryotic cells, therefore the interactions detected in yeast cells are very likely to be reproduced in mammalian cells (Geyer and Brent, 2000). Hence the isolated protein ligands can be used as tools to study cell signaling pathways and validate drug targets, since most of them inhibit protein function (Colas *et al.*, 2008). Additionally, the minimal affinity detection of the assay lies within the micromolar range, ensuring the isolation of protein ligands with high affinity for its targets. Different protein ligands have been isolated using the yeast-two-hybrid assay that interact with targets as varied as Cdk2 (Colas *et al.*, 1996), Nr-13

(Nouvion *et al.*, 2007), the Hepatitis B Virus core protein (Butz *et al.*, 2001), the SH3 domain of RasGAP (Pamonsinlapatham *et al.*, 2008) and BCL-6 (Chattopadhyay *et al.*, 2006).

The yeast-two-hybrid assay can be used for phenotypic screenings that aim to find a protein ligand that causes a specific phenotype, instead of interacting with a particular protein. Once the protein ligand that causes the desired phenotype is isolated, it is used as a bait to screen a cDNA library. This step allows the identification of the protein responsible for the desired phenotype, which can be used to gain insights into biological mechanisms. This approach was used to identify proteins and protein-protein interactions involved in the yeast mating α-factor response pathway (Geyer and Brent, 1999). In another study, protein ligands (peptide aptamers) that inhibited mammalian cell proliferation *in vitro* were isolated and found to interact with Calcineurin (de Chassey *et al.*, 2007).

2.1.3.2. Components of the yeast-two-hybrid assay

In order to apply the yeast-two-hybrid assay to the screening of combinatorial of protein ligands, three features must be present (**Figure 2.1**) (Hoppe-Seyler *et al.*, 2004). First, the target protein, referred also as bait, must be expressed fused to a LexA DBD. The second feature consists of a combinatorial library of protein ligands. The source of diversity are chemically synthesized oligonucleotides, that encode codons for all standard twenty amino-acids in a variable region of five to twenty amino-acids. Finally, both the bait plasmid and the library (in plasmid form) are inserted into yeast strains of opposite mating type (MATa and MATα). The transformation of the vector library into the prey yeast strain is the limiting step of the generation of the library diversity, since transformations in yeast cells are not as effective as in other organisms like *E. coli*, and a significant part of the original diversity found in the oligonucleotides is lost (Geyer and Brent, 2000).

One of the yeast strains that will express the fusion proteins contains reporter genes under the control of LexA operators, which are expressed as the result of the interaction between the prey and the bait, as this leads to a reconstitution of the transcription factor. They can be auxotrophic genes, which supply defective metabolic pathways in yeast mutant strains, allowing the formation of colonies in culture medium that lacks specific nutrients when the bait and prey interact. For example, the yeast strains (both MATa and MATα) used in the assay

cannot grow in medium that lacks leucine, because they have a *leu2* mutation that impairs their ability to produce this amino acid. However, if a bait and a prey that interact with each other are expressed, the reconstituted transcription factor will induce the expression of the *LEU2* gene (under the control of a LexA operator), allowing the cells to grow. Colorimetric reporter genes are also used.

The gene *lacZ* is the most widely used (also under the control of a LexA operator). It encodes the β-galactosidase protein, which causes colonies to turn blue in the presence of the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). Green fluorescent protein (GFP) has been used as well (Fashena *et al.*, 2000). When the two strains that carry the bait and the prey are put in contact in rich media (such as YPDA), they will form diploids. Diploids express both the bait and the prey fusion proteins. If they interact, reporter genes will be expressed, allowing the diploids to grow in auxotrophic media, thus the interaction can be tested (Golemis *et al.*, 2001).

2.1.4. Classification of scaffolds

Protein scaffolds developed until this date lie within a framework of defined structures and folds (**Figure 2.2**). Different protein scaffolds have been employed for the development of protein ligands that interact with different targets.

2.1.4.1. Scaffolds for the presentation of a single loop

Scaffold proteins that present one random peptide loop are the simplest design (**Table 2.1**). The binding site for these protein ligands is formed by a short, extended, constrained peptide stretch with varying sequence, usually presented as an exposed loop (Nygren and Skerra, 2004). It has been suggested that the conformational constraint of the random peptide increases its binding affinity by reducing the entropy of binding upon its target (Ladner, 1995). This scaffold facilitates library construction since it is only necessary to randomize one section of the scaffold, instead of different regions.

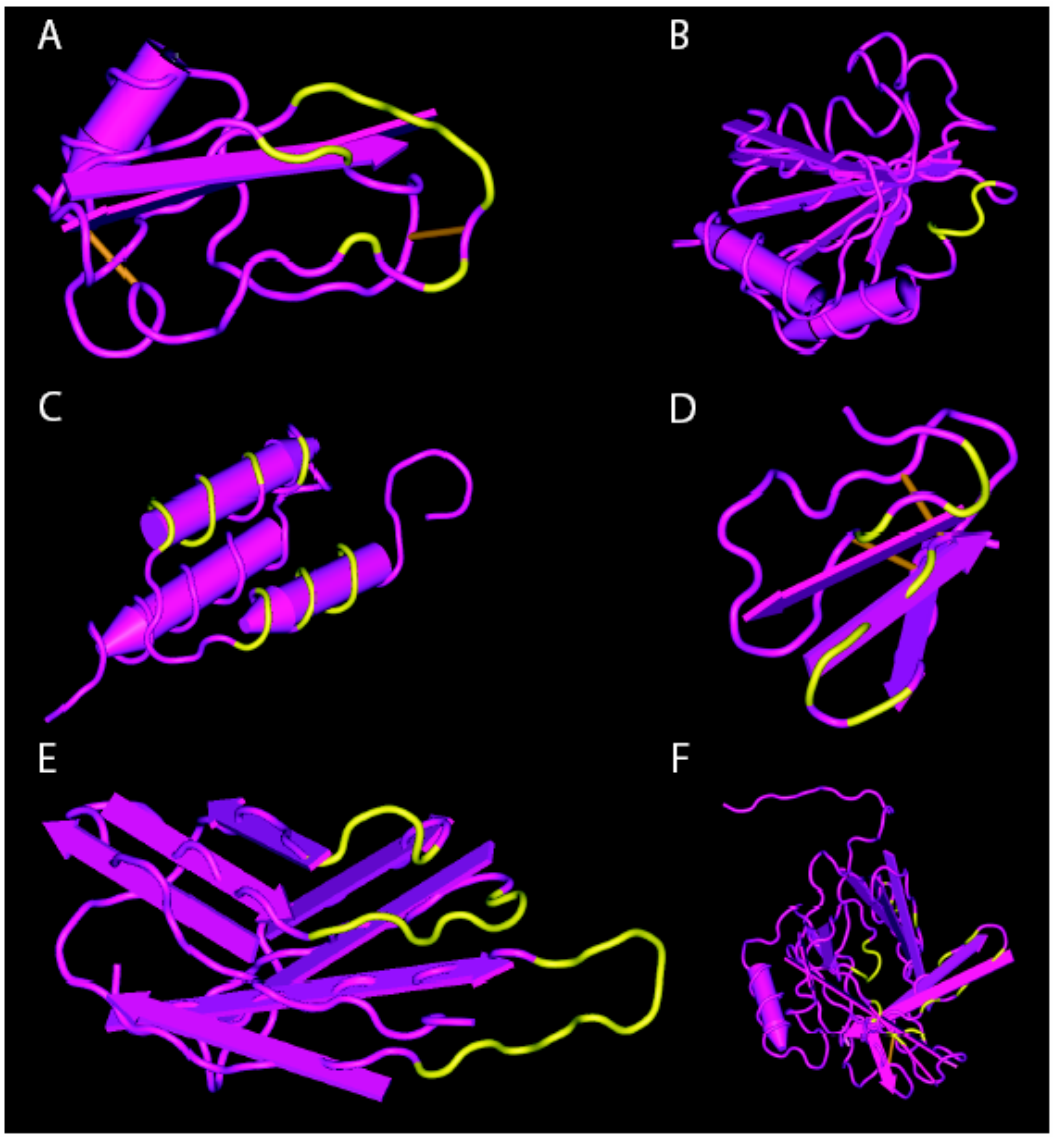


Figure 2.2. Representative protein scaffolds selected for the development of protein ligands.

The binding region of scaffolds A-B consist of single loops, secondary elements for scaffolds C-D and multiple loops that resemble immunoglobulins for scaffolds E-F. The protein backbone including α-helices and β-sheets is depicted in purple; straight bars are disulfide bonds and the positions subjected to randomization are shown in yellow. The PDB IDs used to generate this figure are shown in parenthesis. **(A)** Kunitz domain (1AAP) **(B)** Thioredoxin A (1XOB) **(C)** Affibody: Z-domain of protein A (1Q2N) **(D)** Knottin: cellulose binding domain from cellobiohydrolase Cel7A (1CBH) **(E)** 10^{th} Fibronectin III (1TTG) **(F)** Anticalin FluA: bilin-binding protein (1T0V) with cavity randomization for fluorescein binding.

Table 2.1. Examples of validated scaffolds that display the binding site as a single loop.

Name	Origin	Structure	Target	Selection method	Source
Kunitz	Bovine pancreatic trypsin inhibitor	Stable, 60 residues, 3 disulfide bonds	Human neutrophil elastase	Phage display	Roberts *et al.*, 1992
	Alzheimer's amyloid beta-protein precursor inhibitor (APPI)		Human TF.FVIIa (tissue factor/factor VIIa) complex	Phage display	Dennis *et al.*, 1994
	Thioredoxin A (TrxA)	Short, highly soluble, forms constrained loop	Three monoclonal antibodies	*E. coli*'s flagellum display	Lu *et al.*, 1995
Aptamer			Cyclin dependant kinase 2 (Cdk2)	Yeast-two-hybrid assay	Colas *et al.*, 1996
Peptamers	Staphylococcal nuclease	Small, spontaneous fold, exposed loop on surface	Cell division and yeast pheromone response pathway proteins	Yeast-two-hybrid assay	Norman *et al.*, 1999

2.1.4.2. Scaffolds with structurally variable loops

The binding site for this type of scaffold is formed by a number of hyper-variable loops, displayed on secondary elements from different regions of the scaffold (**Table 2.2**). The best example is the immunoglobulin antibody molecule, in which the variable light and heavy chain domains form a sandwich of two antiparallel β-sheets forming the structurally conserved framework (Padlan, 1994). Several variations to the natural antibody have been proposed, as well as new scaffolds whose binding site is based on the arrangement of multiple loops.

Table 2.2. Examples of validated scaffolds that display the binding site as variable loops.

Name	Origin	Structure	Target	Selection method	Source
VHH domains	Dromedary heavy chain antibodies	Two pairs of heavy chains, only 3 CDRs protruding from one edge	Bovine erythrocyte carbonic anhydrase	Animal immunization	Lauwereys *et al.*, 1998
			TEM-1 and BcII β-lactamases		Conrath, 2001
	Neocarzinostatin	Chromoprotein, consists of seven beta-strands in two sheets forming a β-sandwich	Testosterone	Phage display	Heyd *et al.*, 2003
New Antigen Receptor (NAR)	Immunoglobulin from shark	Two chains, each with a variable and five constant domains	Gingipain K	Phage display	Nuttall *et al.*, 2001
FN3 (Fibronectin Type III)	10th domain of 15 repeating FN3 units in human fibronectin	β-strands with three loops connecting the strands at one end of the β-sandwich	Ubiquitin	Phage display	Koide *et al.*, 1998
			TNF-α	mRNA display	Xu *et al*., 2002
	DNA-binding domain of human retinoid-X-receptor alpha	Two zinc fingers protruding from an hydrophobic core	ATP	mRNA display	Cho & Szostak, 2006
Anticalins	Lipocalin fold from bilin-binding protein (BBP)	Four loops supported by a central β-barrel	Fluorescein	Phage display	Beste *et al*., 1999

2.1.4.3. Scaffolds providing interfaces on secondary structural elements

For this type of scaffold, protein-protein interactions are mediated by rigid super-secondary structures (**Table 2.3**). The binding surface is large and flat. These structures involve several α-helices or β-strands brought together and side chains protruding from two or more of these elements form the interface for complex formation with the target protein (Nygren and Skerra, 2004).

Table 2.3. Examples of validated scaffolds that display the binding site as a secondary structure.

Name	Origin	Structure	Target	Selection method	Source
Affibodies	Protein A	Three. α-helix bundle module of ~60aa	Taq DNA Polymerase, human insulin, apoliprotein A-1	Phage display	Nord *et al.*, 1995
Ankyrin Repeats (AR)	Artificial scaffold with a consensus sequence	Domains built from repeats composed by a β-turn, two antiparallel α-helices and a loop	Maltose-binding protein (MBP)	Ribosome display	Binz *et al.*, 2004
PDZ domain	PDZ domain of AF-6	Binding site formed by second alpha-helix and second beta-strand and a carboxylate binding loop	Drl and Ephrin-B1	Yeast-two-hybrid assay	Schneider *et al.*, 1999
1ICA29	Insect Defensin A	29aa, forms an alpha-helix and two beta-strands, stabilized by two disulfide bonds	TNF-α and TNF receptors	Phage display	Zhao *et al.*, 2004
Knottins	C-terminal CBD domain of cellobio hydrolase I	Triple stranded anti-parallel beta-sheet, stabilized by a cysteine knot	Bovine alkaline phosphatase	Phage display	Smith *et al.*, 1998

2.1.4.4. Cyclic peptide ligands

Cyclic peptides are independent of a protein scaffold. They are peptide chains whose N− and C− termini are linked through a peptide bond, forming a circular chain. For cyclic peptides, similarly to peptides embedded in scaffold, the constraint of the peptide increases the affinity for its target. Cyclic peptides are resistant to cellular exoproteases, because they do not have exposed termini, and are easy to synthesize (Horswill and Benkovic, 2005). Cyclic peptides that act as inhibitors of proteins have been found in nature. Some examples include the immunosuppressant Cyclosporin and the antimicrobiall θ-defensin-1 (Craik, 2006). Cyclic peptides that interact with specific target proteins have been isolated from combinatorial

libraries using phage display. In this technique, the peptides are constrained at disulfide bonds formed by cysteines at their ends (Giebel *et al.*, 1995; Meyer *et al.*, 2006).

New possibilities for the isolation of cyclic peptides that interact with proteins have arisen with engineered inteins. Inteins are segments of a protein that are able to excise themselves and join the remaining segments (exteins) through a peptide bond (Gogarten, 2002). Scott and colleagues (1999) have altered the position of the exteins within an intein, generating the ligation of the exteins into a cyclic peptide. This system has been used isolate cyclic peptides from combinatorial libraries to interact with a variety of targets such as the Dam Methyltransferase (Naumann *et al*., 2008) and the protease ClpXP (Cheng *et al.*, 2007). Nonetheless, it is not possible to isolate cyclic peptides that interact with specific proteins using selection systems such as the yeast two-hybrid assay, since there is no N− or C− terminus to fuse a transcription activation domain. The yeast-two-hybrid assay allows isolation of protein ligands that interact with multiple surfaces of a desired target protein, instead of proteins ligands that disrupt an interaction with another protein or a small molecule. The Geyer lab has recently developed a novel genetic assay to isolate lariat peptide inhibitors using the yeast two-hybrid assay. The original intein system that generates cyclic peptides (Scott *et al.*, 1999) has been further engineered to halt the cyclic peptide reaction at an intermediate step (Geyer RC, unpublished data), which produces a *lariat peptide* that contains a transcription activation domain covalently attached through an amide bond to a lactone-cyclized peptide (**Figure 2.3**). This modification allows the screening of combinatorial libraries of lariat peptides using the yeast-two-hybrid assay.

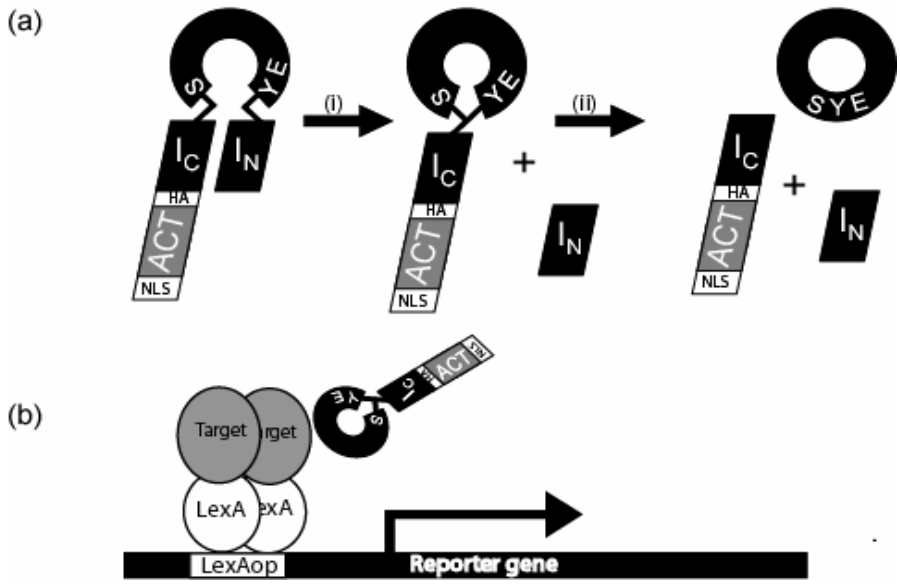


Figure 2.3. Engineered intein to isolate cyclic peptides using the yeast-two-hybrid assay.

a) Intein-mediated peptide cyclization. Steps involved in the cyclization reaction. (i) Folded active intein catalyzes a N-to-S acyl shift to form a thioester intermediate, which undergoes a transesterification reaction with the serine (S) nucleophile to release the I_N fragment and produce the lariat intermediate. (ii) Asparagine side chain cyclization releases the I_C-hemaglutinin (HA)-B42 activation domain (ACT)-nuclear localization sequence (NLS) and generates a peptide cyclized through a lactone linkage. The lactone undergoes an energetically favoured rearrangement to form the final peptide product cyclized through a lactam linkage. **(b) Lariat peptide yeast two-hybrid assay**. In the yeast two-hybrid assay, if the lariat peptide interacts with the target protein, the attached transcription activation domain (ACT) activates reporter genes. The lariat intermediate is inhibited from proceeding to the cyclic peptide product by blocking the asparagine side chain cyclization reaction in step (ii) using an asparagine to alanine mutation. The plasmids encoding the target protein fused to LexA (bait) and the lariat peptide are present in different yeast strains; when mated, the resulting diploids express the two plasmids and the yeast-two-hybrid assay can be tested.

2.1.5. Applications of protein ligands

Protein ligands can serve many purposes. They can be exploited at the intracellular level to play important roles in elucidating signaling pathways and validate drug targets. At the extracellular level, they play important roles in methodologies such as bioseparation and detection of molecules, and they are suitable for proteomic-scale studies (Uhlen, 2008).

2.1.5.1. Protein ligands for manipulation of signaling pathways

Protein ligands can be used to study and manipulate signaling pathways. Due to the modular nature of proteins signaling pathways are very flexible. Among the proteins involved in a signaling pathway adaptor proteins mediate specific protein-protein interactions. It also means they can be broadly manipulated, which has been documented in cases of pathogenic proteins that have acquired the capacity to mediate ectopic protein-protein interactions, such as the E6 protein of Human Papilloma Virus (HPV) and the TIR protein from enteropathogenic *E. coli* (Pawson and Warner, 2007). Altered adaptor proteins have also been observed in oncogenes, especially in the case of chromosomal aberrations, such as BCR-ABL, a chimera that possesses unregulated kinase activity and access to new interaction domains (Ren, 2005). Artificial *re-wiring* or manipulation of signaling pathways has been achieved. Howard *et al.* (2003) created artificial adaptors that joined two different signaling pathways, one apoptotic (apoptosis by death receptors) and the other pathway involved in cell growth, differentiation and survival (*i.e.* the Ras-mitogen-activated protein (MAP) kinase/PI 30-kinase pathways). In particular, the Fas-associated death domain (FADD) links receptors such as Fas to caspase-8/10, and thus links them to apoptotic pathways. On the other hand, the SH2 or phosphotyrosine-binding (PTB) domains of the GRB2 or SHCA adaptor proteins couple tyrosine kinases to signaling pathways involved in cell growth, differentiation, and survival. The chimeric proteins DED-SH2 or DED-PTB were found to complex with caspases and were recruited to activated receptor tyrosine kinases. This resulted in caspase activation, likely as a consequence of clustering of the receptor-associated complex, therefore in cell death (Pawson and Warner, 2007). Protein ligands can be used as adaptors that could integrate any protein into a desired signaling pathway, as a chimeric protein fused to domains with enzymatic activity or other adaptors. This opens the possibility to manipulate signaling pathways, and correct aberrant signaling that occurs in many diseases.

Another mechanism to manipulate signaling pathways involves translocation of target proteins. Colas *et al.* (2000) expressed protein ligands (peptide aptamers) with affinity for CDK2 as fusions to functional domains, to modify and transport CDK2. In particular, the CDK2-interacting TRXA protein ligand was fused to a *hect* domain from ubiquitin-ligases, in order to form a complex with ubiquitin ligase to target proteins for degradation. It was shown that CDK2 is effectively ubiquitinated by the aptamer chimera, although this modification did

not target the protein for degradation. Furthermore, the authors also showed that an anti-CDK2 aptamer fused to a nuclear-localization-signal (NLS) could transport CDK2 into the nucleus.

2.1.5.2. Protein ligands for validation of drug targets

The discovery of new drugs has become a major challenge for the pharmaceutical industry, and many potential pharmaceuticals do not pass clinical trials. Despite major achievements in biomedical sciences and biotechnology, fewer drugs are currently introduced to the market than twenty years ago (Kola and Landis, 2004). This trend can be attributed to lack of efficacy and excessive toxicity of the isolated compounds. Criteria to select target proteins for therapeutic intervention have been based on experimental results from studies using gene knock-outs or downregulation of protein expression. These approaches mainly abolish the entire repertoire of functions of the protein. Perturbation of target proteins can be better achieved with protein ligands, as they bind to specific surfaces of a protein, thus inhibiting function or blocking essential interactions. Abed *et al.* (2007) compared the effect of different reverse genetic approaches such as gene disruption, dominant negative allele over-expression and a protein ligand (a peptide aptamer that interacts with FUR) on the transcriptional network regulated by the FUR (ferric uptake regulator) protein of *E. coli*, used as a case study. This protein regulates processes as varied as iron homeostasis, metabolism and stress responses. Overall, gene knock-out was found to induce more significant perturbations than the other two approaches. This can be explained by the fact that the anti-FUR peptide aptamer and the dominant negative only inhibited the DNA binding-dependent activities of FUR but not its alternative iron sequestration function (Abed *et al.*, 2007).

There is a discrepancy between the perturbations introduced by most target validation methods and by therapeutic molecules, resulting in target proteins that are either falsely validated or falsely rejected (Baines and Colas, 2006). Pharmaceutical drugs are essentially small molecules that interact with a specific surface of a protein, thus altering a specific function. This is the same mechanism by which protein ligands exert their function. Therefore protein ligands can be used to validate drug targets. First, most of the scaffolds can be expressed intracellularly, so it can be determined whether their expression induces the desired effect, such as killing cancerous cells. On the other hand, the same protein ligands can be used

as leads for the discovery of small molecules. In this manner, a library of small molecules can be screened to isolate competitive inhibitors of the interaction between the target protein and the protein ligand. A modified version of the yeast-two-hybrid system was employed to isolate small-molecule compounds that inhibited the interaction between FKBP12 and transforming growth factor beta receptor (TGFβ-R), leading to the discovery of inhibitors of the calcineurin-dependent signaling in T cells (Joshi *et al.*, 2007).

2.1.5.3. Protein microarrays

High-throughput analysis of gene and protein expression is gaining popularity given the availability of entire genome sequences and advances in methodologies that allow the detection of nucleic acids and proteins. DNA microarrays take the lead in this field, since their protein counterpart still presents technical limitations. The majority of protein microarray technologies that are based on antibodies perform poorly. Antibodies immobilized on surfaces present significant cross-reactivity with other proteins, are not chemically stable, and their binding specificity is biased to linear antigens motifs from denatured proteins (Haab *et al.*, 2001). An alternative to protein microarrays based on antibodies can be found in protein ligands. A scaffold based on human Stefin A (named STM) was developed to be employed in extracellular applications (Woodman *et al.*, 2005). The scaffold was engineered to reduce its cross-reactivity with other proteins and was inserted a binding site previously shown to bind CDK2, displayed on a TRXA aptamer (Colas *et al.*, 1996). The STM scaffold was further mutated to introduce a single cysteine residue that allowed the oriented attachment of the scaffold to a gold surface, via the exposed sulfhydryl group that forms an S−Au bond (Davis, *et al.*, 2007). Dual-polarization interferometry was used to detect the interaction between the STM_{cys+} scaffold and Cdk2 from yeast cell lysates with high affinity and specificity, demonstrating the usefulness of the scaffold in protein microarray methodologies (Johnson *et al.*, 2008).

2.2. POOLING DESIGNS FOR SCREENING OF LIBRARIES

In biology, a library can be defined as a collection of different molecules in a stable form. They can be of any chemical nature, ranging from peptides to cDNA molecules present as recombinant DNA. They are maintained as a mixture or in an array. Protein ligands are isolated from combinatorial protein libraries, by employing methodologies that are used for large-scale selections. These libraries contain proteins that consist of an underlying constant scaffold and a randomized variable region that differs among members of the library. A screening, or a protein-protein interaction test with all the members of the library, is conducted to detect a protein ligand that interacts specifically with a target of interest. Screenings are employed with technologies that test for protein-protein interactions such as phage display and the yeast-two-hybrid assay. In a matrix-based library screening using the yeast-two-hybrid assay, several arrays containing the protein library are created, where every spot contains from one to hundreds of library members (**Figure 2.4**). A yeast-two-hybrid assay is performed with the target protein against all the spots in the arrays. In this way, all possible combinations are tested and a comprehensive coverage of the library is achieved. Nonetheless, this approach is both time-consuming and labour-intensive, since thousands of mating operations must be performed (Zhong *et al.*, 2003).

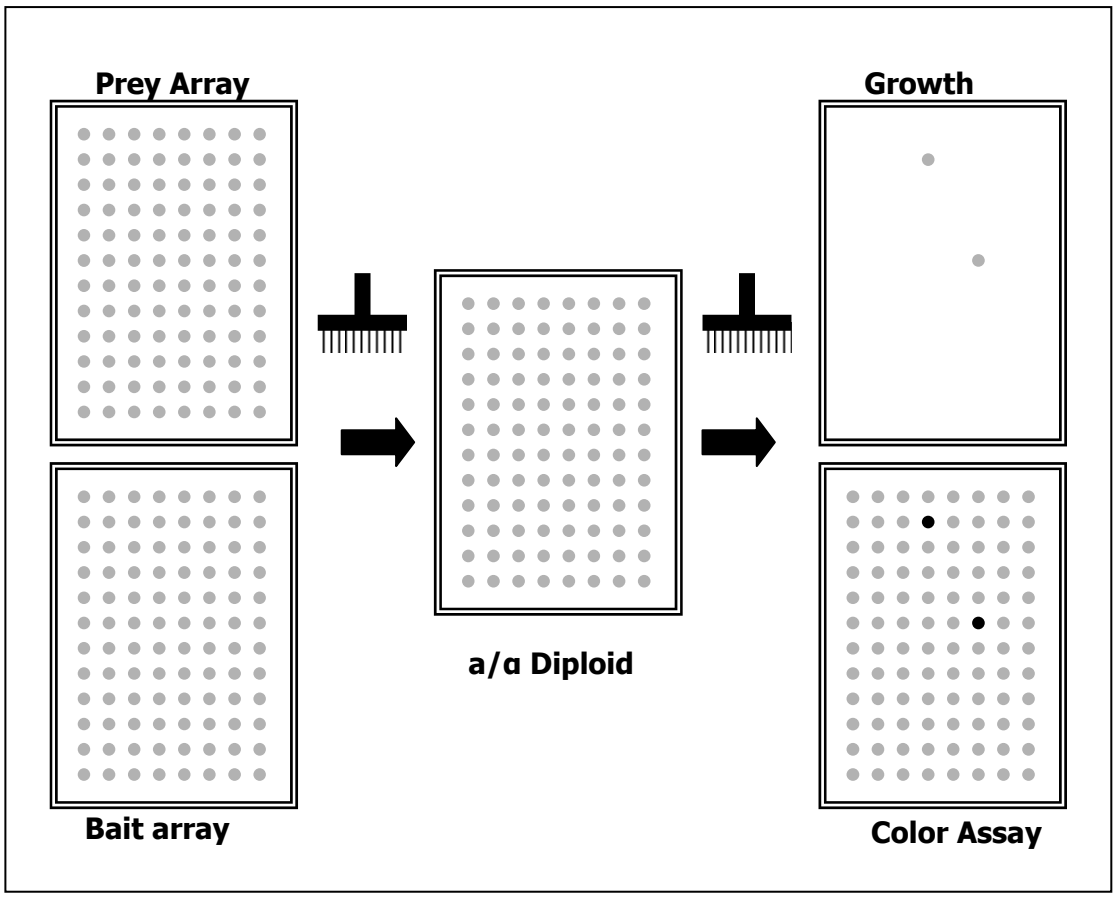


Figure 2.4. Matrix-based yeast-two-hybrid assay.

An array is formed with yeast colonies expressing a bait or a prey. Every component of the array is defined as a spot. For the prey array, every spot contains one to hundreds of different library members. The bait array contains a yeast strain expressing a single target protein (bait) for each spot. The two arrays are plated into a new array that allows the prey and the bait strain to form diploids. The diploids are then plated to scoring plates in which only the yeast-two-hybrid positive diploids can grow (where an interaction between the prey and the bait occurs).

2.2.1. Limitations of the yeast-two-hybrid assay

When screening a protein library using the yeast-two-hybrid assay, many interactions (positive hits or colonies) can be detected. Many of these positive hits are false interactions, hence it is necessary to reconfirm the detected interactions by isolating the prey plasmid from the positive colonies, reinserting it in a new yeast strain, and performing a second yeast-two-hybrid assay (Geyer and Brent, 2000). A comparison of studies that aimed to map the

interactome of *Saccharomyces cerevisiae* revealed that many protein interactions were missed. For example, among the 1,813 interactions that were identified in three different large-scale yeast-two-hybrid screenings, only six interactions are found in common among all three data sets (Ito *et al.*, 2000). The discrepancies between the studies might be explained in part because they covered different fractions of the interactome, which did not overlap substantially. It is not completely understood why the yeast-two-hybrid assay presents such a high degree of both false positives and negatives. The inability of the yeast-two-hybrid to detect some physiological protein-protein interactions can be attributed to improper folding of fused proteins in the yeast cell, proteins not entering the nucleus, or proteins not being post-translationally modified (von Mering *et al.*, 2002). There is also difficulty in determining the interaction of proteins that are toxic to the cell or proteins that auto-activate the yeast-two-hybrid reporters without an interaction with another fused protein.

2.2.2. Pooling and group testing

Screening of libraries in different areas of biology seeks a common goal: To locate or isolate one member of the library, which can be a protein ligand that interacts with a target protein. In genome sequencing projects, genomic libraries are available as recombinant fragments of a genome sorted in individual clones. It is often necessary to detect or isolate an specific member of a library (*i.e.* member *A*). To do so, all the members of the library would have to be tested. Instead of testing all the members of the library to determine which one is *A*, it is more efficient to test *pools* of members. A *pool* is a mixture of library members, or a portion of the library. The goal becomes to detect whether a pool contains *A*, and from there, only the members in the pool that contains *A* will be tested to determine which one is *A*. This sort of problem can be addressed as a *group test*, which aims to resolve the status of each member of a library using the minimum number of tests applied in parallel (Balding and Torney, 1996). To address this issue, the members of the library are distributed into every pool according to a *pooling design*. After testing a set of pools constructed according to a pooling design, it should be possible to determine which pools contain the desired member. This way it can be identified and retrieved from the original library. Pooling reduces the number of tests,

allows for test automation, and addresses the problem of experimental variability (Thierry-Mieg, 2006).

Research in pooling designs has been motivated by genome sequencing projects, which store a genome as a library of fragments of recombinant DNA (several kilobases in size) in clones. These projects usually retrieve a specific sequence from the library. To identify a specific sequence, thousands of clones would have to be analyzed, a process both time-consuming and labour-intensive. Analyzing different pools of clones by PCR or hybridization is significantly more effective than analyzing individual samples, and several pooling designs have been developed (Balding *et al.*, 1996). The field of genetics has benefited from the pooling of individual samples to several areas, such as genotyping (specifically measuring the frequencies of SNPs and microsatellites) and linkage studies (Sham, 2002).

2.2.3. Requirements for pooling in biological systems

In order to pool samples for high-throughput screenings, some requirements apply (Thierry-Mieg, 2006). The library must be available individually in a tagged form. For example, a library of cDNA molecules alone is not exploitable; it must exist as recombinant cDNA in separate clones. Furthermore, it must be possible to test a pool of objects in a single assay and obtain a positive readout if at least one of the objects is positive. For example, this is the case when searching for a specific DNA sequence by PCR in a mixture of molecules. A product will be amplified if at least one of the pooled molecules contains the target sequence. Finally, pooling is especially desirable and efficient when the fraction of expected positives is small (at most a few percent). There are other factors involved as well. The size of the pools imposes a limitation on the construction of pools. Very large pools can affect the detection of an object, which depends entirely on the efficiency of the test (Balding *et al.*, 1996). This is especially important for yeast-two-hybrid screenings, in which the sensitivity of the test decays after more than a hundred strains are pooled.

Pooling designs can detect false positives. This is due to the redundancy of the tests, since each variable is present in several pools, hence tested many times. Nonetheless, reducing experimental variability generally requires testing more pools than the optimal number of pools required to identify an object from the library. For most applications, it is necessary to find a

balance between the number of pools and detection of experimental variability. Some pooling designs are very efficient in terms of numbers of tests but lack the robustness and flexibility that most real biological applications require. Others are more adaptable and noise-tolerant but require the testing of a larger number of pools (Thierry-Mieg, 2006).

2.2.4. Classification of pooling designs

One of the most basic pooling designs is the grid design (**Figure 2.5**). If the library is present in an array format, the pools are constructed according to the rows and columns of the array. The library members from one column are mixed into one pool, and another pool is constructed with the library members of the next column, and so on. Later the pools are tested for the presence of a specific library member (Balding *et al.*, 1996). A positive library member in the array is detected as the intersection of a positive row pool and a positive column pool. However, if more than one positive library member is present in the array, complications can arise. This is the case when a positive row pool intersects with a positive column pool from a different positive library member. It suggests the detection of a non-existing positive hit

Pooling designs used in biology are broadly classified as adaptive or non-adaptive group tests. In adaptive (multi-stage) group tests, the value of the objects is determined progressively after several rounds of tests. The choice of test at any stage depends on the outcomes of previous tests (Balding and Torney, 1996). This approach requires fewer tests than non-adaptive group tests, but it is not suitable for automation since the pools have to be reconstructed depending on the outcomes of the previous tests.

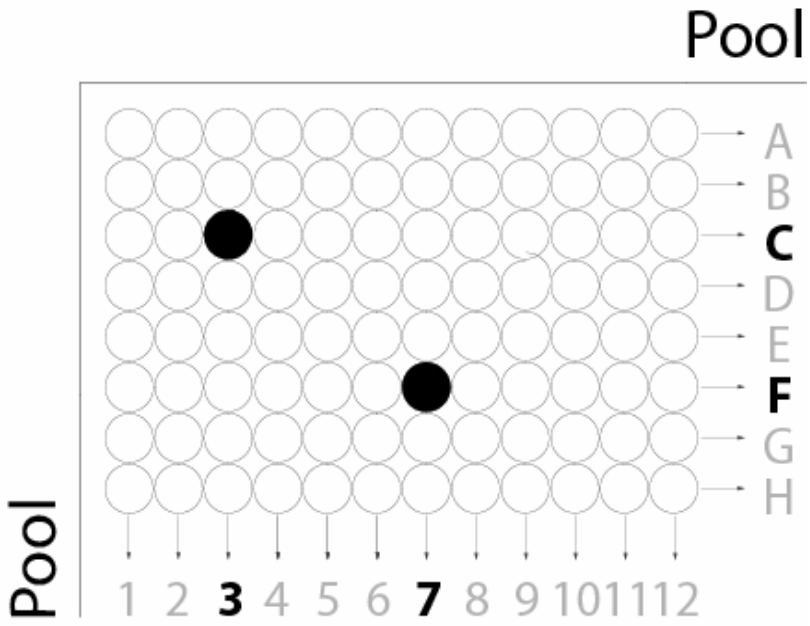


Figure 2.5. Grid pooling design.

The grid is a simple pooling design. The samples are arrayed on a 96-spot plate (grid) and the rows and columns are pooled, creating twenty pools in total. This design is easy to construct although it is susceptible to noise and is not efficient in numbers of tests. In our example, two samples are positive (black circles: row C, column 3 and row F, column 7), and hence the four indicated pools would be positive (bold black letters), in the absence of experimental noise. However two samples can be interpreted as positive (row C, column 7 and row F, column 3), demanding additional tests to confirm the identity of the samples.

In non-adaptive group tests, all tests must be specified without knowing the outcomes of the other tests (Huang and Weng, 2004). This approach is advantageous because the same pools can be tested for all targets and it is automatable. Several non-adaptive pooling designs have been proposed, which are more optimal than the grid design and require strong computational power to construct the pools and decode the outcomes of the tests. Many approaches have been described, which can be broadly classified in two groups, either random or deterministic (reviewed in Ngo and Du, 2000; Balding *et al.*, 1996). In the former, some or all of the entries are randomly determined with parameterized probabilities, which can be optimized based on certain objective function. In deterministic algorithms, the construction of every pool is deterministically specified.

2.2.5.1. An intuitive pooling design: PI-Deconvolution

An intuitive, non-adaptive *transversal* pooling design, named PI-deconvolution, deals very well with the experimental limitations of the yeast-two-hybrid assay (Jin *et al.*, 2006). In a transversal pooling design the pools are divided into separate groups where each group contains all the objects of the library (Du *et al.*, 2006). In the original PI-deconvolution pooling design targets are pooled instead of the library. Several baits are pooled and screened against a cDNA library. Pools are constructed according to *n*-bit binary codes assigned to 2^n baits, ensuring every bait is assigned a *unique* code (**Figure 2.6**). Every bit of the binary code indicates whether a positive hit is present in one of two pools (called pairs), since a specific bait is present in only one pool of the pair; for example, bait *A* will be present in one pool labelled '–' but not in the pool labelled '+'. If an interaction occurs in that pair, the bit will be labelled '–'. After a screening is performed, a code, which is a string of bits, can be deconvoluted for every spot of the array. That code will correspond to a particular bait, since it is known beforehand how the pools were constructed, and there is no possibility that one code correspond to more than one bait (Jin *et al.*, 2006).

In more detail, every pair is constituted by two pools, one referred as positive and the remaining one referred as negative, only for nomenclature purposes. In every pair, half of the baits are pooled in one pool, and the other half is pooled in the other pool; all baits are included in every pair. This process is repeated for every pair, shuffling the baits that are included in the positive or negative pool. A yeast-two-hybrid assay is performed with all pools, mating them with the arrays that contain the library. A positive interaction of a prey with a bait can be detected only in the negative or the positive pair. When these readouts are read along with the other pairs, a code is deconvoluted, which must correspond to a specific bait (Jin *et al.*, 2006). For example, bait *SH3* has been assigned a code '+ – –', and pooled accordingly (**Figure 2.6**). At bit (or pair) 1, denoted by a '+' symbol, bait *SH3* is present only in the '+' pool, and the positive interaction is only detected in that pool, not in the '–' pool.

a)

Bait	Code		
	1	2	3
LexA	+	+	+
CC(aa1-72)	+	+	-
GEF	+	-	+
SH3	+	-	-
SH2	-	+	+
SH1	-	+	-
FABD	-	-	+
Y177	-	-	-

Pair	Pool	Baits
1	+	LexA, CC, GEF, SH3
	-	SH2, SH1, FABD, Y177
2	+	LexA, CC, SH2, SH1
	-	GEF, SH3, FABD, Y177
3	+	LexA, GEF, SH2, FABD
	-	CC, SH3, SH1, Y177

b)

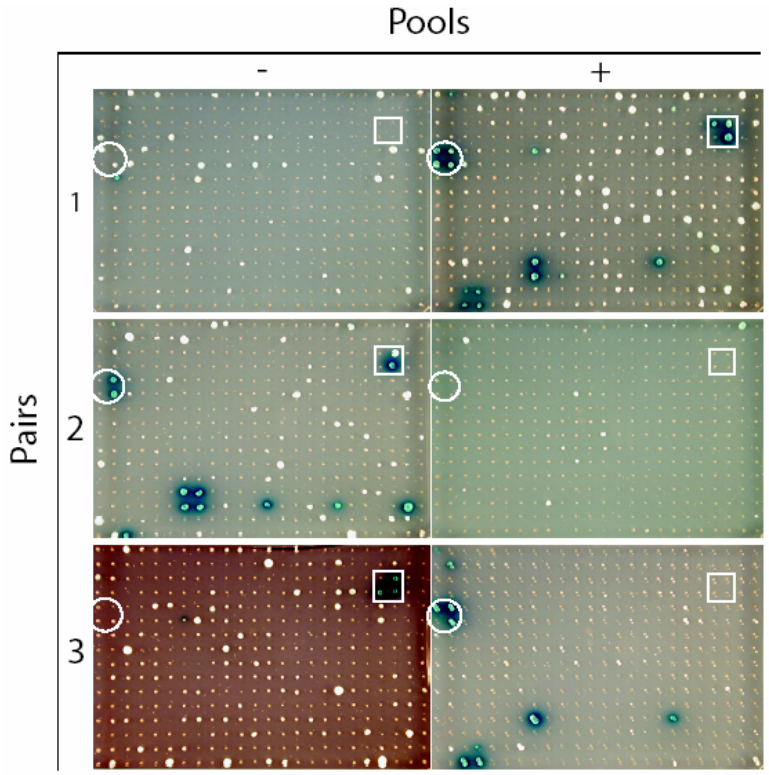


Figure 2.6. PI-deconvolution pooling design.

A set of eight baits are pooled into three different pairs. Each bait is assigned a unique code of three bits (nnn). **b) Yeast-two-hybrid array screening of a lariat combinatorial library by PI-deconvolution.** The '+ – –' profile (squares) allows deconvolution of a single bait (SH3 domain of ABL) within a pool of eight baits, whereas the profile '+ – +' (circles) allows deconvolution of a single bait (GEF domain of BCR). A positive interaction (hit) is assessed when one colony of the array grows and turns blue in sucrose and galactose CSM –HIS –TRP –ADE –LEU +X-Gal plates. In this particular example, every spot of the array was screened against every pool four times.

The efficiency of the approach increases with the number of baits that are pooled. The number of baits that can be pooled (2^n) increases exponentially, whereas the number of pools increases linearly ($2n$) (Jin *et al.*, 2006). Eight baits are pooled into six pools, sixteen baits are pooled into eight pools, thirty-two baits into ten pools, and so on. Additionally, every interaction is tested n times, which improves both coverage and accuracy of the data. Nonetheless, it is possible to observe ambiguous codes, which arise when a positive interaction is detected in the two pools of a pair (false positive), or the opposite case, when no interaction is detected in one pair (false negative). In the latter, the code is not deconvolutable to one bait, although it narrows down the number of possible baits that were responsible for the interaction.

When mapping an interactome, researchers might not be interested in testing more than a few baits. If that is the case, pooling the baits according to PI-deconvolution is not the best choice. A different approach to the screening of libraries with PI-deconvolution consisted of pooling the library instead of pooling the targets (Jin *et al.*, 2007). In their study, an array containing ~6,000 preys (expressing cDNAs) from the yeast genome was first divided into groups of 16, 32 or 64 preys and then every group was pooled according to PI-deconvolution. The arrays were screened against single baits, and the number of interactions detected was similar for the three sizes of groups that were pooled, 16, 32 and 64. Pooling a prey array offers important benefits, such as reducing by about an order of magnitude the number of plates needed to store the array (the original library is contained in sixteen plates of 384 strains each, whereas the same library can be contained in three plates using PI-deconvolution with a pool size of 32), and the possibility of using the pooled arrays for other projects that involve screening of different targets.

PI-deconvolution is a very flexible pooling design. Different applications can obtain different advantages from every pooling design, depending on factors such as variability of the assay, size of the library, expected number of positive hits, and the number of interactions per protein. PI-deconvolution was employed to map a transcription factor regulatory network using the yeast-one-hybrid assay (in this assay the bait is the promoter region located upstream of the reporter genes, which can potentially bind certain transcription factors) is not the best option (Vermeirssen *et al.*, 2007). Deconvolution becomes impossible to solve for baits that interact with multiple transcription factors (TFs), because transcription factors that interact with a bait could be present in the '+' and '−' pools of a given pair. Furthermore, in yeast-one-hybrid

assays many transcription factors can not be detected when diluted more than 27-fold, which limits the size of the pools. Although they could adapt the PI-deconvolution pooling design to the screening of TF libraries by reducing the pool size and performing parallel screens, as in Jin *et al.* 2007, they developed a very effective design based on Steiner Triple System, a block design used in combinatorial mathematics (Vermeirssen *et. al.*, 2007).

The PI-deconvolution pooling design is simple and practical. It is very useful for applications where sensitivity is compromised, such as the yeast-two-hybrid system, and also when the expected number of positive hits is minimal. High numbers of positive hits per pool can lead to ambiguous codes, impeding deconvolution. Its principle that one bait has a unique distribution among the pools has been exploited by other pooling designs (Thierry-Mieg, 2006). More advanced and effective pooling designs have been reported, although the construction of the pools is significantly more complex and they need to be decoded with advanced algorithms.

2.2.5.2. PI-deconvolution for the screening of protein ligand combinatorial libraries

We reasoned that the PI-deconvolution pooling design could be used to screen protein ligand combinatorial libraries using the yeast-two-hybrid system, since it shares two features with the screening of cDNA libraries: The number of expected positive hits per pool is very low (about one hit per one hundred thousand combinatorial prey strains) and false positives are encountered. One major difference in the screening of these two types of libraries is the number of library members that are screened. Whereas the yeast interactome comprises 6,000 proteins, a combinatorial library ideally should comprise more than one million library members to ensure the detection of positive hits. The larger the size of the library, the better chances are to detect protein ligands for the desired targets. It is very useful to apply PI-deconvolution to the screening of combinatorial libraries because several targets can be tested in one run. It presents opportunities for development in proteomics and the validation of drug targets. The ability to screen many baits at the same time in a high-throughput assay could be used to isolate protein ligands for a wide variety of targets, perhaps hundreds, which would be valuable to understand deeply how signaling pathways work and the function of specific proteins. Finally, when there is interest to determine what domains of a protein are responsible for its pathological activity,

or what proteins from a signaling pathway are involved in a disease, it becomes very useful to be able to find protein ligands against a variety of targets.

In order to apply PI-deconvolution to the screening of combinatorial libraries the construction of arrays had to be addressed. Arraying individual strains from a combinatorial library is not possible. At least 10,416 96-well plates are needed to array one million strains (although it would not ensure a complete coverage of the library, some library members will be pooled more than once as well as empty vectors), a quantity that is not achievable for most laboratory infrastructures. Therefore, the library we used had to be pooled in order to handle a smaller number of plates. However, this would bring sensitivity issues, since the yeast-two-hybrid assay does not detect interactions when the interacting prey is diluted more than one hundred-fold in a pool of non-interacting preys. It became necessary to develop a way to screen a condensed combinatorial library in a practical and effective way. The solution to the problem was to make pools of the combinatorial library (without any specific distribution) of about 1,000 yeast cells, so that the arrays could be condensed about 1,000-fold. However, at this dilution it is possible to miss a positive hit because there are many other non-interacting strains present. Therefore, we decided to test every combinatorial pool of the array four times, an empirically validated number of replicates needed to detect at least one positive interaction between the combinatorial protein ligand and one of the baits we pooled according to PI-deconvolution, so a code could be obtained and be deconvoluted to a specific bait.

Our combinatorial protein ligand library is based on an engineered intein scaffold that produces cyclic peptides intracellularly (Scott *et al.*, 1999). Previous work performed in our lab has led to an engineered intein scaffold that displays cyclic peptides that can be linked to the yeast-two-hybrid system (through the formation of a lariat peptide to which attach an activation domain), allowing the screening of cyclic peptides with this method (**Figure 2.3**). We screened a lariat ligand combinatorial library against a set of seven domains of BCR-ABL and a negative control, as a proof of principle for the application of PI-deconvolution to the screening of combinatorial libraries using the yeast-two-hybrid system. In addition, protein ligands that interact with any of the domains and inhibit the oncogenicity of BCR-ABL would be potential candidates for the development of new drugs for the treatment of chronic myelogenous leukemia.

2.3. ROLE OF BCR-ABL IN CHRONIC MYELOGENOUS LEUKEMIA

2.3.1. Chronic Myelogenous Leukemia and BCR-ABL

Chronic myelogenous leukemia (CML) is a myeloproliferative disorder characterized by an excessive production of granulocytic cells (Ren, 2005). The Philadelphia chromosome is associated with 95% of CML cases and results from the reciprocal exchange of DNA between the long arms of chromosomes nine and twenty-two. This chromosomal aberration produces a fusion between the Breakpoint Cluster Region gene (BCR) and the Abelson kinase gene (ABL) (Wong and Witte, 2004), named BCR-ABL. ABL tyrosine kinase is tightly regulated in normal cells, but in BCR-ABL$^+$ cells it becomes constitutively active and brings new regulatory domains/motifs to ABL (**Figure 2.7**) (Ren, 2005). Other mechanisms by which BCR-ABL contributes to the malignant phenotype of myeloid leukemic cells include alteration of the adhesion of the hematopoietic stem cell (HSC) to the matrix and stromal cells of the bone marrow, activation of mitogenic pathways, like Jak/Stat and Ras/Raf/MAPK, and inhibition of apoptotic pathways, such as the PI3K/AKT (Deininger *et al.,* 2000).

A milestone in drug discovery has been the development of inhibitors of the kinase domain of ABL, which effectively controls the CML disease progression. The first to be developed was imatinib, also called gleevec, a compound generated through rational design to inhibit the kinase activity of ABL through occupation of its active site (Capdeville *et al.*, 2002). However, two problems remain to be solved. First, the drug does not eliminate BCR-ABL$^+$ hematopoietic stem cells, therefore the disease is never fully eradicated. On the other hand, mutations of BCR-ABL that affect imatinib binding but maintain the kinase activity are likely to develop, generating resistance to the drug (Hantschel and Superti-Furga, 2004).

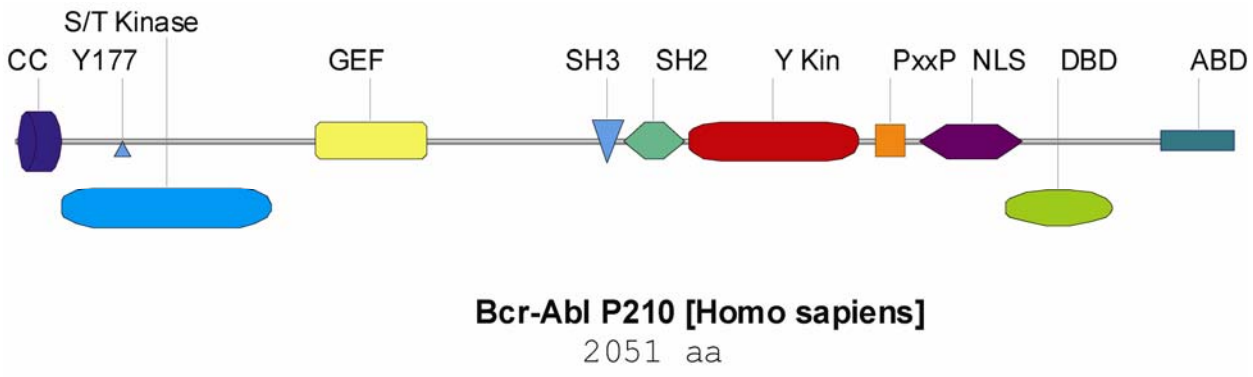


Figure 2.7. Structural organization of domains present in BCR-ABL.

The BCR-ABL protein is the result of a fusion between BCR and ABL that results from a chromosomal translocation. Specifically, the coiled-coil oligomerization domain plays a role in inducing auto-phosphorylation of the kinase. Tyrosine 177 when phosphorylated by BCR-ABL can interact with Grb2, activating the Ras signaling pathway. The GEF domain interacts with Rac1, Cdc42 and RhoA, suggesting a regulatory role for this interaction. As for ABL, the Tyrosine Kinase (Y Kin, SH1 in the text) activates many signaling pathways involved in suppression of apoptosis and proliferation. SH3 and SH2 domains are involved in determining the structural conformation of the kinase and interact with downstream mediators of cell-signaling pathways. Proline rich sequences (PxxP) of BCR-ABL can interact with SH3 domains and Crkl. BCR-ABL has altered cell-adhesion properties, and also possesses other domains, such as Nuclear Localization Domain, DNA-binding domain, and Nuclear-export. Finally in the C-terminal region is located an F-actin binding domain (ABD, FABD in the text), which has been shown to play a role in suppression of apoptosis.

2.3.2. Domains present in BCR-ABL

BCR-ABL is a large chimeric protein. Important domains in BCR include: Coiled-coil (CC) domain, serine/threonine kinase domain, Grb2 binding domain (tyrosine 177, Y177), and a guanidine-exchange factor (GEF) domain. The coiled-coiled domain induces oligomerization of BCR-ABL, which induces autophosphorylation and provides a mechanism for activating the ABL kinase (Smith *et al.*, 2003; Hantschel *et al.*, 2004), as well as activates SH2-binding sites in BCR for interactions with signaling components such as GRB2 and PI3K (Tauchi *et al.*, 1997). The serine/threonine kinase domain has an impaired function in BCR-ABL, where it is phosphorylated at Tyrosine 360, becoming inactive (Wong and Witte, 2004). The phosphorylation of BCR at tyrosine 177 creates an SH2 binding site for GRB2, which in turn creates a complex with SOS and GAB2 through its SH3 domains (Ren, 2005). Formation of

this complex activates the Ras/MAPK signaling pathway through the activation of RAS by SOS. SHP2 and PI3K are further recruited, which in turn activate mitogenic and anti-apoptotic pathways (Pendergast *et al.*, 1993; Puil *et al.*, 1994; Sattler *et al.*, 2002).

The GEF domain of BCR has homology to several RHO specific GEFs (RhoGEFs), including CDC24, DBL, and VAV (Barnes and Melo, 2002). RhoGEFs activate members of the Rho family of small GTPases by catalyzing the exchange of GDP for GTP. Rho family members are involved in regulation of the actin cytoskeleton (Tapon and Hall, 1997), apoptosis (Nishida *et al.*, 1999) and proliferation (Olson *et al.*, 1996). The GEF domain of BCR is responsible for the motility of p210 BCR-ABL leukemic cells (Daubons *et al.*, 2008). It also activates RHOA (Harnois *et al.,* 2003) and has been shown to interact with XPB, a protein associated with increased sensitivity to sunlight and reduced DNA damage repair. This interaction reduces the helicase and ATPase activities of XPB (Takeda *et al.*, 1999).

As most non-receptor tyrosine kinases, the conformation of ABL is very flexible and its kinase activity depends on the conformation of the active site, which can be open (active) or closed (inactive) (Hantschel *et al*., 2004). The SH3 domain of ABL functions as a negative regulator of kinase activity in both active and inactive kinase conformations. Interaction of the SH3 domain with the SH1 kinase domain keeps ABL in an inactive conformation (Hantschel *et al.*, 2003; Barila and Superti-Furga, 1998; Nagar *et al.*, 2003). In the open active kinase conformation, the SH3 domain interacts with proteins that negatively regulate kinase activity (ABI proteins) (Dai and Pendergast, 1995). Similarly, the SH2 domain of ABL functions as a dual regulator of kinase activity in both active and inactive kinase conformations. The SH2 domain has a positive or negative role depending on the conformation of BCR-ABL. It negatively regulates ABL kinase activity by binding to the SH1 domain and maintaining ABL in an inactive conformation (Hantschel *et al.*, 2003; Nagar *et al.*, 2003). Additionally, the SH2 domain contributes to the activation of the RAS pathway through binding SHC, which results in phosphorylation and binding of GRB2 (Goga *et al.*, 1995).

The tyrosine kinase activity of ABL is the most crucial function for BCR-ABL mediated transformation. The SH1 domain has a high degree of homology to the SRC kinase domain, although the mechanism of regulation of its kinase activity by its N-terminal domains is quite different from SRC regulation (Hantschel and Superti-Furga, 2004); these N-terminal domains include N-terminal myristoylation, SH3, and SH2 domains (Van Etten, 1999).

Perturbation of these regulatory domains, as is the case in BCR-ABL where the myristoylation domain is deleted, results in constitutive activation of the SH1 kinase activity. Numerous substrates are tyrosine phosphorylated by BCR-ABL demonstrating the importance of this domain in BCR-ABL transformation (Goldman and Melo, 2003; Goss *et al.*, 2006).

The last exon of ABL contains many proline-rich domains (PxxP), which function as binding sites for SH3 domains of adaptor proteins like CRK, CRKL, GRB2, and NCK (Feller *et al.,* 1994; Smith *et al.*, 1999; Sattler *et al.*, 1996). These complexes activate RAS through the CRKL/SHC pathway (Oda *et al.*, 1994; Pelicci *et al.*, 1995), which in turn activates mitogenic and anti-apoptotic pathways. Furthermore, ABL contains other domains with a variety of activities, such as DNA binding, nuclear export (NES) and localization (NLS), and actin binding (ABD or FABD). The DNA binding domains (three in total) may be involved in initiating transcription, in DNA damage response, and in meiotic processes (Kharbanda *et al.*, 1998; Yuan *et al.*, 1997; Miao and Wang, 1996). The three nuclear localization signals (NLS) and one nuclear export signal (NES) (Taagepera *et al.*, 1998) allow ABL to shuttle between the nucleus and cytoplasm in response to various signals. The F-actin binding domain of BCR-ABL enhances F-actin and cytoskeletal association (Hantschel *et al.*, 2005) and has been shown to play a role in suppression of apoptosis (Underhill-Day *et al.,* 2006).

2.3.3. Inhibition of BCR-ABL domains

It is better to isolate protein ligands from combinatorial libraries against individual domains of a protein, rather than the entire protein. If the entire protein is used, some of the surfaces that could bind protein ligands will not be exposed. Secondly, protein ligands might be isolated against a specific domain that binds them more easily, generating a bias for specific domains. Given its role in CML, BCR-ABL was chosen as a very interesting target for protein ligands that inhibit its function. Its many domains could be screened individually at the same time using the PI-deconvolution pooling design applied to the screening of combinatorial protein ligand libraries. Moreover, the protein ligands isolated against BCR-ABL could inhibit the growth of BCR-ABL$^+$ cell lines, validating specific domains as a target for the development of small molecule inhibitors. In this way, the isolated protein ligands would serve two

functions: Determine which domains are *druggable*, and lead the development of small molecules by competitive binding screenings.

The most important domains for BCR-ABL oncogenesis were selected to be used as baits in the screening of combinatorial protein ligand libraries using PI-deconvolution pooling designs. Only domains that affected the overall structure of ABL kinase (CC, SH3, SH2), were involved in signaling pathways (Tyrosine Y177) or possessed specific functions (GEF, FABD) were considered, along with the ABL kinase domain (SH1). The remaining domains with less important functions, such as the proline-rich regions (PxxP), the DNA-binding domain and the Nuclear localization signal were omitted. PI-deconvolution imposes a limit in the number of domains that can be pooled, therefore seven BCR-ABL domains were chosen to perform a deconvolution screening of a protein ligand combinatorial library using the yeast-two-hybrid assay.

3. OBJECTIVE

The overall objective of this study was to adapt the PI-deconvolution pooling design to the screening of combinatorial protein ligand libraries using the yeast-two-hybrid assay.

To achieve this goal, the following steps were undertaken.

1. Cloning of different BCR-ABL domains into a yeast two-hybrid expression vector

2. Determine the conditions necessary to perform a screening of a combinatorial protein ligand using the PI-deconvolution pooling design (Jin *et al.*, 2006).

3. Perform a screening of a combinatorial protein ligand library using the PI-deconvolution pooling design to isolate lariat ligands that bind to selected BCR-ABL domains using the yeast two-hybrid assay.

4. Develop parameters to evaluate the performance of the screening of a combinatorial protein ligand library based on PI-deconvolution of the baits.

4. MATERIALS AND METHODS

4.1 Common procedures

4.1.1. Cell lines and strains

The strain used in the yeast-two-hybrid assay were *S. cerevisia*e strain EY111 (*MAT*á *his3 trp1 ura3*::*URA3-LexA8op-lacZ ade2*::*URA3-LexA8op-ADE2 leu2*::*LexA6op-LEU2*) and *S. cerevisia*e strain EY93 (*MATa ura2 his3 trp1 leu2 ade2*::*URA3*). The *E. coli* strain MC 1061 (Invitrogen) was used for in making multiple copies of the plasmid of interest, and the strain BL21 (Stratagene) was used to express proteins of interest.

4.1.2 List of Materials

Table 4.1. Table of reagents used in this study.

Item	**Supplier**
Agarose	Invitrogen
Acetic Acid, glacial	EMD
Acrylamide.Bis-Acrylamide [37.5:1] (30%)	Bio-Rad
Adenine	Alfa Aesar
Ammonium persulphate	EMD
Ampicillin, sodium salt	Shelton
Blocking Buffer	Li-Cor
Bromophenol blue	Bio-Rad
Chloroform	BDH
Coomassie Brilliant Blue R-250	Fisher
Disodium Borate	EMD
Dimethyl formamide	Sigma
Dimethyl sulfoxide (DMSO)	Sigma
dNTP set	Fermentas
Ethidium Bromide	BDH

Ethyl Alcohol 95%	Commercial Alcohols
EDTA (ethylenediaminetraacetic acid)	EM Science
Glycerol	EMD
Kanamycin sulphate	Sigma
Lithium Acetate	Sigma
Magnesium Sulfate	EMD
2-Mercaptoethanol	Sigma
PEG (Polyethylene Glycol) 8000	EMD
Phenol:chloforom [1:1]	EMD
Potassium Chloride	EMD
Salmon sperm	Sigma
Sodium carbonate	EMD
Sodium chloride	EMD
SDS (sodium dodecyl sulfate)	EMD
Sodium Hydroxide	BDH
Sodium Phosphate Dibasic	EMD
Sodium Phosphate	EMD
Synthetic Oligonucleotides	Integrated DNA technologies
TEMED	EM Science
TRIS [Tris (hydroxymethyl) aminomethane]	EMD
Triton X-100	Sigma
Trypan Blue Stain	Gibco
Tween-20 (polyoxyethylene (20) monolaureate sorbitan)	EMD
X-Gal (5-bromo-4chloro-indoyl-β-D-galactopyranoside)	Invitrogen

Table 4.2. Table of reagents used in this study.

Item	**Supplier**
10X Buffer EcoRI	Li-cor
10X Buffer Taq Polymerase (platinum, high fidelity)	Invitrogen
Anti-LexA antibody	Invitrogen
Anti-rabbit IRDye 800CW secondary antibody	NEB
EcoRI	NEB
XhoI	Invitrogen
Taq DNA Polymerase (platinum, high fidelity)	NEB

Table 4.3. Table of reagents used in the preparation of media.

Item	Supplier
Agar	BD Biosciences
CSM, drop out supplements	Biol 101 Inc.
Dextrose	BD Biosciences
Galactose	BD Biosciences
Peptone	BD Biosciences
Sucrose	EMD
Tryptone	BD Biosciences
Yeast Extract	BD Biosciences
Yeast Nitrogen without Amino Acids	BD Biosciences

Table 4.4 Table of reagents used in the preparation of media.

Item	Supplier
96-1.2ml tube rack	Molecular Bioproducts
Beckman J2-MI Centrifuge	Beckman
Bransonic 220 Ultrasonic Cleaner	Branson
Centrifuge 5810	Eppendorf
Glass beads, 450-600µm in diameter	Sigma
MaxQ 4000 Shaking Incubator	Barnstead
Microfuge 18 Centrifuge	Beckman Coulter
Micropipettors	Gilson
Micro-pulser Electroporator	Bio-Rad
QIAgen miniprep kit	QIAgen
QIAgen Gel extraction kit	QIAgen
Spectramax 340PC Microplate reader	Molecular Devices
Smartspec 3000 Spectrophotometer	Bio-rad
Ultra-Tech WJ 301D Incubator	Baxter
VP 384F Replicator	V&P Scientific
VP 408FH Replicator	V&P Scientific

4.2. Molecular biology and common procedures

4.2.1. Polymerase-Chain Reaction (PCR)

Plasmid DNA from bacteria was isolated with a commercial kit from QIAgen. Two different polymerases were used to serve different purposes. High Fidelity Platinum *Taq* Polymerase (Invitrogen) was used to amplify DNA used in cloning procedures. See **Table 4.5** for the list of primers used for the amplification of BCR-ABL domains and **Table 4.6** for the amplification of the Ssp-Ssp intein lariat ligand coding region. PCR mixtures in a final volume of 50 µL consisted of 0.4 µM of each primer, Buffer 1X, High Fidelity Platinum *Taq* Polymerase (Invitrogen), 1.5 mM MgSO4, 0.2 mM of each dNTP, 1 unit of Platinum *Taq* DNA Polymerase High Fidelity, ~100 pg of plasmid DNA. *Taq* Polymerase was used to amplify DNA fragments by PCR for purposes different from cloning, such as verifying the presence of a DNA insert within a vector. It was purified by Kris Barreto according to the protocol described by Pluthero (1993). PCR mixtures in a final volume of 50 µL consisted of 0.4 µM of each primer, Buffer 1X (10mM Tris-Cl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 0.2 mM of each dNTP, 2.5 U of *Taq* DNA Polymerase and ~100 pg of plasmid DNA (Ausubel *et al.*, 1988).

Table 4.5. Primers used for cloning of BCR-ABL domains into pEG202 by homologous recombination in yeast.

Primer name	Sequence (5'-3')	Construct (in pEG202)
P1 BCR CC	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc atg gtg gac ccg gtg ggc tt	LexA-CC
P2 BCR CC	tcg ccc gga att agc ttg gct gca ggt cga ctc gag ttac cgg tca tag ctc ttc ttt t	
P1 BCR Y177	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc gac gcc gag aag ccc ttc	LexA-Y177
P2 BCR Y177	tcg ccc gga att agc ttg gct gca ggt cga ctc gag tta aaa ctc gac gtt cac gta	
P1 BCR GEF	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc gtc ctg tcg gga atc ctg gc	LexA-GEF
P2 BCR GEF	tcg ccc gga att agc ttg gct gca ggt cga ctc gag tta att gat gct gga cag gaa gt	
P1 ABL SH1	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc atc acc acg ctc cat tat cc	LexA-SH1
P2 ABL SH1	tcg ccc gga att agc ttg gct gca ggt cga ctc gag tta cat tgt ttc aaa ggc ttg gt	
P1 ABL SH2	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc aac agt ctg gag aaa cac tc	LexA-SH2
P2 ABL SH2	tcg ccc gga att agc ttg gct gca ggt cga ctc gag tta gtt gcg ctt tgg ggc tgg at	
P1 ABL SH3	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc ctt ttc gtt gca ctg tat ga	LexA-SH3

P2 ABL SH3	tcg ccc gga att agc ttg gct gca ggt cga ctc gag tta gac tgg cgt gat gta gtt gc	
P1 ABL FABD	gcg gtt ggg gtt att cgc aac ggc gac tgg ctg gaa ttc tca acc cga gtg tct ctt cg	LexA-FABD
P2 ABL FABD	tcg ccc gga att agc ttg gct gca ggt cga ctc gag tta cct ctg cac tat gtc act ga	

For colony-PCR, a sterile plastic tip was used to pick one CFU from the transformed cells plated on selective media; the CFU was stirred in 50 µL of PCR reaction and then briefly stirred in 5 mL of LB containing the corresponding antibiotic. After the PCR as completed, products were visualized on an agarose gel at 1% (w/v) in order to verify the presence of the insert of interest in the plasmid.

Table 4.6. Primers used for cloning of lariat ligands into pJG4-5 by homologous recombination in yeast and sequence and DNA sequencing

Primer name	Sequence (5'-3')
P1 pJ4-5 pre-EcoRI	tgc gca ccg gac agg aga
P2 pJG4-5 post-XhoI	gca agg tag aca agc cga caa c

4.2.2. Agarose gel electrophoresis

For visualization of DNA, samples were mixed with loading buffer 6X (50% glycerol, 0.2M EDTA pH 8.3, 0.05% bromophenol blue) to a final concentration of 1X, and 5 µL of the mixture were added per lane for PCR products and 1 µL of the mixture for plasmids (50-150 ng/µL). Samples were resolved on the agarose gel at 200 V for 15 to 30 min on SB Buffer (5 mM disodium borate decahydrate, pH 8.0), then visualized and photographed using a UV light Transilluminator (Bio-Rad). Agarose gels consisted of 0.5 to 1.5% (w/v) agarose (ultrapure agarose, Invitrogen), 1X SB Buffer, and 0.5 µg/ml ethidium bromide. For purification of PCR products a kit from QIAgen was used.

4.2.3. Transformation of bacterial cells by electroporation

The procedure for preparing bacterial cells competent for electroporation is described in Ausubel *et al.*, 1988. A colony-forming-unit (CFU) was grow overnight at 37°C in a shaking incubator, in 5 mL of LB. The following day, the cells were grown in 500 mL of LB until an OD_{600nm} of 0.6 was reached. At that point, the cells were centrifuged at 4,000 rpm for 20 min at

4°C and resuspended in ice-cold water. This step was performed twice. The cells were then resuspended in 5 mL of ice-cold glycerol 10% (v/v) and aliquots of 50 µL were stored at −80°C.

Bacterial cells were transformed using electroporation. 1 µL of plasmid (50-150 ng/µL) was mixed with 50 µL of competent cells, and the mixture was transferred to an electroporation cuvette. Cells were then electroporated with field strength of ~12.5 kV/cm, mixed with 500 µL of LB and incubated at 37°C for 30 min in a shaking incubator. Cells were then plated in LB agar plates containing ampicillin (100 µg/µL).

4.2.4. DNA sequencing

DNA samples were sequenced at the Plant Biotechnology Institute, National Research Council of Canada. Plasmids were diluted to 0.05 µg/µL and primers were provided at a concentration of 3.2 pmol/µL.

4.2.5. Western Blot of yeast lysates with LexA

Cell lysates were prepared following the procedure suggested by Kushnirov, (2000). A yeast CFU was grown overnight in CSM −HIS in a shaking incubator at 30°C. Next day, 1 mL of the culture was centrifuged for 1 min at 13,000 rpm, and the pellet was resuspended in 100 µL of ddH_2O, followed by the addition of 0.2 M NaOH. The mixture was incubated for 5 min, and then centrifuged for 1 min at 13,000 rpm. The pellet was resuspended in 50 µL of SDS sample buffer (0.06 M Tris-Cl pH 6.8, 5% Glycerol, 2% SDS, 4% β-mercaptoethanol, 0.0025% bromophenol blue) and boiled for 3 min. SDS-PAGE was performed with a 10% acrylamide gel, where 10 µL of each cell lysate were loaded per lane. The samples were resolved on the acrylamide gel at 200 V for 45 min, and visualized and photographed using a GelDoc imager (Bio-Rad). Resolving denaturing gels consisted of 0.375 M Tris-Cl pH 8.8, 0.1% SDS, 10% degassed acrylamide/bis-acrylamide 37.5:1. For polymerization, 5 µL of TEMED and 50 µL of ammonium persulphate were added to every 10 mL gel mixture. Stacking denaturing gels consisted of 0.125 M Tris-Cl pH 8.8, 0.1% SDS, 10% degassed acrylamide/bis-acrylamide 37.5:1. For polymerization, 10 µL of TEMED and 50 µL of ammonium persulphate were added

to every 10 mL of gel mixture. Afterwards the proteins were transferred at 15 V for 20 min from the gel to a 0.45 μm nitrocelulose membrane (Bio-Rad) using a transblot dry Transfer Cell (Bio-Rad). The membrane was incubated in blocking buffer (Li-Cor) for 1 hr, rocking at room temperature. Afterwards it was incubated with Lex-A antibody (Invitrogen) diluted 1:2,500, rocking at 4°C overnight. Next day, the membrane was washed 3 times with PBS+Tween 20 0.05% (v/v) for 5 min. The membrane was further incubated with the secondary antibody (anti-rabbit from Li-cor, fluoresces at 800nm) diluted 1:10,000 for 1 hr. The membrane was later washed 3 times with PBS + Tween 20 0.1% (v/v) for 5 min. Finally, the membrane was visualized with the Li-Cor Imager at 800nm.

4.3. Protocols for manipulation of yeast cells

4.3.1. Yeast mini-preparation

Plasmids were isolated according to the protocol described in the yeast minipreparation protocol (Geyer and Brent, 2000). A yeast CFU was grown overnight in 2 mL of complete supplement mixture (CSM) media (lacking specific amino-acids, such as histidine or tryptophan) in a shaking incubator at 30°C. Cells were harvested by centrifugation at 13,000 rpm for 30 sec. The cell pellet was resuspended in 200 μL of breaking buffer [2% (v/v) Triton X-100, 1% (v/v) SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA], followed by the addition of 300 μg of glass beads (425-600 nm diameter) and 200 μL of phenol-chloroform-isoamyl alcohol (25:24:1, v/v/v). Cell walls were disrupted by vortexing the mixture for 2 min. Next, the mixture was centrifuged for 5 min at 13,000 rpm, and 50 μL of the aqueous layer were stored. Typically 1 μL of the aqueous solution was used for transforming bacterial cells by electroporation.

4.3.2. Transformation of yeast frozen competent cells

Yeast frozen competent cells were prepared according to the procedure reported by Gietz and Schiestl, (2007). A yeast CFU was grown overnight in 25 mL of YPDA 2X at 30°C in a shaking incubator. Next day the OD_{600nm} of the culture was determined and 2.5 x 10^9 cells

were added to 500 mL of YPDA 2X. The cell suspension was incubated in a shaking incubator at 30°C until an OD_{600nm} of 2 was reached. Then it was centrifuged for 5 min at 4000 rpm, and the pellet was resuspended in 250 mL of sterile ddH_2O. A second centrifugation step under the same conditions was performed, and the pellet was resuspended in 5 ml of sterile ddH_2O. A third centrifugation step under the same conditions was performed, and the pellet was resuspended in 5 ml of sterile frozen competent cell (FCC) solution (5% v/v glycerol, 10% v/v DMSO). 50 µL aliquots were dispensed in 1.5 mL microcentrifuge tubes and stored at −80°C, temperature at which the cells could be stored for a year with little loss of transformation efficiency. DNA transformation consisted of thawing a FCC aliquot at 37°C, following by centrifugation at 13,000 rpm for 2 min. The pellet was resuspended in 260 µL PEG 8000 (50%w/v), 36 µL lithium acetate 1 M, 50 µL single-stranded DNA (2 mg/mL) and 14 µL of 150-300 ng of plasmid (digested or undigested) and PCR product (if *in vivo* cloning was performed) in water. The mixture was vortexed vigorously and incubated in a thermal block at 42°C for 45 min, followed by centrifugation at 13,000 rpm for 30 sec. The pellet was resuspended in 500 µl of the appropriate synthetic dropout media, and different volumes were plated in agar plates containing the appropriate synthetic dropout media, which were incubated at 30°C. Transformants were recovered after 4 days.

The transformation procedure was also employed to clone DNA fragments into a yeast vector by *in vivo* homologous recombination in yeast (Hua *et al.*, 1997). DNA fragments were amplified by PCR, with primers that were 59 bp long, with 39 bp complimentary to the vector from the digestion site and 20 bp complimentary to the region to be inserted. Additional nucleotides were included if necessary to preserve the appropriate reading frame. The plasmid was digested with one or two restriction enzymes, followed by gel purification to remove non-digested plasmids.

4.3.3. Transformation of yeast cells in 96-well format

Transformation of yeast cells in 96-well format was performed according to Gietz and Schiestl, (2007). A yeast CFU was dissolved in 10 mL of sterile ddH_2O and 100 µL of the cell suspension were transferred to each well of a 96-well plate. A replicator (VP 408FH) was used to transfer a small amount of liquid from the 96-well plate to a YPDA plate, so the cells could grow. The resulting array was incubated overnight at 30°C. Next day, cells from the array were

transferred to a 96-well plate containing 100 µL of ddH_2O using the replicator. The plate was centrifuged at 1,800 rpm for 5 min and the supernatant was removed by using a single shake of the plate into a large sink. 50 µL of transformation mix (For 100 reactions: 1.5 mL 1 M lithium acetate, 2 mL single-stranded DNA (2 mg/mL), 0.05 µg plasmid/well in 1.5 mL ddH_2O) were added to each well, shaken at room temperature for 5 min in a shaking incubator. 100 µL PEG 50% (v/v) were added into each well, and then shaken at room temperature for 5 min in a shaking incubator, until the cell suspension became homogenous. The 96-well plate was incubated for 1 hr at 42°C, followed by centrifugation at 1,800 rpm for 10 min. The supernatant was removed by using a single shake of the plate into a large sink, and the pellet was resuspended in 50 µL of ddH_2O. 10 µL of the resulting cell suspension were plated in the appropriate synthetic drop-out plates using a multichannel pipette, and incubated at 30°C. Transformants were recovered after 3 days.

4.3.4. β-Galactosidase Assay

A yeast CFU was grown overnight in complete supplement mixture (CSM) media in a shaking incubator at 30°C. Next day, the culture was diluted to an OD_{600nm} of 0.15 and the cells allowed to grow to an OD_{600nm} of 0.6. 1.2 mL of cells were centrifuged at 13,000 rpm for 5 min, and the pellet was resuspended in Buffer Z (sodium phosphate dibasic (Na_2HPO_4) 0.06 M, sodium phosphate (NaH_2PO_4) 0.04 M, potassium chloride (KCl) 0.01 M, magnesium sulphate ($MgSO_4$) 0.01 M). Cells were centrifuged again as in the previous step, followed by the addition of: 150 µL Buffer Z (plus β-mercapto-ethanol 0.05 M), 50 µL chloroform, and 20 µL SDS 0.1%. The mixture was vortexed, followed by the addition of 700 µL of Buffer Z + ONPG (*ortho*-nitrophenyl-β-galactoside, 1 mg/mL) were added. After 60 to 90 min, the reaction was stopped by adding 500 µL of sodium carbonate (Na_2CO_3 1 M), and then centrifuged for 10 min at 13,000 rpm. 200 µL of the mixture were added to a well of 96-well plate and the absorbance read in a microplate reader (Spectramax 340PC) at OD_{420nm}. Miller units were calculated according to the formula:

$$\text{Miller unit} = \frac{(OD_{420nm})(1000)}{(\text{time})(\text{volume})(OD_{600nm})}$$

Where *time* is reaction time (min) and *volume* is reaction volume (mL).

4.4. High-throughput yeast-two-hybrid assay

4.4.1. Cleaning and use of the replicator

A replicator (V&P Scientific) was used to make high density arrays of yeast CFUs with 96, 384 and 1536 spots per array (VP408FH for 96-spot spot format and VP384F for 384-spot format). To make 96-well format arrays, the replicator was dipped in 96-well round-bottom plates containing 20 to 100 μL of yeast cells, and then pinned into an agar plate. 384-well format arrays were made by pinning four times the cells in suspension or colonies in agar plates, both in 96-well format. To sterilize the replicator between different arrays, it was first dipped in 10% bleach for 30 sec, then dipped in and out in water, and later dipped in 95% ethanol for 1 min, then dried with hot air.

4.4.2. Pooling of baits and preys

For both the baits and the preys, a yeast CFU was grown overnight in the respective synthetic media. Next day, the cells were diluted in the respective synthetic media. After measuring the OD_{600nm} for each culture, the number of cells per mL was estimated. For encoded baits, an equal amount of cells of every bait was mixed, following the PI-deconvolution pooling design (**Figure 2.6**). The mixture was centrifuged for 5 min at 4,000 rpm, and resuspended in freeze-down solution (65% glycerol, 0.1 M $MgSO_4$, 25mM Tris-Cl pH 8.0), then stored at –80°C.

For yeast-two-hybrid assays in which an interacting pair of preys and baits was diluted, we employed the PR domain of RIZ1 fused to the LexA DBD, which interacts with a lariat ligand named 6-49, isolated in previous combinatorial library screenings carried out in our lab. To perform the dilution of baits and preys in a pool of non-interacting strains, the number of cells of the strains was equalized, and then 1:2 or 1:4 dilutions were performed in a 96-well plate.

4.4.3. Selection of positive interactions with the yeast-two-hybrid assay

For both the baits and the preys, a yeast CFU was grown overnight in the respective synthetic media. Next day, the cells were diluted in the respective synthetic media. When the culture reached an OD_{600nm} of 0.5-0.7, the replicator was employed to transfer the cell suspension of the bait or the prey to YPDA plates. After the suspension dried out, a suspension of bait or prey was added in the same position of the previously pinned culture. Plates were incubated at 30°C. After 2 days, in order to select diploids, the YPDA plates were replicated to glucose CSM –HIS –TRP agar plates, keeping a 384-spot format. After another 2 days, the plates were replicated to galactose CSM –HIS –TRP –ADE –LEU agar plates to select positive yeast-two-hybrid interactions. After another 8 days, the plates were replicated to galactose and sucrose CSM –HIS –TRP –ADE –LEU + X-Gal agar plates to select yeast-two-hybrid interactions according to growth and blue color formation of the spots.

4.4.4. Estimation of the probability of detecting at least one positive interaction with different number of replicates using the yeast-two-hybrid assay

The probability of detecting at least one positive interaction with different number of replicates using the yeast-two-hybrid assay for an interacting bait and prey diluted in pools of non-interacting baits and preys, was calculated with binomial coefficients. Experimentally obtained values obtained with the yeast-two-hybrid assay indicated the number of positive and negative interactions observed in a number of spots. The number of spots tested is the total number of spots. The probability of detecting a positive interaction is equal to the ratio of positive interactions over the total number of spots. A binomial coefficient was used to calculate the total number of combinations of spots for different number of replicates, given by the equation:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \; ;$$

where *n* is the total number of spots tested and *k* is the number of replicates.

Second, the number of possible combinations of negative spots were calculated for different number of replicates, that is, combinations that would not contain any positive interaction; according to the combinatorial formula:

$$\binom{r}{k} = \frac{r!}{k!(r-k)!};$$

where *r* is the number of negative spots (in which no interaction was detected), and *k* is the number of replicates.

Once these two values were estimated for every replicate (for 2,3,4…n) the number of combinations of spots that would show at least one positive hit could was calculated as the difference between the total number of combinations of spots and the number of combinatiors of negative spots. With these three values, the ratio of the number of combinations containing at least one positive spot over the total number of combination of spots is equivalent to the probability of detecting at least one positive hit for a certain number of replicates at a given dilution of an interacting prey in a pool of non-interacting preys.

4.5. Screening of a combinatorial lariat ligand library for targets pooled according to PI-deconvolution

4.5.1. Sorting of combinatorial lariat library into independent tubes

The scaffold of the combinatorial library used in our screening was an Ssp-Ssp intein engineered to display lariat peptides as a fusion to the B42 AD (**Figure 2.3**), cloned in the pJG4-5 vector, and transformed into the EY93 yeast strain. The library contains seven randomized amino-acids in the lariat, as well as two extra EY residues. It was created by Kris Barreto and will be referred here after as the *R7 library*. In order to sort the R7 library into individual tubes, 500 µL of the library were dissolved in 100 mL of CSM –TRP +ADE media, and incubated for 1 hr at 30°C in a shaking incubator. Viability and cell density were

determined under a microscope with a hematocytometer, staining the cells with trypan blue. Cells were mixed in a 1:20 ratio with 0.4% trypan blue. The library was diluted in CSM –TRP +ADE media to a density of 62.5 viable cells/μL in a final volume of 400 mL. While the cells were being stirred with a sterile magnetic stirrer, 16 μL of the cell suspension were dispensed in 1.2 mL tubes in a 96-well rack containing 400 μL –TRP +ADE, transferring approximately 1,000 cells per tube. Every cell would ideally contain a different seven amino acid sequence displayed as a constrained peptide. The 96-well tube racks were incubated at 30°C in a shaking incubator for 2 days. 16 μL of the R7 library cell suspension were plated in CSM –TRP +ADE agar plates to determine the number of CFUs.

Bait	Code		
	1	2	3
LexA	+	+	+
CC(aa1-72)	+	+	-
GEF	+	-	+
SH3	+	-	-
SH2	-	+	+
SH1	-	+	-
FABD	-	-	+
Y177	-	-	-

Pair	Pool	Baits
1	+	LexA, CC, GEF, SH3
	-	SH2, SH1, FABD, Y177
2	+	LexA, CC, SH2, SH1
	-	GEF, SH3, FABD, Y177
3	+	LexA, GEF, SH2, FABD
	-	CC, SH3, SH1, Y177

Figure 4.1. Pooling of BCR-ABL domains according to PI-deconvolution.

A set of eight baits, consisting of seven domains of BCR-ABL and a negative control, the DBD of LexA, were pooled according to PI-deconvolution. Every bait was pooled into only one pool of every pair, in three different pairs. Every bait is assigned a unique code of three bits (nnn).

4.5.2. Pooling of BCR-ABL baits according to the PI-deconvolution pooling design

A yeast CFU containing a bait plasmid was grown overnight in 10 mL CSM –HIS. The next day 7 mL of culture were transferred to 25 mL CSM –HIS, and incubated in a shaking incubator for 7 hours at 30°C. Afterwards, cell density was determined under a microscope with a hematocytometer, diluting the cells 1:20 with 0.4% trypan blue. For every pool, equal amounts of bait cells (more than 10 million cells for each bait) were transferred to 250 mL CSM –HIS. The pools were incubated at 30°C in a shaking incubator for 3 days, in order to obtain high-density cultures.

4.5.3. High-throughput yeast-two-hybrid assay to isolate interacting lariat ligands against BCR-ABL baits

In order to obtain a high concentration of cells, every *R7* 96-1.2 mL tube rack was centrifuged at 1,800 rpm for 5 min, 150 μL of media was removed from the tube, and the pellet was resuspended in the remaining liquid and transferred to a 96-well round bottom plate. The same principle was applied for the 4-bait pools, which were diluted 1:10 the day before to a final volume of 300 mL CSM −HIS and incubated in a shaking incubator at 30°C. The six pool cultures were centrifuged for 10 min at 4,000 rpm, and the pellet was resuspended in 50 mL of sterile ddH_2O, concentrating the overnight cultures 6-fold. Afterwards, the *R7* Ssp-Ssp intein library was mated with the 4-bait pools in YPDA agar plates, generating an array of 384 spots, each spot containing approximately 1,000 R7 library members (the exact number of CFUs that grow on CSM −TRP plates at the time the library was sorted) and 4 baits. Every 96-well plate containing the R7 library was pinned using a sterile replicator (VP408FH) in six different YPDA agar plates (one for each pool). Every pool culture was pinned using a sterile replicator (VP408FH) in one of the previously plated YPDA agar plates with the *R7* library. The arrays were condensed to 384 spots per plate instead of 96 spots per plate, by pinning the replicator in four different positions. The YPDA plates were incubated at 30°C for 2 days. Afterwards the YPDA plates were pinned to CSM –HIS –TRP plates and incubated at 30°C for 2 days, in order to exclusively allow the growth of diploids. The CSM –HIS –TRP plates were later pinned to sucrose and galactose CSM –HIS –TRP –ADE –LEU plates, and incubated at 30°C for 10 days.

Afterwards, the sucrose and galactose CSM –HIS –TRP –ADE –LEU plates were pinned to sucrose and galactose CSM –HIS –TRP –ADE –LEU + X-Gal plates, and incubated at 30°C for 3 days. Spots of the arrays that showed growth and turned blue were determined.

4.6. Screening of a combinatorial protein ligand library with individual baits using the yeast-two-hybrid assay

A yeast CFU containing a bait plasmid was grown overnight in CSM −HIS media. Next day, the cells were diluted in 8 mL of the same media. When the OD_{600nm} was 0.5-0.6, 6.5 mL of culture was centrifuged at 4,000 rpm for 5 min. Additionally, 500 µL of *R7* Ssp-Ssp intein library was diluted in 50 mL of CSM –TRP +ADE. The OD_{600nm} of the R7 library was measured and a number of cells, equivalent to one tenth (1/10) of the number of bait cells, were transferred to the bait cells pellet. The mixture was centrifuged for 5 min at 4,000 rpm, and the pellet was resuspended in a volume of sterile water equivalent to the volume of the cell pellet, typically less than 200 µL. The cell suspension was spread uniformly in YPDA agar plates and incubated at 30°C overnight. Next day, the cells growing on the surface of the plate were collected with a sterile glass slide, and centrifuged for 5 min at 4,000 rpm. The pellet was resuspended in glycerol freeze-down solution (65% glycerol, 0.1 M $MgSO_4$, 25mM Tris-Cl pH 8.0) and stored at –80°C. Next day, an aliquot of the frozen cells was diluted 10-fold serially and plated in CSM –TRP and CSM –HIS –TRP, then incubated at 30°C overnight. After two days, colonies from the CSM –TRP and CSM –HIS –TRP plates were counted. The mating efficiency was calculated as the ratio of colonies from CSM –HIS –TRP plates over the number of colonies from the CSM –TRP plates. The number of diploids per µL was estimated from the number of colonies from the CSM –TRP plates.

4.7. Preparation of media

Indications to make synthetic plates and liquid media in the yeast two-hybrid assay are described in Geyer and Brent (2000). Synthetic plates and liquid media were prepared in a 1 L final volume as described below.

YPDA plates and liquid media: Yeast extract 10 g, Peptone 20 g, Agar (plates only) 20 g, H_2O 950 ml, Adenine 40 mg. Autoclave YPD media for 20 min at 1.05 kg/cm^2 and add 50 mL 40% (w/v) dextrose.

Dropout plates and liquid media: Yeast nitrogen base without amino acids (BD Biosciences) 6.7 g, Dropout base (Biol 101 Inc.) *x* g, Agar (plates only) 20 g. H_2O: For dextrose (plates or media) 950 ml or galactose/sucrose (plates or media) 925 mL. Where *x* is equal to the amount of the appropriate CSM (complete supplement mixture) dropout base: 0.77 g CSM –TRP; 0.74 g CSM –HIS; 0.62 g CSM –HIS –LEU –TRP; or 0.61 g CSM –ADE –HIS –LEU –TRP. Autoclave media for 20 min at 1.05 kg/cm^2 and add either 50 mL of 40% (w/v) dextrose, or 50 ml of 40% (w/v) galactose and 25 mL of 40% (w/v) sucrose.

X-Gal plates: Yeast nitrogen base without amino acids 6.7 g, Dropout base *x*, Agar 20 g. H_2O: (glucose plates or media) 850 mL or (galactose/saccharose plates or media) 825 mL. Where *x* is equal to the amount of the appropriate CSM dropout base: 0.74 g CSM –HIS, or 0.62 g CSM –HIS –LEU –TRP, 0.61 g CSM –ADE –HIS –LEU –TRP. Autoclave media for 20 min at 1.05 kg/cm^2 and add either 50 ml of 40% (w/v) glucose, or 50 ml of 40% (w/v) galactose and 25 mL of 40% (w/v) sucrose. Cool media to 55°C and add 10X BU salts (see below). Add 4 mL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal; 20 mg/mL dissolved in dimethyl formamide).

BU Salts: $Na_2HPO_4 \cdot 7H_2O$ 70 g, NaH_2PO_4 30 g, H_2O 900 mL. Adjust 10X BU salts to pH 7.0 and autoclave BU salts for 20 min at 1.05 kg/cm^2.

5. RESULTS

5.1. Cloning of domains of BCR-ABL into a yeast-two-hybrid expression vector

BCR-ABL is an oncogene that results as a consequence of the fusion of ABL and BCR through a chromosomal rearrangement. The tyrosine kinase of BCR-ABL^{+} becomes constitutively active and BCR brings new regulatory domains/motifs to ABL (Ren, 2005). BCR-ABL is responsible for most cases of chronic myelogenous leukemia.

The most important domains for BCR-ABL oncogenesis were selected to be screened against a combinatorial protein ligand library using the PI-deconvolution pooling design. Domains that affect the overall structure of ABL kinase (CC, SH3, SH2), play a role in signaling pathways (Tyrosine Y177) or possess specific functions (GEF, FABD, SH1 tyrosine kinase) were cloned. The position of each domain in BCR-ABL was based on reported data of structural studies and mutation analysis of the activity of the different domains (**Figure 5.1a**). The seven chosen domains of BCR-ABL were cloned as a fusion to LexA DNA-binding domain (DBD) in a yeast-two hybrid expression vector (pEG202). Domains were amplified by PCR from a plasmid containing human BCR-ABL^{p210} and cloned *in vivo* by homologous recombination in yeast (Hua *et al.*, 1997). None of the constructs were found to auto-activate the yeast-two-hybrid reporters and their sequences were verified. The FABD domain contained a mutation of Threonine to Serine in amino-acid position 1,084 of ABL that is not conserved among species distantly related phylogenetically. Furthermore, the expression of the fused domains was confirmed by immunoblot using a LexA antibody (**Figure 5.1b**). The observed molecular weight of all domains corresponds to their expected molecular weight, suggesting the fusion proteins are expressed in-frame and with the expected molecular weight.

a)

Protein	Domain	Position in BCR or ABL*	Size of LexA-domain fusion (KDa)	Reference
BCR	Coiled-coil (CC)	1-72	31.2	Zhao *et al.*, 2002
	Y177	171-182	24.3	Puil *et al.*, 1994
	Guanidine-exchange factor (GEF)	501-589	43.8	Chuang *et al.*, 1995
ABL isoform 1b	Src-Homology 3 (SH3)	65-119	28.7	Gosser *et al.*, 1995
	Src-Homology 2 (SH2)	120-220	34.1	Overduin *et al.*, 1992
	Src-Homology 1, Tyrosine kinase (SH1)	229-515	55.8	Schindler *et al.*, 2000
	F-Actin binding domain (FABD)	1026-1149	36.2	Hantschel *et al.*, 2005

*NCBI accession numbers. BCR: gi29241, ABL: gi62362411.

b)

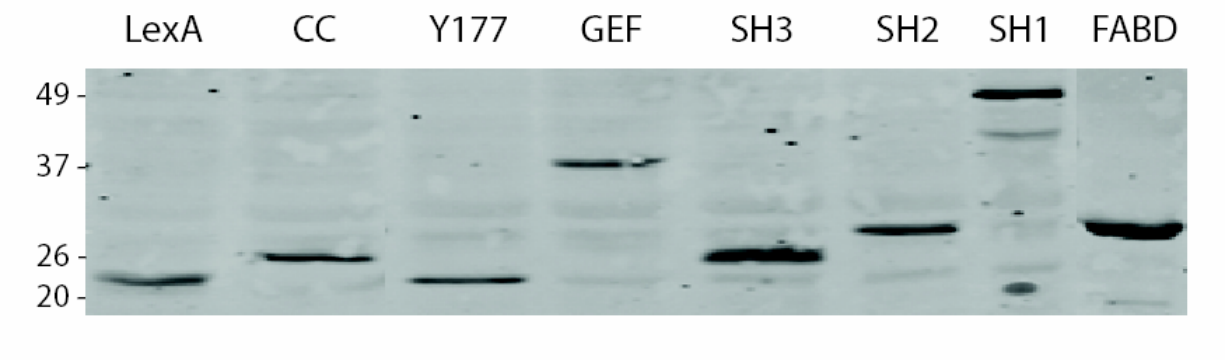


Figure 5.1. Expression of LexA- fusion BCR-ABL domains.

a) List of domains of BCR-ABL that were cloned as a LexA-domain fusion into a yeast-two-hybrid expression vector. Position within the protein sequence of BCR or ABL is shown, as well as the expected molecular weight of the fusion proteins. **b)** Western Blot with an anti-LexA antibody of yeast lysates from cells expressing LexA-domain fusions. The molecular weight of the fusion proteins is shown in panel a). The molecular weight of a protein ladder is shown to the left of the blot. All the domains showed the expected molecular weight.

5.2. Detection limits of the yeast-two-hybrid assay

Two issues must be addressed in order to screen protein ligand combinatorial libraries using PI-deconvolution. The first issue involves pooling the baits according to PI-deconvolution; it pre-determines the number of baits to be pooled following the exponential 2^n, (*i.e.* 8, 16, 32, 64, 128, etc.). The second component is a matrix-based screening of the protein ligand combinatorial library. The large number of members a library of this kind contains, around 10^7 transformants, makes it necessary to pool library members for every 'spot' of the array (**Figure 2.4**). Otherwise, thousands of 96- or 384-spot arrays would be needed to perform one single screening. Therefore, the number of cells per spot for every array should be condensed (or pooled) at least 100-fold, so fewer plates are required to be assayed. The ideal scenario would be to sort the library in an array in which every spot would contain between 100 and 2,000 members. This way, the library could be stored in a reasonable number of 384-spot arrays. The detection of protein-protein interactions using the yeast-two-hybrid assay decreases when the interacting bait and prey are diluted more than 100-fold in a pool of non-interacting baits or preys. We needed to determine how diluting an interacting prey in a pool of non-interacting preys would affect the performance of the yeast-two-hybrid assay. Note that in study the baits would be pooled according to PI-deconvolution, whereas the library members would not be pooled according to any pooling design. They would be pooled randomly using a certain density of preys per spot on the array.

To establish the dilution threshold of the yeast two-hybrid assay we used a previously known interacting prey and bait. More specifically, a lariat ligand referred to as 6-49 (prey) that interacts with the PR domain of RIZ1 (bait). It was isolated previously in the Geyer lab from a combinatorial lariat ligand library screening using the yeast-two-hybrid assay. We determined the sensitivity of the yeast-two-hybrid assay by diluting the prey 6-49 and the RIZ1 bait in a pool of strains containing non interacting preys (pJG4-5) and baits (pEG202-SH1) (**Figure 5.2a**). A yeast-two-hybrid assay was performed to identify the dilutions where the interactions are no longer detectable. The results show how diluting the interacting baits or preys affects the number of positive hits detected in a negative way (**Figure 5.2b**). Nonetheless, the assay does detect the interaction of RIZ1 with 6-49 when the 6-49 prey is diluted between 500- and 1,000-fold and the RIZ1 bait diluted between four- and eight-fold per spot. The data confirms that the

yeast-two-hybrid assay loses sensitivity when the interacting bait and strain are diluted in a pool of non-interacting baits and preys.

a)

PREY DILUTION: 6-49 (aptamer vs. RIZ) diluted in pJG4-5 (empty prey vector)													
	Riz/SH1	1	2	3	4	5	6	7	8	9	10	11	12
6-49/pJG4-5		1	(1:2)	(1:4)	(1:8)	(1:16)	(1:32)	(1:64)	(1:128)	(1:256)	(1:512)	(1:1024)	0
A	1	(+)											
B	(1:2)												
C	(1:4)												
D	(1:8)												(-)
E	(1:16)												
F	(1:32)												
G	(1:64)												
H	0	(-)											

b)

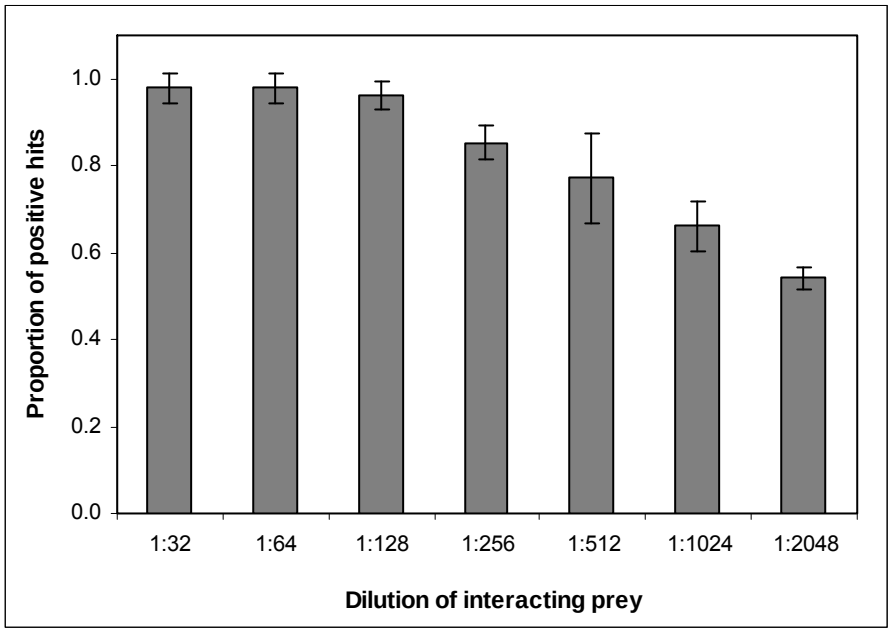


Figure 5.2. Sensitivity of the yeast-two-hybrid assay.

a) A prey strain containing a lariat peptide (6-49) that interacts with RIZ1 was diluted 2-fold in an equivalent number of non-interacting strains (prey harbouring the pJG4-5 vector, does not express a lariat peptide). In a similar fashion, a RIZ1-expressing bait was diluted 2-fold in a bait expressing SH1, which does not interact with the lariat peptide 6-49. The number of positives hits detected decrease as you move diagonally in the array (arrow). **b**) Percentage of positive hits detected for a serially diluted interacting prey in a pool of non-interactors. Interacting bait was diluted 1:4 in a pool of non-interactors. Experimental values from experiment shown in a).

5.3. Probability of detecting at least one positive hit

We have shown that a yeast-two-hybrid assay with an interacting prey and bait diluted in a pool of non-interacting strains can be performed. The possibility to also pool a library as hundreds of library members per spot enforces our aim to develop a high-throughput combinatorial screening of lariat ligand libraries. In order to unambiguously deconvolute a code to one target, the value of all the bits from the code must be determined. In other words, an interaction must be detected for every pair in only one of two pools. If the interaction of a diluted bait and prey is tested many times, or several replicates are tested, the probability of detecting it increases. We wanted to determine how many times (or replicates) an interaction should be tested in order to detect at least one positive interaction.

For the experimental data shown on **Figure 5.2b**, the proportion of positive hits is equal to the probability of detecting an interaction for only one spot (replicate). For example, for the 1:024 dilution a proportion of 0.659 was observed, because fifty-eight spots were positive out of eighty-eight (**Figure 5.3a**). If the experiment is repeated again once for that dilution, it can be seen as taking only one spot of the eighty-eight spots tested. If various spots are tested again, it can be seen as picking certain combinations of those eighty-eight spots, depending on the number of spots tested (replicates). If three replicates were tested, it can be seen as taking any three spots from the 88 spots originally tested. Whereas some combinations will contain positive hits and some will contain only negative hits, in our case we want to detect spots that contain at least one positive interaction. To do so, we used a binomial coefficient to determine all the possible combinations of spots for those three replicates. Afterwards, we determined all the possible combinations of spots that would contain only negative interactions, also using a binomial coefficient. When the total number of negative spots is subtracted from the total number of spots the result is all the possible combinations of spots that contain at least one positive hit. This value was divided by the total number of spots to estimate the proportion of positive hits, the same as the probability of detecting at least one positive hit per spot for an interacting bait and prey for replicates of an interaction. (**Figure 5.3a**). It can be seen that after eight replicates all the dilutions tested reach a probability equal to one, which means that having eight replicates for one spot will ensure detecting at least one positive hit, a condition

necessary to apply the PI-deconvolution pooling design to the screening of combinatorial libraries (**Figure 5.3b**).

a)

Number of replicates	Total combinations of spots	Combinations of negative spots	Combinations of at least one positive spot	Probability of detecting at least one positive spot
1	88	30	58	0.659
2	3828	435	3393	0.886
3	109736	4060	105676	0.963
4	2331890	27405	2304485	0.988
5	39175752	142506	39033246	0.996
6	541931236	593775	541337461	0.999
7	6348337336	2035800	6346301536	1.000
8	64276915527	5852925	64271062602	1.000

b)

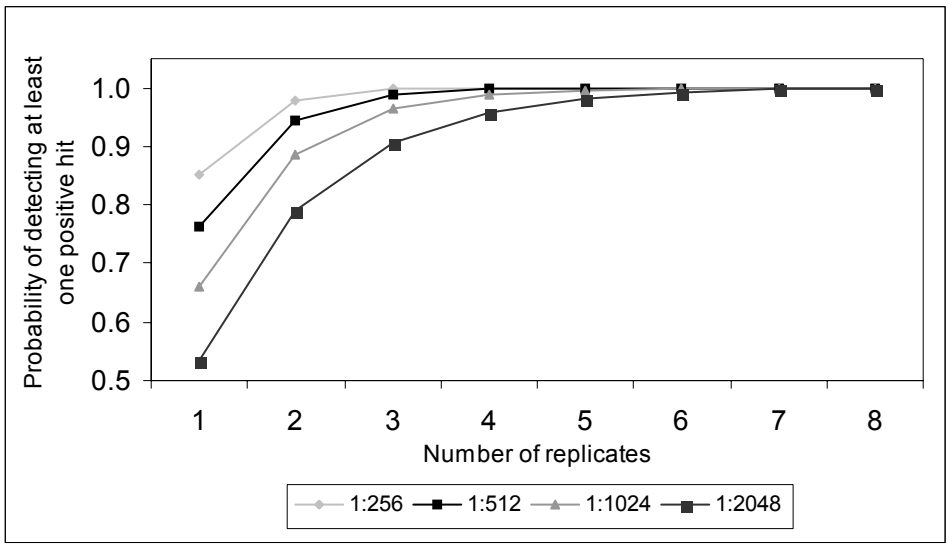


Figure 5.3. Probability of detecting at least one positive hit for different dilutions of an interacting prey in a pool of non-interacting preys with different replicates of a spot.

a) Example of the combinatorial algorithm, values obtained when a yeast-two-hybrid assay performed with the interacting bait (RIZ1) was diluted 1:4 and the interacting prey (6-49) was diluted 1:1024. **b)** The number of replicates for every spot needed to detect at least one positive hit for different dilutions of an interacting prey in a pool of non-interacting preys and an interacting bait diluted 1:4. Values calculated using a combinatorial formula.

We wanted to determine the optimal dilution of preys to perform our screening and the number of replicates it required. We aimed to screen one million library members. In a screening, there is a strict relationship between the dilution of preys, the number of replicates, and the number of arrays. The fewer arrays needed to contain one library, the more efficient the screening will be in terms of cost and time (**Figure 5.4**). We chose to pool about 1,000 cells per spot. In this dilution, detecting at least one (+) hit with a probability higher than 0.98 per spot requires testing at least four replicates. Following these parameters, one million library members would be contained in sixty-one plates (arrays). We could also have chosen the 1:2,048 dilution, however this choice would imply the need for five replicates to detect a positive hit with a probability of 0.98, thus the construction of the arrays would be difficult because they must be constructed in series of two or four. For example, one 96-well array containing 96 different samples can be converted into four replicates by condensing it 4 times into a 384-spot array. Making four replicates per spot is easier than constructing other number of replicates, because one 384-spot array is made from condensing four 96-spot plates using a pin tool. This way one 96-well plate with 1,000 library strains per well has to be replicated four times into an array plate. Overall, the conditions for the screening of a combinatorial protein ligand library against a pool of four baits had been established: Approximately 1,000 cells from the library would be used per spot, and every spot would contain four replicates.

5.4. High-throughput yeast-two-hybrid assay with an interacting bait pooled according to PI-deconvolution

Once the optimal parameters for the screening a combinatorial protein ligand library using four pooled baits (from the 1:4 dilution) were established, we performed a small-scale mock screening with lariat ligand 6-49 and RIZ1. In this case, RIZ1 was pooled according to the PI-deconvolution pooling design for eight baits, which determined that the interacting bait should be diluted 4-fold in a pool of non-interacting baits, instead of diluting 2-fold serially as in previous experiments. Positive hits are found in the pools in which RIZ1 is present, unambiguously deconvoluting to the code '− + −', which corresponds to RIZ1 (**Figure 5.5**). Dilution of the preys to more than 1:64 decreases the number of positive hits detected. Almost no interactions can be detected when diluting the preys more than 1:2,048.

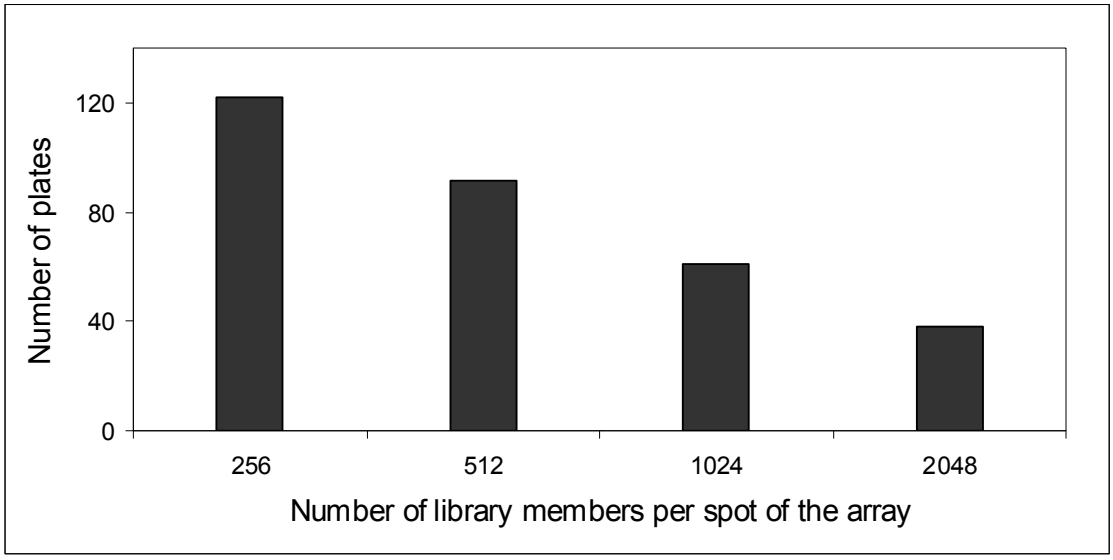


Figure 5.4. Number of plates required to screen one million members of a combinatorial library.

The number of plates required for every step of the screening of pools of four baits against one million members of a combinatorial lariat ligand library was calculated based on the number of replicates needed to detect at least one positive interaction with a probability of 0.98. A step is defined as the media in which the cells will be replicated, for example from glucose CSM –HIS –TRP to galactose and sucrose CSM –HIS –TRP –ADE –LEU plates.

Nonetheless, we observed some positive interactions in spots in pools where the interacting bait (RIZ1) was not present. This is likely due to appearance of false positives. However, as the 6-49 lariat ligand is diluted more than 1:128, positive hits are not detected where the bait is not present (**Figure 5.6**). This suggests that pooling several lariat ligands in one spot contributes to the elimination of false positives. The results confirm that several baits can be effectively screened with combinatorial protein ligand libraries using the PI-deconvolution pooling design. The parameters we have established are appropriate to ensure the detection of all the interactions between an interacting bait and prey when they are diluted in a pool of non-interacting baits and preys.

a)

Bait	Pair		
	0	1	2
LexA	+	+	+
GFP	+	+	-
JH1 (JAK2)	+	-	+
GST	+	-	-
SH2 (ABL)	-	-	+
SH1(ABL)	-	-	-
CDK2	-	+	+
RIZ1	-	+	-

b)

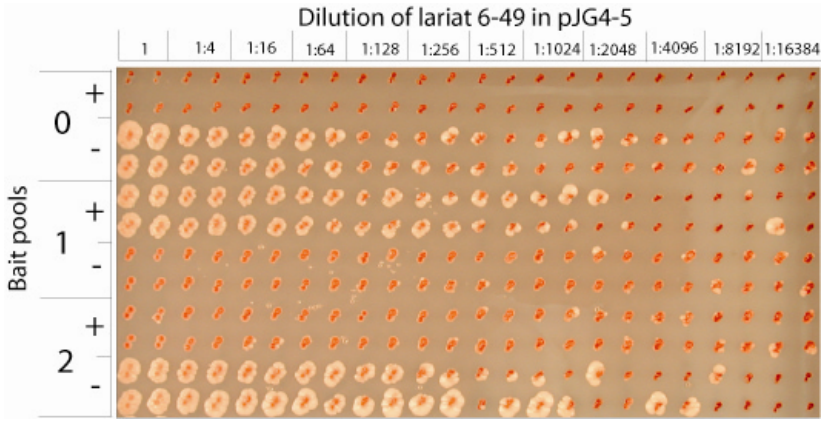


Figure 5.5. A mock screening of a known interaction between RIZ1 and peptide lariat 6-49 pooled with non-interacting baits and preys using PI-deconvolution.

(a) Eight baits were pooled-encoded into 3 pairs of (+) and (−) according to the PI-deconvolution pooling design. RIZ1 was encoded as '− + −'. Other proteins expressed as fusion to the LexA DNA-binding domain included green fluorescent protein (GFP), JH1 domain of janus kinase 2 (JAK2), glutathione-S-transferase (GST) and cyclin-dependent kinase 2 (CDK2) **(b)** A strain expressing a lariat peptide (6-49) that interacts with RIZ1, was serially diluted 4-fold and 2-fold in a pool containing a non-interacting strain (pJG4-5). The pools that grew in auxotrophic media are encoded as '− + −', which deconvolutes to RIZ1. Detection of (+) hits decreases as the strain expressing 6-49 is diluted. Each spot of the array was tested four times.

5.5. PI-deconvolution screening of a lariat ligand combinatorial library

We screened a lariat ligand combinatorial library against eight baits pooled into six pools according to PI-deconvolution with the yeast-two-hybrid assay (**Figure 4.1**). Baits screened included domains of BCR-ABL (CC, Y177, GEF, SH3, SH2, SH1, and FABD) and the DNA-binding domain of LexA. The DBD of LexA was used as a negative control to identify lariat ligands that interact against the DNA-binding domain of LexA, not with BCR-ABL domains. The scaffold of the lariat ligand combinatorial library was an engineered Ssp-Ssp intein that displays cyclic peptides as lariats. It contained seven randomized positions in addition the fixed amino-acids EY in the cyclic peptide formed by the intein (**Figure 2.3**).

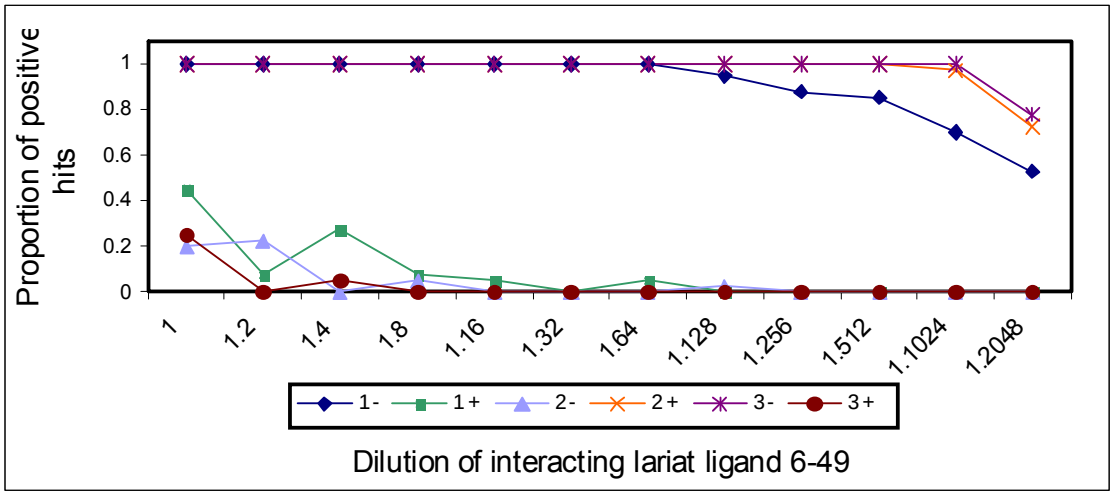


Figure 5.6. Diluting an interacting lariat ligand in a pool of non-interacting preys eliminates the false positives that arise during a high-throughput yeast-two-hybrid assay with baits pooled according to PI-deconvolution.

Proportion of (+) hits from a mock deconvolution combinatorial library screening for PI-encoded pools (4 baits per pool) with different dilutions of the interacting prey. A yeast-two-hybrid high-throughput assay was performed with 6 pools, each containing 4 baits. Baits were pooled according to PI-deconvolution. Pools consisted of three pairs, each containing a '+' pool and a '– ' pool, labelled 1–, 1+, 2–, 2+, 3–, 3+. RIZ1 was only present in the pools 1–, 2+ and 3–. It interacts with the lariat ligand 6-49, diluted serially in a pool of non-interacting preys. Positive interactions should be only observed in the pools where RIZ1 was included, although some positive interactions were detected in the pools were RIZ1 was not present, due to the intrinsic noise of the assay. However, as the 6-49 lariat ligand is diluted, the unexpected interactions are no longer detected.

The screening process is outlined in **Figure 5.7**. Approximately one million members of the combinatorial library were screened. The library was sorted in 1,536 spots from sixteen 96-tube racks, where every spot of the array contained approximately 750 cells per spot. Every spot was plated four times, thus four replicates of every spot were screened. The sorted library was mated with every pool (6 pools in total), and diploids that carried the bait and the prey plasmids were selected. The diploid colonies were plated to agar plates containing galactose and sucrose CSM –HIS –TRP –ADE –LEU, where only colonies with positive interactions (hits) are able to form colonies. Afterwards, the arrays were plated to plates containing galactose and sucrose CSM –HIS –TRP –ADE –LEU +X-Gal, where positive interactions were selected by appearance of blue colonies.

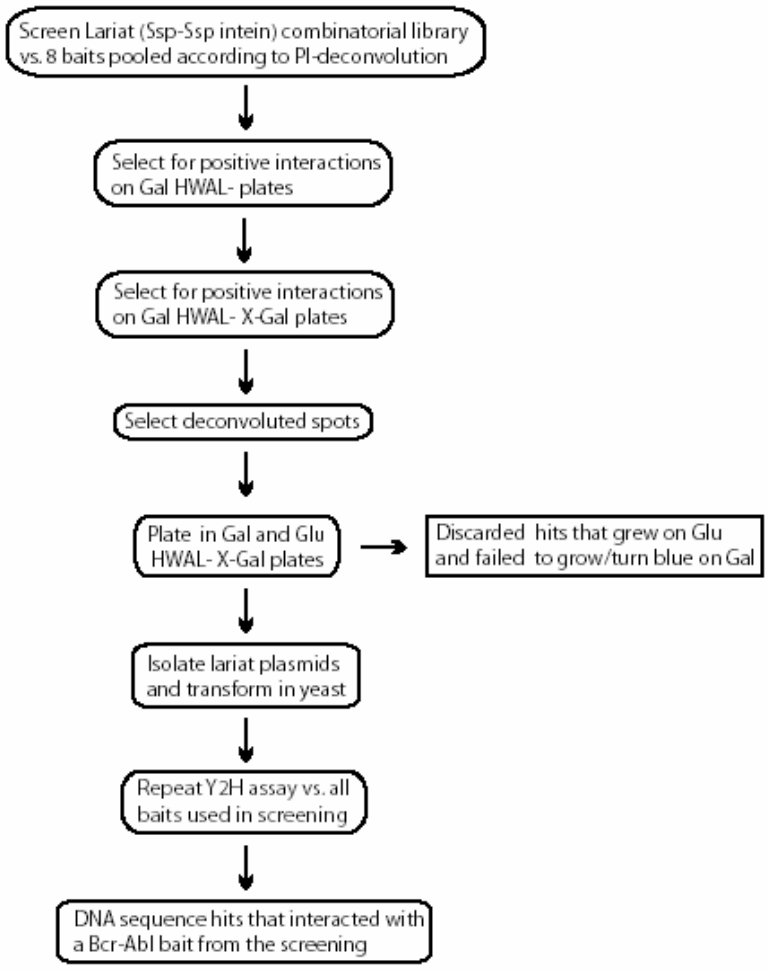


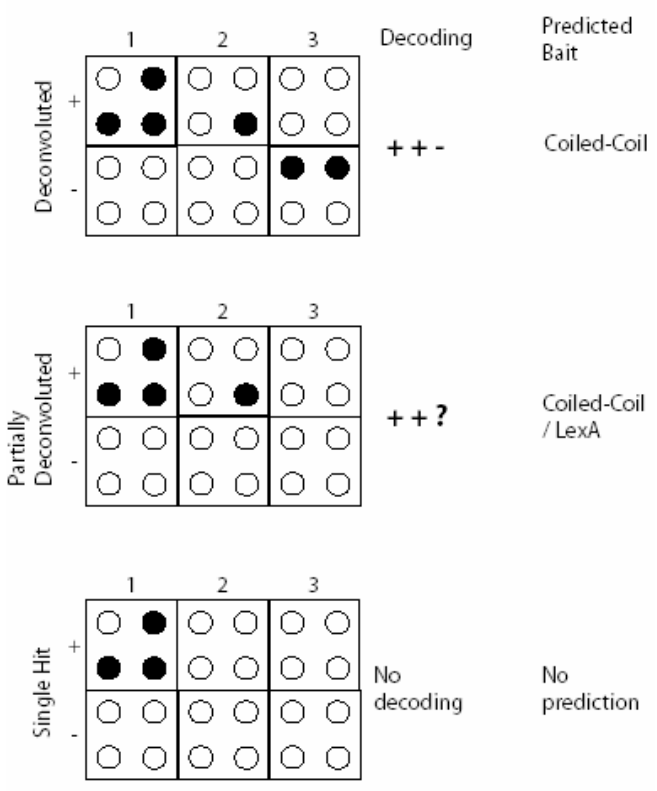
Figure 5.7. Screening of a combinatorial library using PI-deconvolution.

A lariat ligand combinatorial library is mated to six pools that contain eight baits pooled following PI-deconvolution. Positive interactions (hits) are selected by growth in plates lacking leucine and adenine, and later in plates containing X-Gal, in which the positive interactions turn blue. The results from the –ADE –LEU and X-Gal selection plates are combined and from there those spots that show PI-deconvolution are selected. Positive hits from those spots are plated into sucrose and galactose (gal) CSM –HIS –TRP –ADE –LEU +X-Gal and glucose (glu) CSM –HIS –TRP –ADE –LEU +X-Gal to eliminate false positives; colonies plated in Glu CSM –HIS –TRP –ADE –LEU +X-Gal will not grow if the activation of the reporter genes depends on the interaction of the bait and the prey. Furthermore, hits grown in sucrose and galactose CSM –HIS –TRP –ADE –LEU +X-Gal should grow and turn blue again. The lariat plasmids are isolated and retransformed into an empty yeast strain, followed by a second yeast-two-hybrid assay to confirm the previously detected interactions. Finally, the DNA sequence of the lariat encoding region is obtained from the confirmed positive interactions.

Positive interactions from both galactose and sucrose CSM –HIS –TRP –ADE –LEU plates with and without X-Gal were determined. Bear in mind that one spot could present up to

four positive hits, because every spot was tested four times (four replicates). In order to analyze the data, the positive hits from all the spots (four hits per spot) were grouped in either one of three categories (**Figure 5.8a**): Single hits, partially deconvoluted hits, and deconvoluted hits. Single hits appeared in only one out of the six pools. Partially deconvoluted hits appeared in only two out of the three pairs (*i.e.* following the code '− + ?', or '+ + ?', '?' represents no interaction detected). Deconvoluted hits appeared in all three pairs and only one pool of each pair (either in the – or + pair, but not both), hence showing a deconvoluted pattern (*i.e.* + – +, + + –) that can be uniquely associated with one of the eight baits included in the pools. A fourth category, named *not deconvoluted*, was assigned to positive interactions in spots in more than one pool of the same pair; these are potential false positives. A total of 1432 positive interactions (positive hits) were found in 775 spots. The majority of positive hits were present as single hits, and a minor proportion (53 out of 775), were deconvoluted (**Figure 5.8b**). Deconvoluted positive hits are likely true interactors because they show a reproducible and consistent deconvoluted pattern (Jin *et al.*, 2006).

a)



b)

	Deconvoluted	Partially deconvoluted	Single hits	Not deconvoluted	Total
Spots	53	127	559	36	775
(+) Hits	321	311	631	169	1432

Figure 5.8. PI-Deconvolution categories.

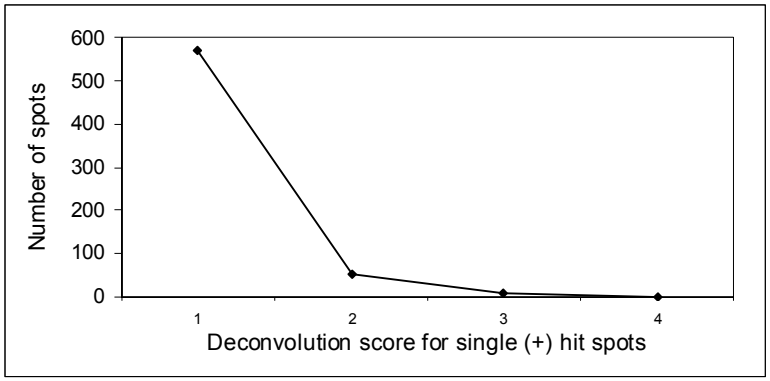
Criteria developed to evaluate the outcome of the screening of a combinatorial lariat ligand library using PI-deconvolution. **a) Deconvolution categories.** A spot is labelled *deconvoluted* after presenting at least one positive hit in only one pool (+ or –) or every pair (1/2/3); this way the spot can be assigned a 'code', in the example being '+ + –', which decoded to the coiled-coil domain. If one of the pairs lacks one positive hit, it is considered *partially deconvoluted*. In this case, deconvolution can be applied and could be decoded to two baits (out of eight). A third case, if a spot presents a positive hit in only one pool out of the six pools, it is considered *single hit*, which can not be assigned any code, and most probably represent a false positive. **b) Summary of deconvolution categories of a PI-deconvolution screening of a combinatorial protein ligand library**. The number of spots for each category and the number of hits (replicates) that appeared in every spot are shown. Every spot could present up to four hits. A fourth category consists of *not deconvoluted* spots, in which a positive hit was observed in both pools of one pair, in the '+' and the '–' pool.

5.6. Deconvoluted spots showed high deconvolution scores

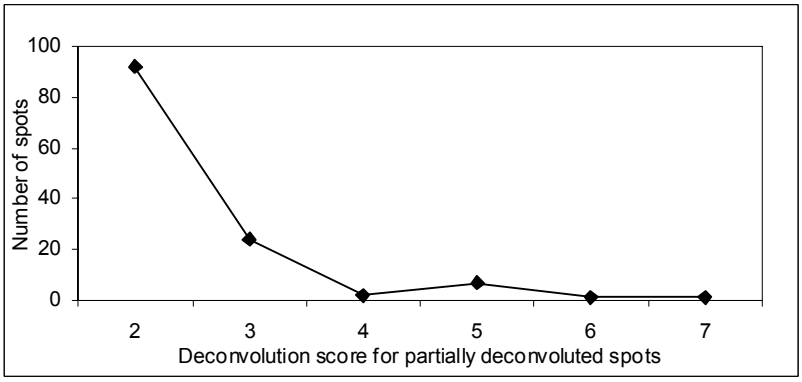
Another parameter that we developed to evaluate the outcome of the screening of protein ligand combinatorial library is the *deconvolution score*. For the screening of a combinatorial lariat ligand library, we reasoned that for a deconvoluted spot, the more hits that were detected, the more likely they would be a true interaction. In order to verify this hypothesis, every spot that showed positive interactions was assigned a deconvolution score. This score reveals the number of positive hits detected per spot, since every spot has four replicates. For the conditions used in our screening, the optimal score would be twelve, since for one spot, which is represented in three pools, there would be four positive hits per pool (from the four replicates). Therefore, the *deconvolution score* would be twelve. Analyzing the deconvolution score of all the spots could provide more information about the positive hits isolated in the screening rather than focusing on the deconvolution pattern alone (**Figure 5.9**). Single hits comprised the majority of the hits detected in the screening. The vast majority were present as isolated single hits, with very low deconvolution scores. The same behaviour was

observed for the partially deconvoluted spots, in which the vast majority of the spots showed an interaction only in one or two of the replicates. For deconvoluted spots a different trend is observed, since spots with high deconvolution scores were observed.

a)



b)



c)

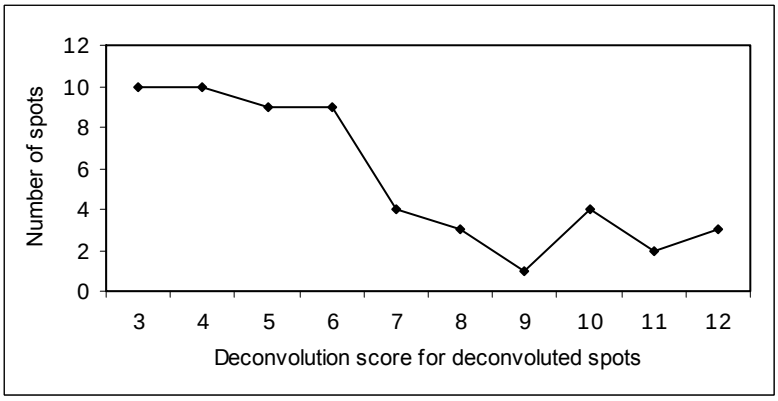


Figure 5.9. Distribution of the deconvolution score for spots that showed positive interactions after a screening of a combinatorial lariat ligand library for different deconvolution categories.

Deconvolution score depends on the deconvolution category. For example, for single hits, the score ranges from 1 to 4 since for one spot there are only 4 replicates. For partially deconvoluted hits, the score ranges from 2-8, since there are two spots and 4 replicates per spot; also there must be at least 2 hits (one hit in every pool) for the spot to be classified as partially deconvoluted. **a**) Distribution for single hits. **b)** Distribution for partially deconvoluted hits. **c)** Distribution for deconvoluted hits.

5.7. Confirmation of the interaction of deconvoluted positive spots with their predicted target

Due to the appearance of false positives in yeast-two-hybrid screenings, it is necessary to confirm detected interactions, a process that is both time-consuming and labour-intensive. To confirm a yeast-two-hybrid interaction, the prey plasmid must be isolated from the colonies (diploids) that showed a positive interaction, inserted again in a prey yeast strain, followed by a second yeast-two-hybrid assay with the bait it interacted with in the screening. Elimination of false positives is one of the reasons that lead us to adapt the PI-deconvolution pooling design to the screening of combinatorial protein ligand libraries using the yeast-two-hybrid assay. Therefore, we tested detected interactions from the screening to evaluate the usefulness of the approach and verify that positive hits isolated were indeed true positives. A 96-spot array was created with randomly selected colonies from deconvoluted spots (arrayed in duplicate). Spots that were not be deconvoluted were also included in the array, the case of 6[4H], 10[10D], 12[7H]; or partially deconvoluted, such as 6[8F] and 7[10D] (**Table 5.1**). Positive and negative controls were included as well to fill all 96-spots of the array (not shown). The 96-spot format facilitated handling of the samples with a 96-pin replicator or a multichannel pipette.

Galactose-dependence yeast-two-hybrid expression of the reporters was tested with the spots from the array. The expression of the lariat ligands is only activated by galactose and inhibited by glucose; therefore colonies that grow on plates that contain glucose are not expected to be true yeast-two-hybrid interactors. The array was replicated to galactose and sucrose CSM –HIS –TRP –ADE –LEU plates, and glucose CSM –HIS –TRP –ADE –LEU plates. Only spots that grew on galactose and sucrose CSM –HIS –TRP –ADE –LEU plates but not in glucose CSM –HIS –TRP –ADE –LEU plates were considered yeast-two-hybrid positives, therefore the interaction of the lariat ligand with its target was expected to be

reproduced. Colonies from spots that grew on glucose CSM –HIS –TRP –ADE –LEU plates or failed to grow in galactose CSM –HIS –TRP –ADE –LEU plates were not expected to be reproducible yeast-two-hybrid positive hits. After the confirmation of galactose-dependent expression of yeast-two hybrid reporters, the interactions of the 96 spots of the array were further confirmed by isolating the prey plasmid and reinserting it in a prey strain, followed by a second yeast-two-hybrid assay. Only the interactions of colonies from thirty-one spots that passed the galactose-dependency test were positive for this second yeast-two-hybrid assay (**Table 5.1**). The interactions of the spots that showed gal-dependence were confirmed. These spots were either deconvoluted or non deconvoluted positive spots (**Table 5.1**). Spots were found to interact with the GEF domain of BCR, and SH3, SH1, and FABD domains of ABL. In particular, 6[4H] and 7[10D] spots interacted with the FABD domain of ABL. 1[2E] and 4[11B] interacted with the SH3 domain of ABL and 11[9A] with the SH1 domain of ABL. Remaining spots interacted with the GEF domain of BCR. No confirmed positive hits were found to interact with other BCR-ABL domains present in the pools, although some were predicted to interact with the coiled-coil domain (3[12B]), Y177 (2[12H]) of BCR, or LexA (13[1C]. We did not isolate lariat ligands that interact with LexA, although other screenings performed in our laboratory have identified lariat ligands that interact with this domain. It is possible that the fraction of the library screened was not diverse enough to contain a LexA lariat ligand.

Nonetheless, we found an association between the deconvolution score of the deconvoluted spots and whether the interaction of these hits with their targets would be re-confirmed (**Figure 5.10**). Deconvoluted spots whose interactions with their target were not confirmed showed low deconvolution scores, whereas the opposite occurred for deconvoluted spots with high deconvoluted scores: All their interactions with their targets were confirmed. All interactions from deconvoluted spots with deconvolution scores higher than eight were confirmed, and only two interactions were not detected for deconvoluted spots with deconvolution scores higher than six. Overall, the deconvolution score can be used as a useful parameter to evaluate whether the interactions found in deconvoluted spots will be confirmed or not. Considering only deconvoluted spots with high deconvolution scores (such as eight) could eliminate completely the need to confirm positive interactions observed after a screening of protein ligand combinatorial libraries using the yeast-two-hybrid system.

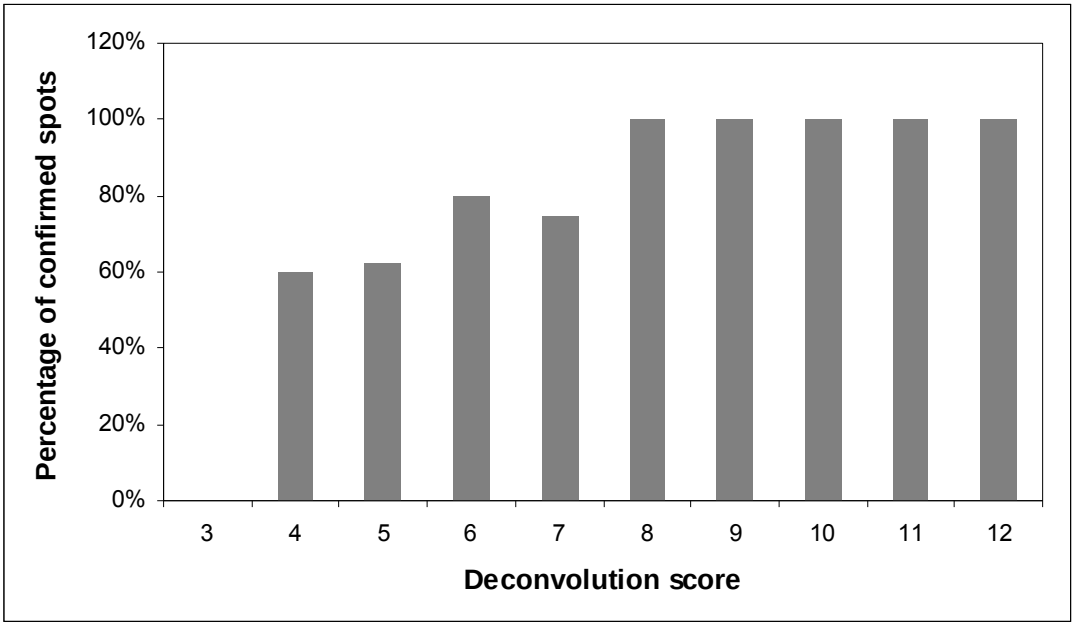


Figure 5.10. Distribution of the deconvolution score for deconvoluted spots whose interactions were subjected to reconfirmation

The deconvolution score was calculated for the deconvoluted spots from the array that was re-confirmed. For confirmed spots the interaction with their predicted bait was observed when a yeast-two-hybrid assay was repeated with retransformed prey plasmid. The deconvolution score ranges from three to eight because there must be at least three. Only deconvoluted spots were considered, spots that were confirmed but were not deconvoluted or partially deconvoluted were not included.

5.8. Sequencing of the lariat region of the confirmed deconvoluted positive hits

One advantage of performing combinatorial screenings with genetic systems rather than chemical-based systems is the ability to determine the identity of the interacting agents by amplifying their plasmids and obtaining their DNA sequence. **Figure 5.11** shows predicted amino-acid sequences of the lariat ligand coding region of reconfirmed positive hits (**Table 5.1**), since they are likely true interactors as suggested by the yeast-two-hybrid assay. All the sequences obtained corresponded to the Ssp-Ssp lariat intein scaffold (**Figure 2.3**), in which seven amino-acids were randomized for the construction of the combinatorial library. We also

show the regions adjacent to the variable region of this scaffold (on the N-terminus –AS-, followed by –CLS- on the C-terminus and the –EY- residues, which also constitutes the lariat. Most of the sequences contained seven residues in their variable region as expected (the only exception was 15[8H], which presents a stop codon in the 7th residue). Some of the isolated hits that contain the –ALS sequence instead of –CLS do form an *inactive* version of the intein, instead of the full cyclic peptide displayed by the lariat (**Figure 2.3**); however, these hits interacted with their targets in the yeast-two-hybrid assay. Altogether, the determination of the sequence of the isolated hits offers new perspectives to evaluate whether these hits have biological or inhibitory activity. It also allows the chemical synthesis of cyclic peptides with the nine amino-acid sequence, as well as mutational analysis.

a)

Spot	pre-7X	Variable region	post-7X	Predicted Target
4 [11B]	AS	PHSVFGQ	EY-CLS	SH3
11 [9A]	AS	SGGCWAQ	EY-ALS	SH1
1 [11C]	AS	MIPWCTY	EY-CLS	GEF
1 [9F]	AS	LMWWMPH	EY-CLS	GEF
10 [1A]	AS	DWWRGRR	EY-CLS	GEF
12 [7H]	AS	LRPGARR	EY-ALS	GEF
13 [1G]	AS	TLWFLHG	EY-ALS	GEF
13 [6F]	AP	KMWFFES	EY-CLS	GEF
14 [10E]	AS	VLWFFSG	EY-CLS	GEF
14 [5A]	AS	LLWFWPG	EY-CLS	GEF
14 [5G]	AS	YNRGARR	EY-CLS	GEF
15 [11F]	AS	YSRGGRR	EY-CLS	GEF
15 [8H]	AS	EGHPCD.	EY-CLS	GEF
15 [9E]	AS	EFWQFSH	EY-CLS	GEF
16 [7F]	AS	GFWWLTG	EY-CLS	GEF
16 [8A]	AS	FVCGARR	EY-ALS	GEF
2 [6F]	AS	RWFAVFN	EY-CLS	GEF
3 [2B]	AS	LVPGARR	EY-ALS	GEF
3 [8F]	AS	YMWWWSG	EY-CLS	GEF
4 [2H]	AS	WVRNSRR	EY-ALS	GEF
5 [12H]	AS	QLWFWPR	EY-CLS	GEF
6 [3E]	AS	TLWFLHG	EY-ALS	GEF
6 [5A]	AS	RLWMLTA	EY-CLS	GEF
7 [5A]	AS	RLWLFMR	VY-CLS	GEF
9 [5G]	AS	YMWWWSG	EY-CLS	GEF
6 [4H]	AS	GLSWPAE	EY-CLS	FABD
7 [10D]	AS	GWVSQKG	EY-ALS	FABD

b)

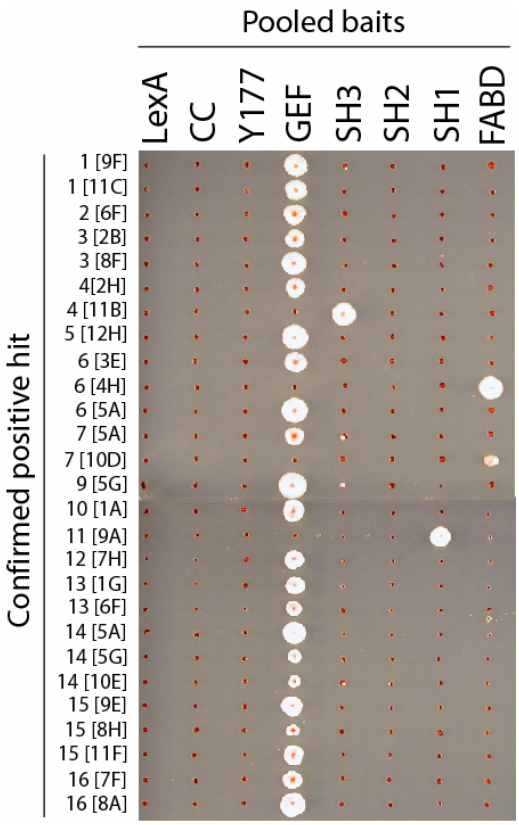


Figure 5.11. Predicted amino acid sequence of the variable region of the confirmed lariat ligands.

a) The DNA sequence of the lariat ligands from spots whose interactions were confirmed is shown, divided in three segments. The pre-7X and post-7X refers to the residues adjacent to the variable region of the lariat ligand, which is contained in the Ssp-Ssp intein scaffold (see

Figure 2.3). The variable region refers to the region of the Ssp-Ssp intein that forms the lariat cyclic peptide, along with two fixed EY residues. The variable region was randomized for all twenty standard amino acids. It is also shown the BCR-ABL domain the lariat ligands interacted with according to the yeast-two-hybrid assay. **b**) Yeast-two-hybrid assay showing the interaction of the confirmed spots that were sequenced with their target. Every spot showed an interaction with one domain of BCR-ABL exclusively. Colonies showed growth in sucrose and galactose CSM –HIS –TRP –ADE –LEU plates.

We reasoned that the isolated lariat ligands should share structural similarities that would be revealed by an analysis of their amino-acid sequence through the generation of a phylogenetic tree. Although the lariats are not evolutionary related, the generation of a phylogenetic tree provided a good way to group or cluster the most similar lariats. A tree was generated using the software MegAlign (DNAstar Inc., 1997) using default conditions. The amino-acid sequences used for the alignment were variants of the AHAS-X_7EY-CLSFG sequence, where X_7 represents the variable region of the lariat ligand. Remarkably, the sequences of the isolated lariat ligands are grouped into well defined clusters (**Figure 5.12**). Lariat ligands (4[11B], 6[4H], 7[10D] and 11[9A]) that showed interaction with a domain different from the GEF domain clustered independently from other sequences and are not similar to the rest of the sequences as well. On the other hand, the GEF lariat ligands grouped into well defined clusters, showing specific consensus sequences. One of these clusters is defined by the XXXGARR consensus (where X represents any amino-acid) derived from the five lariat ligands 12[7H], 3[2B], 16[8A], 14[5G] and 15[11F]. Another cluster of five sequences is defined by the consensus X-X_{NP}-W-W/F-X_{NP}-X-G (where X represents any amino-acid and X_{NP} represents a non-polar amino-acid), derived from the lariat ligands 13[1G]/6[3E], 16[7F], 3[8F]/9[5G], 14[5A], 5[12H]. The remaining lariat ligands did not group into a specific cluster, as is the case for 13[6F], 7[5A]/9[12E], 1[11C] and 15[8H]. Overall, the well defined clustering of some of the isolated lariat ligands suggests they are true interactors, not artifacts of the yeast-two-hybrid assay. If they were not true interactors, they would not show a defined phylogenetic pattern, since there would not be a structural consensus underlying their sequences.

Table 5.1. Confirmation of interactions from deconvoluted, partially and not deconvoluted spots.

Spot	Deconvoluted code						Deconv. score	Galactose dependence	Interaction confirmed
	1 -	1+	2 -	2+	3 -	3+			
1[2E]		1	3		1		5	+	+
1[3F]		2	3		1		5	+	+
1[9F]		1	3			1	5	+	+
1[11C]		4	4			4	12	+	+
2[5F]		1	1		1		3	-	-
2[6F]		1	2		1		4	+	+
2[12H]	1		2		1		4	-	-
3[2B]		4	3			2	9	+	+
3[8F]		2	2			1	5	+	+
3[12B]		2		2	1		5	-	-
4[2A]	1		1		1		3	-	-
4[2H]		4	1			3	7	+	+
4[11B]		3	2		3		8	+	+
5[12H]		4	4			2	10	+	+
6[2C]		1	1		1		3	-	-
6[3E]		4	3			3	10	+	+
6[4H]	3	1	3			1	ND	+	+
6[5A]		4	1			1	6	+	+
6[8F]	1		1				PD	-	-
7[5A]		3	2			3	8	+	+
7[10D]	1					2	PD	+	+
9[5G]		4	4			3	11	+	+
9[12E]		1		2		1	3	-	-
9[12H]	1			1		1	3	-	-
10[1A]		3	2			1	6	+	+
10[1B]		1	2		2		5	-	-
10[10D]		1		1	1	1	ND	-	-
11[9A]	1			3	2		6	+	+
12[3C]		1		1	1		3	-	-
12[7H]	1	1	3			1	ND	+	+
12[12H]		4	1		2		7	-	-
13[1C]		1		1		1	3	-	-
13[1G]		3	4			4	11	+	+
13[4H]		1	3			2	6	-	-
13[6F]		4	4			4	12	+	+
13[9B]		3	1			1	5	+	+
13[12B]		2	2			2	6	+	+
14[5A]		4	4			4	12	+	+
14[5G]		2	1			1	4	+	+
14[10E]		1	1			2	4	+	+
14[11C]		2	1		1		4	-	-
15[9E]		2	1			2	5	+	+
15[8H]		3	3			1	7	+	+
15[11F]		4	3			3	11	+	+
16[7F]		2	2			3	7	+	+
16[8A]		3	3			4	11	+	+

A 96-spot array was created with colonies from 46 deconvoluted, partially deconvoluted and not deconvoluted spots. Colonies were passed to CSM –HIS –TRP –ADE –LEU plates containing sucrose and galactose or glucose. True interactions should occur in plates containing galactose but not glucose. A second yeast-two-hybrid assay with lariat ligands from deconvoluted spots expected to be reproducible was performed against the eight targets pooled in the screening of a combinatorial lariat ligand library with PI-deconvolution. The variable region of the lariat ligand was re-cloned into the prey vector by homologous recombination in yeast. Colonies were plated in galactose and sucrose CSM –HIS –TRP –ADE –LEU plates. Deconvolution score is estimated only for deconvoluted spots; not deconvoluted (ND) or partially deconvoluted (PD) do not ajust to this criterium**.**

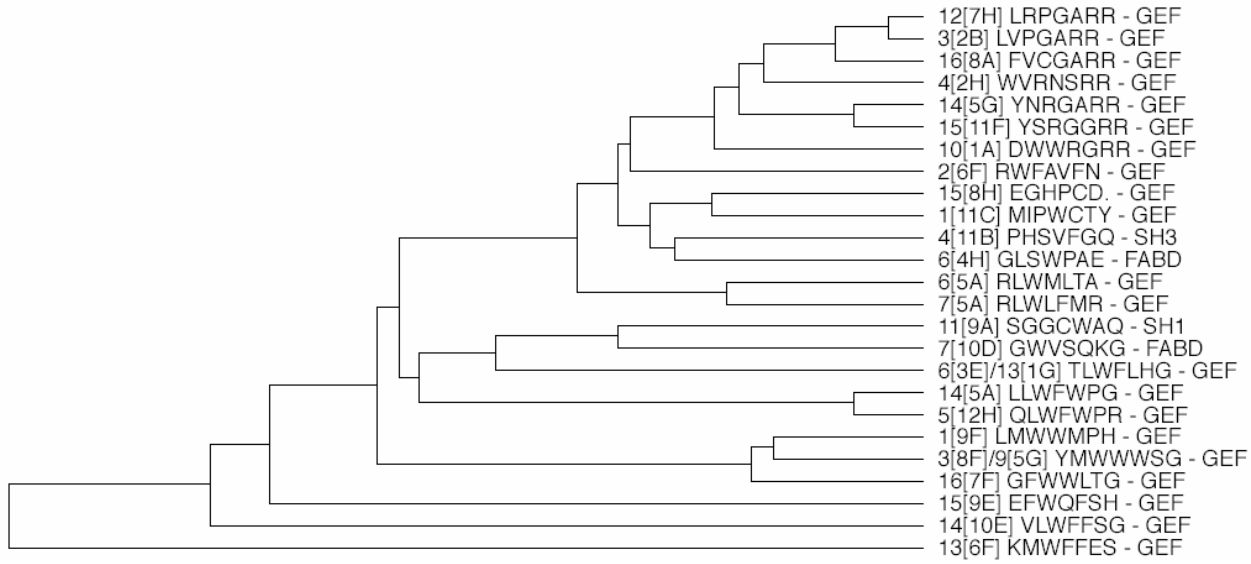


Figure 5.12. Phylogenetic tree of deconvoluted lariat ligands.

The confirmed deconvoluted hits and their confirmations were DNA sequenced, and a phylogenetic tree was generated using the region of seven amino-acids that forms the lariat motif. The sequences shown interact with the GEF domain of BCR in a yeast-two-hybrid assay, with the exception of 4[11B] that interacts with SH3, 11[9A] that interacts with SH1 and both 6[4H] and 7[10D] that interact with the FABD domains of ABL.

5.9. β-Galactosidase assay to measure the strength of the interaction of the confirmed lariat ligands

In order to measure the affinity of the different confirmed positive hits for their targets, we performed a β-galactosidase assay. The β-galactosidase assay measures the strength of the interaction between a lariat and their target. It is based on the spectrophotometric detection of

ortho-nitrophenol, a compound formed after the hydrolysis of ONPG (*ortho*-nitrophenyl-β-galactoside) by β-galactosidase. It measures the amount of β-galactosidase expressed in yeast cells, as a consequence of the interaction of the lariat with its target and thus the activation of the yeast-two-hybrid reporters. It is a semi-quantitative assay. β-galactosidase is one of the three reporters that we employed in the yeast-two-hybrid assay, encoded by the LacZ gene. A liquid β-galactosidase was performed with all the confirmed deconvoluted positive hits. Differences were detected, which provided good criteria for selecting lariat ligands for further analysis (**Figure 5.13**). For all the experiments performed, the values were normalized to the strength of interaction between the RIZ1 and the lariat ligand 6-49. Even though most of the isolated lariat ligands did not show the same strength of interaction as RIZ1 and 6-49, some positive hits presented a remarkable strength of interaction measured by the β-galactosidase assay, around 3– and 2– fold higher than the reference. This is the case of 4[11B] and 5[12H], 6[5A] and 9[5G] lariat ligands. 4[11B] interacts with the SH3 domain of BCR-ABL according to the yeast-two-hybrid assay, whereas the other isolated lariat ligands bind the GEF domain of BCR-ABL with high affinity. Twenty-three lariat ligands performed below the threshold imposed by the reference.

5.10. Screening of a combinatorial lariat ligand library with single targets

From the PI-deconvolution screening, we isolated a large number of positive hits that interacted with the GEF domain (twenty-six), in comparison with the positive hits that interacted with SH3, SH1 or FABD domains of BCR-ABL. The GEF domain of BCR might be more prone to bind lariat ligands than the other domains from BCR-ABL. In order to test this hypothesis, we screened a combinatorial lariat ligand library (based on the Ssp-Ssp Intein scaffold, the same library used for the PI-deconvolution screening) against individual targets using the yeast-two-hybrid assay, and counted how many positive hits appeared (**Figure 5.14**). Positive hits were selected in galactose and sucrose CSM –HIS –TRP –ADE –LEU + X-Gal plates. More positive hits were isolated against the GEF domain than against other domains, which supports the idea that the GEF domain is prone to interact with lariat ligands. However, the differences are not as striking as they were for the PI-deconvolution screening, for which 23

positive hits interacted with the GEF domain, and only one positive hit interacted with the SH3 or the SH1 domain.

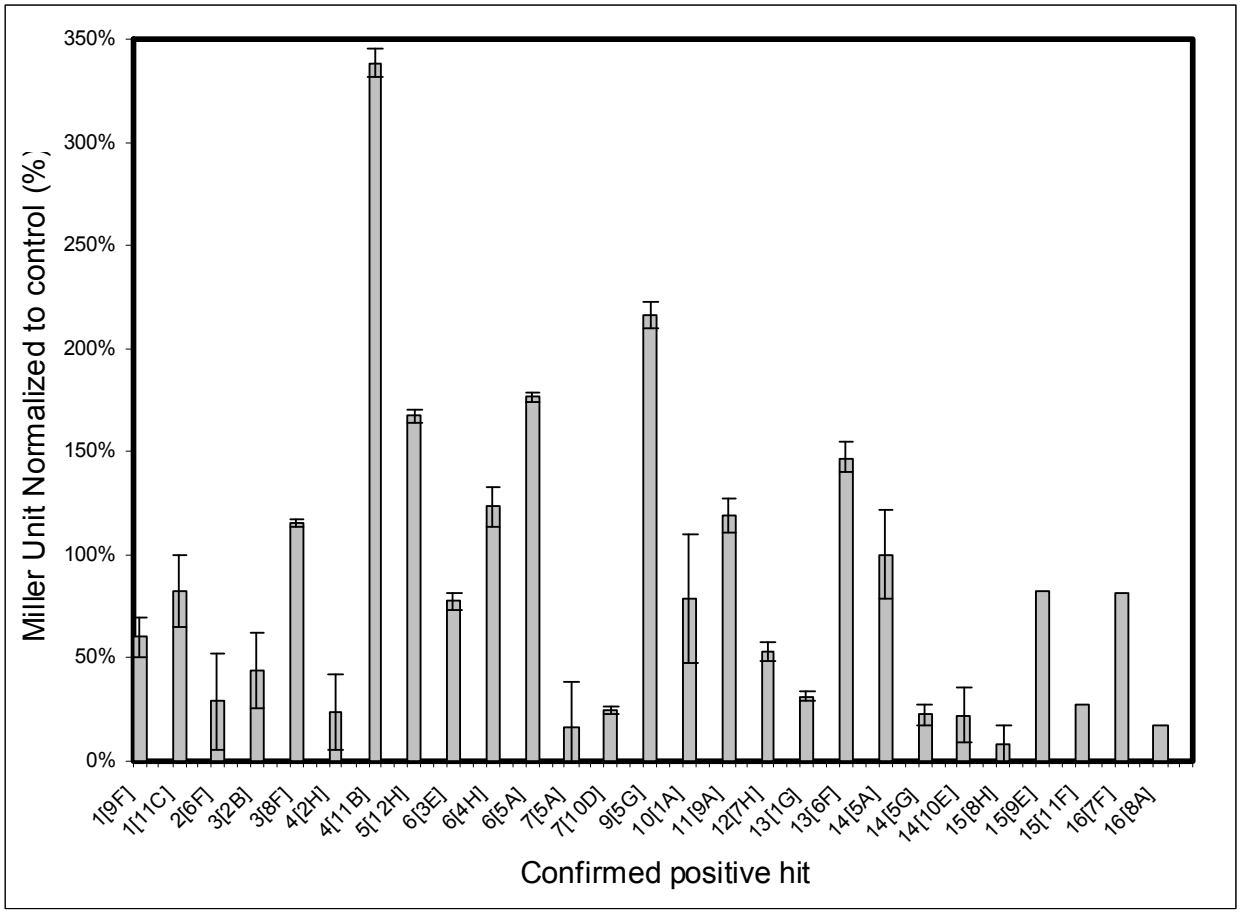


Figure 5.13. β-galactosidase assay with confirmed deconvoluted lariat ligands.
The values from the liquid β-galactosidase assay were calculated as Miller units (see methods), and normalized to a reference interaction, the interaction between RIZ1 and the 6-49 lariat ligand. The value of this interaction was considered 100%.

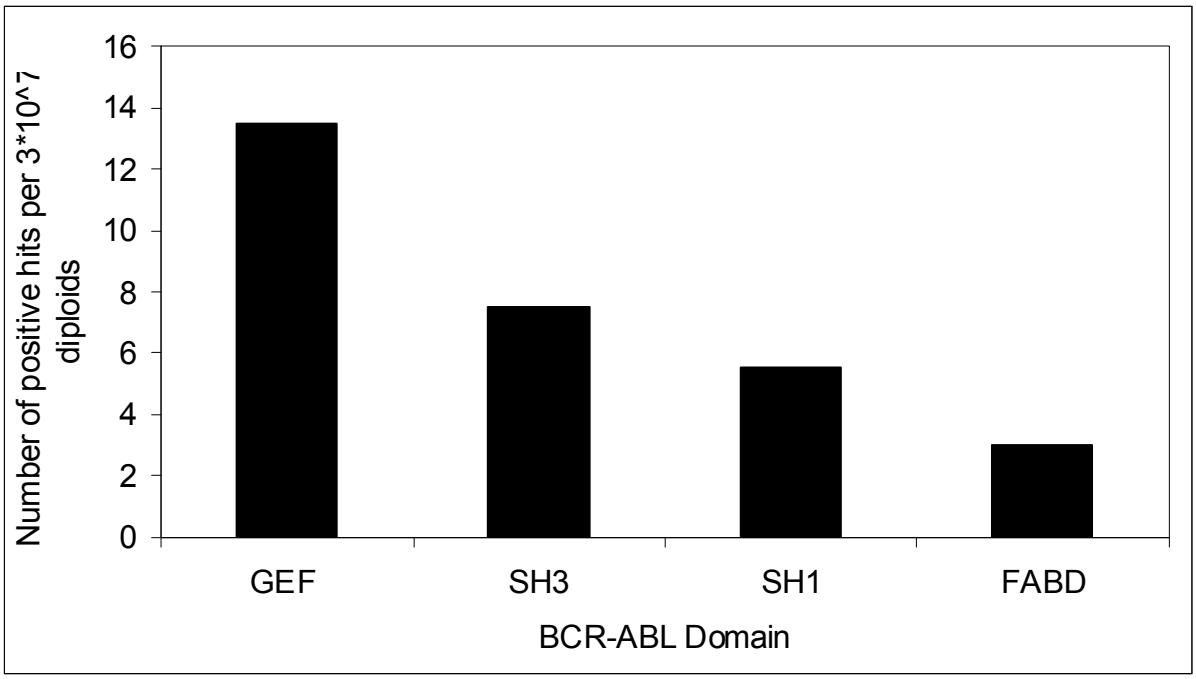


Figure 5.14. Number of positive hits isolated after a screening of a combinatorial lariat ligand library against single BCR-ABL domains.

The four domains of BCR-ABL for which positive hits were confirmed in the PI-deconvolution were screened individually against a combinatorial lariat ligand library (Ssp-Ssp intein) using the yeast-two-hybrid assay. A total of 25 million diploids were plated on galactose and sucrose CSM –HIS –TRP –ADE –LEU +X-Gal plates and the number of positive hits were determined.

6. DISCUSSION

Protein ligands that interact with a specific target protein can be developed. They can block its function or prevent it from associating with other protein partners. Protein ligands can be used in a wide array of applications, ranging from diagnostics to the study of cell signaling and as leads to identify small-molecule inhibitors. They can be isolated by artificial means from combinatorial libraries using the yeast-two-hybrid assay. However, the assay detects false positives, which imposes the need to confirm identified interactions of the target protein and the protein ligand. To do so, the plasmid expressing the protein ligand must be isolated and reinserted in a yeast prey strain, followed by a second yeast-two-hybrid assay with the bait it originally interacted. This is a time-consuming and labor-intensive task, especially when dozens of positives hits have to be reconfirmed.

Pooling samples can dramatically improve the screening of biological libraries, such as protein ligand, cDNA and genomic libraries. Pooling designs dictate how members of the library or the targets must be distributed in different pools. Instead of testing an entire library, or individual samples, pools are tested, and the goal becomes to detect a specific event in the pools, which in our case is a protein-protein interaction. Certain pooling designs can detect errors, or false positives, a very useful property when screening a combinatorial protein ligand library. We chose to adapt the *PI-deconvolution* pooling design to screening of combinatorial libraries (Jin *et al.*, 2006). According to the conditions imposed by PI-deconvolution, we aimed to screen a combinatorial protein ligand library against eight proteins, seven target domains from BCR-ABL and one negative control (LexA). BCR-ABL is an oncogene that results as a consequence of the fusion of ABL and BCR through chromosomal rearrangements. The chimera has been shown to be responsible for most cases of chronic myelogenous leukemia (Ren, 2005). The seven domains that we chose to screen are involved in the pathogenesis of BCR-ABL or play a role in the unregulated kinase activity of ABL.

Our screening of protein ligand libraries using the yeast-two-hybrid assay consisted of a combination of a matrix-based screening and the PI-deconvolution design. We determined the optimal conditions to perform a screening using a combinatorial protein ligand library. The yeast-two-hybrid assay fails to detect interactions when an interacting bait and the prey are diluted in a pool of non-interacting strains. In our case, the baits were pooled according to PI-

deconvolution, and the preys were randomly pooled. Using the interaction between RIZ1 and the protein lariat ligand 6-49 as a reference, we determined the probability of detecting an interaction when the bait and the prey are diluted serially, from 2-fold to 1000-fold. This probability was found to decrease dramatically when the interacting preys are diluted more than 100-fold. These results were consistent when RIZ1 was pooled according to PI-deconvolution. Probably, as the interacting bait and prey are diluted, the chanc es of mating with each other decrease because they can mate to a non-interacting strain, thus not forming the diploid in which the bait-prey interaction would occur. Furthermore, it is likely that the replicator tool fails to transport all the possible combinations of diploids formed in one spot, and in this way the diploids in which the interaction occurs are not taken to the next plate where the yeast-two-hybrid selection occurs.

We performed a screening of a protein ligand combinatorial library with targets pooled according to PI-deconvolution. Seven domains of BCR-ABL were cloned as LexA DBD fusions in the pEG202 plasmid, namely the Coiled-coil (CC), a peptide containing the tyrosine Y177, Guanidine exchange factor (GEF), SH3, SH2, Tyrosine kinase (SH1) and F-Actin binding domain (FABD). These seven targets, along with an additional negative control, the LexA DBD alone, were pooled according to PI-deconvolution in six pools; every pool contained four targets. Using the yeast-two-hybrid assay, all pools were screened against a combinatorial protein ligand library. The scaffold of the library is an Ssp-Ssp engineered intein that displays lariat peptides; the library contains seven randomized positions in the lariat. The library was arrayed as spots containing approximately 750 different members. Every spot was tested against each of the six pools. Approximately one million different protein ligands were screened against every pool. After performing the yeast-two-hybrid selection, only the positive interactions from pools that fully deconvoluted are expected to be true bait-prey interactions. The interactions from these positive hits were reconfirmed, showing the utility of the PI-deconvolution pooling design.

The majority of the isolated positive hits interacted with the GEF domain of BCR. For the remaining positive hits, one positive hit interacted with the SH3 domain, another with the SH1 domain, and two against the FABD domain. The sequence of the seven amino-acids that constituted these lariat ligands was obtained, and the strength of the interaction between every protein ligand and its target was measured using a spectrophotmetric β-galactosidase assay.

Finally, we developed a criterion to classify all the isolated positive hits, referred to as the *deconvolution score*. From the deconvoluted positive hits, those with low deconvolution scores are likely false positives, whereas all the positive hits that were deconvoluted and showed high deconvolution scores were reconfirmed for their detected interactions.

Altogether, we have shown that PI-deconvolution can be adapted to the screening of protein ligand combinatorial libraries. Our conditions can be adjusted to any screening, given the parameters we evaluated, such as the probability of detecting a positive hit when the interacting bait and prey are diluted. This suggests the number of targets that can be pooled (8, 16, 32, 64, etc) and the number of library members per spot, which ultimately dictates how many plates need to be constructed and what fraction of the combinatorial library can be screened. Furthermore, the positive hits whose interaction with their predicted target was reconfirmed were deconvoluted and showed high deconvolution scores. Future screenings could only focus their attention on those positive hits that show high deconvolution scores, without the need to reconfirm their interactions by tedious molecular biology methodologies. Although performing a screening using PI-deconvolution requires extensive planning and its set-up is labor-intensive, on the long term it offers more benefits than conventional approaches. Our strategy is an effective way to overcome the limitation of the appearance of false positives in the yeast-two-hybrid assays. Furthermore, we performed a screening of the Ssp-Ssp intein combinatorial library using the conventional approach, against the four BCR-ABL baits for which we isolated confirmed deconvoluted positive hits in the PI-deconvolution screening. This screening was important because it allowed a comparison of the performance of PI-deconvolution applied to the screening of protein ligand combinatorial libraries.

Using the PI-deconvolution approach, we isolated a set of lariat ligands against different domains of BCR-ABL: GEF, SH3, SH1 and FABD domains. After our single-bait screening we isolated more lariat ligands, against the same BCR-ABL domains and the SH2 domain. They are potential interactors with BCR-ABL, which might inhibit its oncogenic potential or alter the phenotypes it induces. We generated a phylogenetic tree with the isolated lariats that shows well defined clusters of sequences, according to the domain with which they interact. The sequence similarity of certain ligands suggests they must interact with the same region of a target.

The fact that more lariat ligands were isolated against the GEF domain of BCR than any other domain of BCR-ABL is surprising. In the only study that has reported the isolation of protein ligands (peptide aptamers) that interact with a GEF domain (Schmidt *et al.*, 2002), they found three different inhibitors of the GEF domain of Trio. Their sequences are not similar to the sequences of the lariat ligands that we report in this study, although the TrxA scaffold is different from the Ssp-Ssp engineered intein that displays lariat peptides. They did not mention how many positive hits they found in their study, hence there is no evidence to support that GEF domains are highly prone to interact with peptides. Another study reported the isolation nucleic acid aptamers against the GEF domain of Cytohesin 1 (Mayer *et al.*, 2001). After several rounds of selection of ligands, they reported only five sequences of RNA aptamers that bind to a GEF domain. The other known inhibitor of a GEF domain is the fungal small molecule inhibitor Brefeldin A (BFA), which binds to GEF domain of Arf proteins in the GDP-bound state (Renault *et al.*, 2003). GEF domains are flexible molecules that can cycle between a GDP-bound and a GTP-bound state. One possibility is that the lariat ligands we have isolated interact against different conformations of the GEF domain, either the GDP- or the GTP- bound states.

Relatively few lariat ligands against docking domains such as the SH3 and SH2 domains of ABL were isolated in comparison to the GEF domain. One would expect that these domains, whose main purpose is to bind other proteins, would be highly prone to interact with other molecules. For instance, in our PI-deconvolution screening of a combinatorial protein ligand library only one lariat ligand was isolated against the SH3 domain and none against the SH2. When we conducted a single-bait screening against these domains, we detected more lariat ligands that interacted with the GEF domain than the other BCR-ABL domains. This is evidence that supports that the GEF domain of BCR is more prone to interact with lariat ligands than the other BCR-ABL domains screened. The discrepancies between the lariat ligands isolated from the PI-deconvolution screening and the single-bait screening might not be related to the structure of the domains but rather the conditions under which the screenings were performed. We assumed that all the baits from every pool would be present in equal amounts, a ratio 1:1:1:1. It is possible that some baits grew faster than others and would become the majority of the baits in the pools, which would give them an advantage over the other baits for finding an interactor.

The question arises whether it is better to perform a PI-deconvolution or a single-bait screening to isolate protein ligands from a combinatorial library. It should be noted that the conditions under which we performed a PI-deconvolution were entirely manual, and this type of screening would be more useful if automated conditions were used. This way, a larger library size and more targets could be screened at one time, avoiding the need to re-confirm the observed interactions. The *deconvolution score* is a very practical parameter to validate the observed yeast-two-hybrid interactions. However, we did isolate more lariat ligands in the single-bait screening, in comparison to the PI-deconvolution screening. If only a limited number of targets will be screened, it might be better to perform a single-bait screening. If a considerable number of targets will be screened, more than eight, and there are robotic equipment in the laboratory, PI-deconvolution would offer excellent benefits.

We performed a PI-deconvolution screening of a protein ligand combinatorial library under one set of conditions, namely pooling eight baits, one million library members and an Ssp-Ssp engineering intein lariat ligand library. It would be interesting to evaluate the performance of PI-deconvolution screening using more complex conditions, like pooling more targets, or using a different scaffold for the creation of the combinatorial library. Data from more screenings will support the usefulness of the *deconvolution score* parameter, to determine if it can determine with confidence whether a positive hit will be reconfirmed or not. Altogether, we established the appropriate conditions to screen a protein ligand combinatorial library using the PI-deconvolution pooling design with seven domains of BCR-ABL, and established parameters to evaluate future screenings. Our findings should facilitate the isolation of protein ligands that interact with other targets of interest. Further studies may focus on expanding the number of targets that can be screened at the time, with the use of robotic equipment. The ideal scenario would be to isolate a protein ligand that alters the function of every protein of the genome of an organism, in order to evaluate the function of every protein. Additionally, the approach we employed can be used to adapt PI-deconvolution to the screening of combinatorial libraries with other methodologies, such as phage-display.

We screened a protein ligand combinatorial against a variety of targets, ranging from a short peptide stretch (Y177) to a tyrosine kinase domain to docking domains (SH3 and SH2). We determined the amino-acid sequence of a variety of lariat ligands that interact with domains of BCR-ABL using the yeast-two-hybrid assay. It will be very interesting to determine whether

there lariat ligands interact with BCR-ABL with another technique that detects protein-protein interactions and also whether they can inhibit ABL kinase activity or inhibit its interaction with docking domains. Coimmunoprecipitations and surface plasmon resonance could be used to determine if the isolated lariat ligands do interact with their target, among other methodologies available. It will be also of interest to determine whether these lariat ligands can have an effect on downstream signalling pathways. For the GEF domain, it will be interesting to determine if it can alter the motility of BCR-ABL$^+$ cell lines. Also will be interesting to determine if these lariat ligands can inhibit the proliferation of BCR-ABL$^+$ CML cell lines; in such a case, they could lead the discovery of small-molecule inhibitors of BCR-ABL, which along with Imatinib would constitute a very strong and effective therapy against CML.

7. REFERENCES

Abed, N., Bickle, M., Mari, B., Schapira, M., Sanjuan-Espana, R., Robbe Sermesant, K., Moncorge, O., Mouradian-Garcia, S., Barbry, P., Rudkin, B.B. *et al*. (2007). A comparative analysis of perturbations caused by a gene knock-out, a dominant negative allele, and a set of peptide aptamers. Mol. Cell. Proteomics *12*, 2110-2121.

Ausubel, F.M., and Brent, R., Kingston R.E. (1988). Current Protocols in Molecular Biology. Ed. Wiley. New York.

Baines, I.C., and Colas, P. (2006). Peptide aptamers as guides for small-molecule drug discovery. Drug Discov. Today *7-8,* 334-341.

Balding, D., Bruno, W., Knill, E., and Torney, D. (1996). A comparative survey of non-adaptive pooling designs. In Genetic mapping and DNA sequencing, Speed, T. and Waterman, M. eds., (New York: Springer) pp. 133-154.

Balding, D.J., and Torney, D.C. (1996). Optimal Pooling Designs with Error Detection. Journal of Combinatorial Theory, Series A *1,* 131-140.

Barila, D., and Superti-Furga, G. (1998). An intramolecular SH3-domain interaction regulates c-Abl activity. Nat. Genet. *3,* 280-282.

Barnes, D.J., and Melo, J.V. (2002). Cytogenetic and molecular genetic aspects of chronic myeloid leukaemia. Acta Haematol. *4,* 180-202.

Beste, G., Schmidt, F.S., Stibora, T., and Skerra, A. (1999). Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. Proc. Natl. Acad. Sci. U. S. A. *5,* 1898-1903.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. Nat. Biotechnol. *5,* 575-582.

Binz, H.K., Amstutz, P., and Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. *10,* 1257-1268.

Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S.M., Lee, T., Pope, S.H., Riordan, G.S., and Whitlow, M. (1988). Single-chain antigen-binding proteins. Science *4877,* 423-426.

Brinkmann, U., Reiter, Y., Jung, S.H., Lee, B., and Pastan, I. (1993). A recombinant immunotoxin containing a disulfide-stabilized Fv fragment. Proc. Natl. Acad. Sci. U. S. A. *16,* 7538-7542.

Bruno, W.J., Knill, E., Balding, D.J., Bruce, D.C., Doggett, N.A., Sawhill, W.W., Stallings, R.L., Whittaker, C.C., and Torney, D.C. (1995). Efficient pooling designs for library screening. Genomics *1,* 21-30.

Butz, K., Denk, C., Fitscher, B., Crnkovic-Mertens, I., Ullmann, A., Schroder, C.H., and Hoppe-Seyler, F. (2001). Peptide aptamers targeting the hepatitis B virus core protein: a new class of molecules with antiviral activity. Oncogene *45,* 6579-6586.

Capdeville, R., Buchdunger, E., Zimmermann, J., and Matter, A. (2002). Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. Nat. Rev. Drug Discov. *7,* 493-502.

Causier, B. (2004). Studying the interactome with the yeast two-hybrid system and mass spectrometry. Mass Spectrom. Rev. *23*, 350-367.

Chattopadhyay, A., Tate, S.A., Beswick, R.W., Wagner, S.D., and Ko Ferrigno, P. (2006). A peptide aptamer to antagonize BCL-6 function. Oncogene *15,* 2223-2233.

Cheng, L., Naumann, T.A., Horswill, A.R., Hong, S.J., Venters, B.J., Tomsho, J.W., Benkovic, S.J., and Keiler, K.C. (2007). Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease. Protein Sci. *8,* 1535-1542.

Colas, P. (2000). Combinatorial protein reagents to manipulate protein function. Curr. Opin. Chem Biol. *4*, 54-59.

Colas, P. (2008). The eleven-year switch of peptide aptamers. J. Biol. *1,* 2.

Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J., and Brent, R. (1996). Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. Nature *6574,* 548-550.

Colas, P., Cohen, B., Ko Ferrigno, P., Silver, P.A., and Brent, R. (2000). Targeted modification and transportation of cellular proteins. Proc. Natl. Acad. Sci. U. S. A. *25,* 13720-13725.

Craik, D.J. (2006). Chemistry. Seamless proteins tie up their loose ends. Science *5767,* 1563-1564.

Dai, Z., and Pendergast, A.M. (1995). Abi-2, a novel SH3-containing protein interacts with the c-Abl tyrosine kinase and modulates c-Abl transforming activity. Genes Dev. *21,* 2569-2582.

Daubon, T., Chasseriau, J., Ali, A.E., Rivet, J., Kitzis, A., Constantin, B., and Bourmeyster, N. (2008). Differential motility of p190bcr-abl- and p210bcr-abl-expressing cells: respective roles of Vav and Bcr-Abl GEFs. Oncogene *19,* 2673-2685.

Davis, J.J., Tkac, J., Laurenson, S., and Ko Ferrigno, P. (2007). Peptide aptamers in label-free protein detection: 1. Characterization of the immobilized scaffold. Anal. Chem. *3,* 1089-1096.

de Chassey, B., Mikaelian, I., Mathieu, A.L., Bickle, M., Olivier, D., Negre, D., Cosset, F.L., Rudkin, B.B., and Colas, P. (2007). An antiproliferative genetic screening identifies a peptide aptamer that targets calcineurin and up-regulates its activity. Mol. Cell. Proteomics *3,* 451-459.

Deininger, M.W., Goldman, J.M., and Melo, J.V. (2000). The molecular biology of chronic myeloid leukemia. Blood *10,* 3343-3356.

Dennis, M.S., and Lazarus, R.A. (1994). Kunitz domain inhibitors of tissue factor-factor VIIa. I. Potent inhibitors selected from libraries by phage display. J. Biol. Chem. *35,* 22129-22136.

Deshayes, K., Schaffer, M.L., Skelton, N.J., Nakamura, G.R., Kadkhodayan, S., and Sidhu, S.S. (2002). Rapid identification of small binding motifs with high-throughput phage display: discovery of peptidic antagonists of IGF-1 function. Chem. Biol. *4,* 495-505.

Du, D.Z., Hwang, F.K., Wu, W., and Znati, T. (2006). New construction for transversal design. J. Comput. Biol. *4,* 990-995.

Feller, S.M., Knudsen, B., and Hanafusa, H. (1994). c-Abl kinase regulates the protein binding activity of c-Crk. EMBO J. *10,* 2341-2351.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *6230,* 245-246.

Geyer, C.R., and Brent, R. (2000). Selection of genetic agents from random peptide aptamer expression libraries. Methods Enzymol. 171-208.

Geyer, C.R., Colman-Lerner, A., and Brent, R. (1999). "Mutagenesis" by peptide aptamers identifies genetic network members and pathway connections. Proc. Natl. Acad. Sci. U. S. A. *15,* 8567-8572.

Giebel, L.B., Cass, R.T., Milligan, D.L., Young, D.C., Arze, R., and Johnson, C.R. (1995). Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. Biochemistry *47,* 15430-15435.

Gietz, R.D., and Schiestl, R.H. (2007). Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *1,* 1-4.

Gietz, R.D., and Schiestl, R.H. (2007). Microtiter plate transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *1,* 5-8.

Glockshuber, R., Malia, M., Pfitzinger, I., and Pluckthun, A. (1990). A comparison of strategies to stabilize immunoglobulin Fv-fragments. Biochemistry *6,* 1362-1367.

Goga, A., McLaughlin, J., Afar, D.E., Saffran, D.C., and Witte, O.N. (1995). Alternative signals to RAS for hematopoietic transformation by the BCR-ABL oncogene. Cell *6,* 981-988.

Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002). Inteins: structure, function, and evolution. Annu. Rev. Microbiol. 263-287.

Goldman, J.M., and Melo, J.V. (2003). Chronic myeloid leukemia--advances in biology and new approaches to treatment. N. Engl. J. Med. *15,* 1451-1464.

Golemis, E.A., Serebriiskii, I., Finley, R.L.,Jr, Kolonin, M.G., Gyuris, J., and Brent, R. (2001). Interaction trap/two-hybrid system to identify interacting proteins. Curr. Protoc. Protein Sci. Unit19.2.

Goss, V.L., Lee, K.A., Moritz, A., Nardone, J., Spek, E.J., MacNeill, J., Rush, J., Comb, M.J., and Polakiewicz, R.D. (2006). A common phosphotyrosine signature for the Bcr-Abl kinase. Blood *12,* 4888-4897.

Gosser, Y.Q., Zheng, J., Overduin, M., Mayer, B.J., and Cowburn, D. (1995). The solution structure of Abl SH3, and its relationship to SH2 in the SH(32) construct. Structure *10,* 1075-1086.

Haab, B.B., Dunham, M.J., and Brown, P.O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol. *2,* RESEARCH0004.

Hantschel, O., Nagar, B., Guettler, S., Kretzschmar, J., Dorey, K., Kuriyan, J., and Superti-Furga, G. (2003). A myristoyl/phosphotyrosine switch regulates c-Abl. Cell *6,* 845-857.

Hantschel, O., and Superti-Furga, G. (2004). Regulation of the c-Abl and Bcr-Abl tyrosine kinases. Nat. Rev. Mol. Cell Biol. *1,* 33-44.

Hantschel, O., Wiesner, S., Guttler, T., Mackereth, C.D., Rix, L.L., Mikes, Z., Dehne, J., Gorlich, D., Sattler, M., and Superti-Furga, G. (2005). Structural basis for the cytoskeletal association of Bcr-Abl/c-Abl. Mol. Cell *4,* 461-473.

Harnois, T., Constantin, B., Rioux, A., Grenioux, E., Kitzis, A., and Bourmeyster, N. (2003). Differential interaction and activation of Rho family GTPases by p210bcr-abl and p190bcr-abl. Oncogene *41,* 6445-6454.

Harrison, J.L., Williams, S.C., Winter, G., and Nissim, A. (1996). Screening of phage antibody libraries. Methods Enzymol. 83-109.

Hoppe-Seyler, F., Crnkovic-Mertens, I., Tomai, E., and Butz, K. (2004). Peptide aptamers: specific inhibitors of protein function. Curr. Mol. Med. *5,* 529-538.

Horswill, A.R., and Benkovic, S.J. (2005). Cyclic peptides, a chemical genetics tool for biologists. Cell. Cycle *4,* 552-555.

Hosse, R.J., Rothe, A., and Power, B.E. (2006). A new generation of protein display scaffolds for molecular recognition. Protein Sci. *1,* 14-27.

Howard, P.L., Chia, M.C., Del Rizzo, S., Liu, F.F., and Pawson, T. (2003). Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. Proc. Natl. Acad. Sci. U. S. A. *20,* 11267-11272.

Hua, S.B., Qiu, M., Chan, E., Zhu, L., and Luo, Y. (1997). Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. Plasmid *2,* 91-96.

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y. (2000). Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc. Natl. Acad. Sci. U. S. A. *3,* 1143-1147.

Jin, F., Avramova, L., Huang, J., and Hazbun, T. (2007). A yeast two-hybrid smart-pool-array system for protein-interaction mapping. Nat. Methods *5,* 405-407.

Jin, F., Hazbun, T., Michaud, G.A., Salcius, M., Predki, P.F., Fields, S., and Huang, J. (2006). A pooling-deconvolution strategy for biological network elucidation. Nat. Methods *3,* 183-189.

Johnson, S., Evans, D., Laurenson, S., Paul, D., Davies, A.G., Ferrigno, P.K., and Walti, C. (2008). Surface-immobilized peptide aptamers as probe molecules for protein detection. Anal. Chem. *4,* 978-983.

Joshi, P.B., Hirst, M., Malcolm, T., Parent, J., Mitchell, D., Lund, K., and Sadowski, I. (2007). Identification of protein interaction antagonists using the repressed transactivator two-hybrid system. BioTechniques *5,* 635-644.

Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? Nat. Rev. Drug Discov. *8,* 711-715.

Kushnirov, V.V. (2000). Rapid and reliable protein extraction from yeast. Yeast *9,* 857-860.

Ladner, R.C. (1995). Constrained peptides as binding entities. Trends Biotechnol. *10,* 426-430.

Lobato, M.N., and Rabbitts, T.H. (2004). Intracellular antibodies as specific reagents for functional ablation: future therapeutic molecules. Curr. Mol. Med. *5,* 519-528.

Lu, Z., Murray, K.S., Van Cleave, V., LaVallie, E.R., Stahl, M.L., and McCoy, J.M. (1995). Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. Biotechnology (N. Y) *4,* 366-372.

Mayer, G., Blind, M., Nagel, W., Bohm, T., Knorr, T., Jackson, C.L., Kolanus, W., and Famulok, M. (2001). Controlling small guanine-nucleotide-exchange factor function through cytoplasmic RNA intramers. Proc. Natl. Acad. Sci. U. S. A. *9,* 4961-4965.

Meyer, S.C., Gaj, T., and Ghosh, I. (2006). Highly selective cyclic peptide ligands for NeutrAvidin and avidin identified by phage display. Chem. Biol. Drug Des. *1,* 3-10.

Miao, Y.J., and Wang, J.Y. (1996). Binding of A/T-rich DNA by three high mobility group-like domains in c-Abl tyrosine kinase. J. Biol. Chem. *37,* 22823-22830.

Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., and Kuriyan, J. (2003). Structural basis for the autoinhibition of c-Abl tyrosine kinase. Cell *6,* 859-871.

Naumann, T.A., Tavassoli, A., and Benkovic, S.J. (2008). Genetic selection of cyclic peptide Dam methyltransferase inhibitors. Chembiochem *2,* 194-197.

Ngo, H.Q., and Du, D.Z. (2000). A survey on combinatorial group testing algorithms with applications to DNA Library Screening. Du, D. Z., Pardalos, P. M. and Wang, J. eds., (Detroit: DIMACS Series in Discrete Mathematics and Theoretical Computer Science) pp. 171-182.

Nishida, K., Kaziro, Y., and Satoh, T. (1999). Anti-apoptotic function of Rac in hematopoietic cells. Oncogene *2,* 407-415.

Nord, K., Nilsson, J., Nilsson, B., Uhlen, M., and Nygren, P.A. (1995). A combinatorial library of an alpha-helical bacterial receptor domain. Protein Eng. *6,* 601-608.

Norman, T.C., Smith, D.L., Sorger, P.K., Drees, B.L., O'Rourke, S.M., Hughes, T.R., Roberts, C.J., Friend, S.H., Fields, S., and Murray, A.W. (1999). Genetic selection of peptide inhibitors of biological pathways. Science *5427,* 591-595.

Nouvion, A.L., Thibaut, J., Lohez, O.D., Venet, S., Colas, P., Gillet, G., and Lalle, P. (2007). Modulation of Nr-13 antideath activity by peptide aptamers. Oncogene *5,* 701-710.

Nygren, P.A., and Skerra, A. (2004). Binding proteins from alternative scaffolds. J. Immunol. Methods *1-2,* 3-28.

Oda, T., Heaney, C., Hagopian, J.R., Okuda, K., Griffin, J.D., and Druker, B.J. (1994). Crkl is the major tyrosine-phosphorylated protein in neutrophils from patients with chronic myelogenous leukemia. J. Biol. Chem. *37,* 22925-22928.

Olson, M.F., Pasteris, N.G., Gorski, J.L., and Hall, A. (1996). Faciogenital dysplasia protein (FGD1) and Vav, two related proteins required for normal embryonic development, are upstream regulators of Rho GTPases. Curr. Biol. *12,* 1628-1633.

Overduin, M., Rios, C.B., Mayer, B.J., Baltimore, D., and Cowburn, D. (1992). Three-dimensional solution structure of the src homology 2 domain of c-abl. Cell *4,* 697-704.

Padlan, E.A. (1994). Anatomy of the antibody molecule. Mol. Immunol. *3,* 169-217.

Pawson, T., and Warner, N. (2007). Oncogenic re-wiring of cellular signaling pathways. Oncogene *9,* 1268-1275.

Pelicci, G., Lanfrancone, L., Salcini, A.E., Romano, A., Mele, S., Grazia Borrello, M., Segatto, O., Di Fiore, P.P., and Pelicci, P.G. (1995). Constitutive phosphorylation of Shc proteins in human tumors. Oncogene *5,* 899-907.

Pendergast, A.M., Quilliam, L.A., Cripe, L.D., Bassing, C.H., Dai, Z., Li, N., Batzer, A., Rabun, K.M., Der, C.J., and Schlessinger, J. (1993). BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. Cell *1,* 175-185.

Pluthero, F.G. (1993). Rapid purification of high-activity Taq DNA polymerase. Nucleic Acids Res. *20,* 4850-4851.

Puil, L., Liu, J., Gish, G., Mbamalu, G., Bowtell, D., Pelicci, P.G., Arlinghaus, R., and Pawson, T. (1994). Bcr-Abl oncoproteins bind directly to activators of the Ras signalling pathway. EMBO J. *4,* 764-773.

Ren, R. (2005). Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. Nat. Rev. Cancer. *3,* 172-183.

Renault, L., Guibert, B., and Cherfils, J. (2003). Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. Nature *6966,* 525-530.

Roberts, B.L., Markland, W., Ley, A.C., Kent, R.B., White, D.W., Guterman, S.K., and Ladner, R.C. (1992). Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. Proc. Natl. Acad. Sci. U. S. A. *6,* 2429-2433.

Sattler, M., Mohi, M.G., Pride, Y.B., Quinnan, L.R., Malouf, N.A., Podar, K., Gesbert, F., Iwasaki, H., Li, S., Van Etten, R.A. *et al*. (2002). Critical role for Gab2 in transformation by BCR/ABL. Cancer. Cell. *5,* 479-492.

Sattler, M., Salgia, R., Okuda, K., Uemura, N., Durstin, M.A., Pisick, E., Xu, G., Li, J.L., Prasad, K.V., and Griffin, J.D. (1996). The proto-oncogene product p120CBL and the adaptor proteins CRKL and c-CRK link c-ABL, p190BCR/ABL and p210BCR/ABL to the phosphatidylinositol-3' kinase pathway. Oncogene *4,* 839-846.

Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., and Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science *5486,* 1938-1942.

Schmidt, S., Diriong, S., Mery, J., Fabbrizio, E., and Debant, A. (2002). Identification of the first Rho-GEF inhibitor, TRIPalpha, which targets the RhoA-specific GEF domain of Trio. FEBS Lett. *1-3,* 35-42.

Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S.A., Baechi, T., Moelling, K., and Hovens, C.M. (1999). Mutagenesis and selection of PDZ domains that bind new protein targets. Nat. Biotechnol. *2,* 170-175.

Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C., and Benkovic, S.J. (1999). Production of cyclic peptides and proteins in vivo. Proc. Natl. Acad. Sci. U. S. A. *24,* 13638-13643.

Sham, P., Bader, J.S., Craig, I., O'Donovan, M., and Owen, M. (2002). DNA Pooling: a tool for large-scale association studies. Nat. Rev. Genet. *11,* 862-871.

Sidhu, S.S., Lowman, H.B., Cunningham, B.C., and Wells, J.A. (2000). Phage display for selection of novel binding peptides. Methods Enzymol. 333-363.

Skerra, A., and Pluckthun, A. (1988). Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. Science *4855,* 1038-1041.

Smith, G.P., Patel, S.U., Windass, J.D., Thornton, J.M., Winter, G., and Griffiths, A.D. (1998). Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. J. Mol. Biol. *2,* 317-332.

Smith, J.M., Katz, S., and Mayer, B.J. (1999). Activation of the Abl tyrosine kinase in vivo by Src homology 3 domains from the Src homology 2/Src homology 3 adaptor Nck. J. Biol. Chem. *39,* 27956-27962.

Smith, K.M., Yacobi, R., and Van Etten, R.A. (2003). Autoinhibition of Bcr-Abl through its SH3 domain. Mol. Cell *1,* 27-37.

Taagepera, S., McDonald, D., Loeb, J.E., Whitaker, L.L., McElroy, A.K., Wang, J.Y., and Hope, T.J. (1998). Nuclear-cytoplasmic shuttling of C-ABL tyrosine kinase. Proc. Natl. Acad. Sci. U. S. A. *13,* 7457-7462.

Takeda, N., Shibuya, M., and Maru, Y. (1999). The BCR-ABL oncoprotein potentially interacts with the xeroderma pigmentosum group B protein. Proc. Natl. Acad. Sci. U. S. A. *1,* 203-207.

Tapon, N., and Hall, A. (1997). Rho, Rac and Cdc42 GTPases regulate the organization of the actin cytoskeleton. Curr. Opin. Cell Biol. *1,* 86-92.

Tauchi, T., Miyazawa, K., Feng, G.S., Broxmeyer, H.E., and Toyama, K. (1997). A coiled-coil tetramerization domain of BCR-ABL is essential for the interactions of SH2-containing signal transduction molecules. J. Biol. Chem. *2,* 1389-1394.

Thierry-Mieg, N. (2006). A new pooling strategy for high-throughput screening: the Shifted Transversal Design. BMC Bioinformatics. 28.

Thierry-Mieg, N., and Bailly, G. (2008). Interpool: interpreting smart-pooling results. Bioinformatics *5,* 696-703.

Uhlen, M. (2008). Affinity as a tool in life science. BioTechniques *5,* 649-654.

Underhill-Day, N., Pierce, A., Thompson, S.E., Xenaki, D., Whetton, A.D., and Owen-Lynch, P.J. (2006). Role of the C-terminal actin binding domain in BCR/ABL-mediated survival and drug resistance. Br. J. Haematol. *6,* 774-783.

Van Etten, R.A. (1999). Cycling, stressed-out and nervous: cellular functions of c-Abl. Trends Cell Biol. *5,* 179-186.

Vermeirssen, V., Deplancke, B., Barrasa, M.I., Reece-Hoyes, J.S., Arda, H.E., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stamm, L., Brent, M.R., and Walhout, A.J. (2007). Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. Nat. Methods *8,* 659-664.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *6887,* 399-403.

Wong, S., and Witte, O.N. (2004). The BCR-ABL story: bench to bedside and back. Annu. Rev. Immunol. 247-306.

Woodman, R., Yeh, J.T., Laurenson, S., and Ko Ferrigno, P. (2005). Design and validation of a neutral protein scaffold for the presentation of peptide aptamers. J. Mol. Biol. *5,* 1118-1133.

Worn, A., and Pluckthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. J. Mol. Biol. *5,* 989-1010.

Xu, L., Aha, P., Gu, K., Kuimelis, R.G., Kurz, M., Lam, T., Lim, A.C., Liu, H., Lohse, P.A., Sun, L. *et al*. (2002). Directed evolution of high-affinity antibody mimics using mRNA display. Chem. Biol. *8,* 933-942.

Young, N.M., MacKenzie, C.R., Narang, S.A., Oomen, R.P., and Baenziger, J.E. (1995). Thermal stabilization of a single-chain Fv antibody fragment by introduction of a disulphide bond. FEBS Lett. *2*, 135-139.

Yuan, Z.M., Huang, Y., Ishiko, T., Kharbanda, S., Weichselbaum, R., and Kufe, D. (1997). Regulation of DNA damage-induced apoptosis by the c-Abl tyrosine kinase. Proc. Natl. Acad. Sci. U. S. A. *4*, 1437-1440.

Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V.N., and Kim, P.S. (2002). Structure of the Bcr-Abl oncoprotein oligomerization domain. Nat. Struct. Biol. *2*, 117-120.