Structural Characterisation of Human Kinases using a Library-Based Construct Screening Approach

Hayretin Yumerefendi

Laboratoire Européen de Biologie Moléculaire

Programme Doctorale Internationale

conjoint avec

Université Joseph Fourier – Grenoble I

Ecole Doctorale Chimie et Sciences du Vivant

Thése

Caractérisation Structurale des Kinase Humaines par le Criblage de Bibliothèque des Constructions

Hayretin YUMEREFENDI

Dr. Darren Hart, EMBL-Grenoble	Directeur de thése
Prof. Stefan Knapp, Université de Oxford	Rapporteur
Dr. Michael Dyson, Univesité de Cambridge	Rapporteur
Prof. Marc Jamin, UVHCI	Examinateur
Dr. Imre Berger, EMBL-Grenoble	Examinateur
Prof. Philippe Minard, CNRS	Examinateur

European Molecular Biology Laboratory

International Ph.D. Programme

joint degree with

University Joseph Fourier – Grenoble I

Faculty of Chemistry and Life Sciences

Thesis

Structural Characterisation of Human Kinases using a Library-Based Construct Screening Approach

Hayretin Yumerefendi

Dr. Darren Hart, EMBL-Grenoble	Thesis Director
Prof. Stefan Knapp, University of Oxford	Reviewer
Dr. Michael Dyson, Univesity of Cambridge	Reviewer
Prof. Marc Jamin, UVHCI	Examiner
Dr. Imre Berger, EMBL-Grenoble	Examiner
Prof. Philippe Minard, CNRS	Examiner

Soumise pour l'obtention du diplôme de

DOCTEUR de l'EMBL et l'Université Joseph Fourier – Grenoble I

Spécialité: Biologie Structurale et Nanobiologie

Submitted in partial fulfillment of the requirements for the degree

Doctor of Philosophy of the EMBL and University Joseph Fourier – Grenoble I

Discipline: Structural and Nanobiology

Résumé

Le manque de protéine soluble est souvent un obstacle à la caractérisation structurale des protéines. Une approche courante pour surmonter cela consiste à isoler des domaines protéiques en utilisant des méthodes itératives de création et analyse de constructions. Autrement des banques d'ADN, contenant toutes les constructions possibles d'une même protéine, peuvent être crée par troncation enzymatique et testée à la fois pour l'expression et la solubilité de celle-ci. Dans ce projet, la nouvelle méthode d'Expression des Protéines Solubles par Troncation Aléatoire Incrémentielle (ESPRIT) a été utilisée pour explorer la définition des domaines protéiques. Des protéines kinases multidomaines qui avaient échoué surexpression solubles ont été choisies pour leur intérêt biologique. La méthode a d'abord été optimisée pour améliorer la qualité et l'efficacité des banques permettant un meilleur traitement de la diversité générée. Elle a ensuite été appliquée à une protéine kinase modèle DAPK1, à partir de laquelle une définition précise de domaine a été démontrée. Le criblage des constructions a ainsi permis l'identification de protéines plus stables, et cristallisation des constructions dans les conformations alternatives. La méthode a aussi été appliquée pour identifier des variants solubles du complexe DAPK1 et son partenaire la calmoduline, permettant de mettre en évidence un domaine d'interaction minimal, plus petit que celui décrit auparavant. Alors que plusieurs tentatives pour obtenir le domaine catalytique kinase de IKKβ avaient échoué, des criblages additionnels sur cette protéine avec cette méthode ont permis d'identifier des domaines de régulation solubles et fonctionnels *in vivo*. Enfin, il a été démontré que des constructions du domaine catalytique de p110β, faiblement exprimées dans *E.coli*, pouvaient être exprimées solubles dans des cellules d'insectes avec des rendements plus importants.

Abstract

Structural characterization of proteins is often hindered by insufficient amounts of soluble protein. A common approach is to isolate its separate domains, classically done by time-consuming iterations of design, generation and testing of constructs. An alternative approach is to generate a random library of all possible constructs by enzymatic DNA truncation and test them in one experiment for expression and solubility. In this work, the novel, directed evolution-type method Expression of Soluble Proteins by Random Incremental Truncations (ESPRIT) was used to explore the definition of protein domains. The biological focus was a set of multidomain protein kinases that has previously resisted soluble over-expression and structural characterisation. First, improvements in the method were developed to improve the quality and efficiency of the truncation libraries leading to better coverage of the diversity. Then, its application on a more characterised protein, DAPK1, demonstrated precise domain definition. Small variations of the constructs resulted in significantly different thermal stability and crystal packing, thus providing a means to resolve structures with alternative conformations. The method was also used for the identification of soluble variants of the complex between DAPK1 and its partner calmodulin where the minimal interaction region was shorter than previously reported. The application of the method to a series of difficult-to-express protein kinases resulted in rapid identification of incompatibility with *E. coli* over-expression for some of the targets. While attempts to rescue the low solubility the IKKβ kinase catalytic domain were unsuccessful, additional screens on this protein identified soluble and regulatory domains that showed some functionality *in vivo*. Finally, it was demonstrated that constructs of the PI3Kβ p110β catalytic domain with residual solubility in *E. coli*, identified from screening, could be expressed in insect cells at higher yields in soluble form.

Remerciements

Tout d'abord, je voudrais exprimer ma gratitude à ma famille pour leur soutien. Un point central dans mon voyage avait Prof Ryan Richards grâce à qui j'ai découvert la recherche. Je suis profondément reconnaissant à mon directeur de thèse Dr Darren Hart, qui m'a fait l'honneur d'être son premier étudiant en thèse. Il était un maître enthousiaste avec qui j'ai eu de nombreuses discussions vives. Je tiens également à remercier tous les membres de notre laboratoire merveilleux qui m'a aidé à garder une bonne motivation et m'a rappelé les aspects positifs dans les jours sombres. Un grand merci à Natasa Djordjevic pour être une stagiaire très aimable et joyeuse; Bridgette Connell et Sergiy Avilov pour leurs commentaires utiles à la relecture de ce manuscrit.

Ce travail a été soutenu par le généreux parrainage d'AstraZeneca.

Acknowledgements

First of all, I would like to express gratitude to my family for their support. A pivotal point in my voyage had Prof. Ryan Richards thanks to him I discovered research. I am deeply grateful to my supervisor Dr. Darren Hart who honoured me to be his first PhD student. He was an enthusiastic supervisor with whom I had numerous bright discussions. I also want to thank all the wonderful members of our laboratory who helped me keep good motivation and reminded me the positive aspects of matters in dark days. Further thanks go to Natasa Djordjevic for being a very helpful and joyful trainee; Bridgette Connell and Sergiy Avilov for their helpful comments in proofreading this manuscript.

This work was supported by the generous sponsorship of AstraZeneca.

Préface

Cette thèse est organisée d'une manière modulaire. Chacun des chapitres suivants peuvent être lus en tant que texte indépendant et autonome mais sont aussi interconnectés. Le premier chapitre est une introduction globale de thèse avec un accent sur les protéines kinases, leur importance dans la régulation des processus cellulaires ainsi que les cibles des inhibiteurs. Problèmes de la surexpression de protéine dans le domaine de la biologie structurale et des approches de l'évolution dirigée qui les abordent sont également abordés dans ce chapitre. Chacun des chapitres principeaux contient sa propre introduction, méthodes, résultats et des articles de discussion. Le deuxième chapitre décrit les principales méthodes utilisées pour ce travail, Esprit et son optimisation supplémentaire. Les chapitres suivants illustrent l'application de la méthode sur un ensemble de protéines kinases humaines. Le sixième chapitre est consacré à une discussion globale sur les résultats et les perspectives futures. Enfin, toutes les matières générales, les méthodes et le matériel sera trouvée en annexe.

Cette thèse est écrite en Anglais. Néanmoins, des résumes en Français sont fourni au début du chaque chapitre.

Preface

This thesis is arranged in a modular manner. Each of the following chapters could be read as a self-independent text but are also interconnected. The first chapter is an overall thesis introduction with a focus on protein kinases, their significance in regulating cellular processes and as drug targets. Also covered in this chapter is the problem of soluble protein overexpression in the field of structural biology and various directed evolution approaches addressing it. Each of the main chapters contains its own introduction, methods, results and discussion sections. The second chapter describes the major method in use for this work, ESPRIT, and its further optimisation. The following chapters demonstrate the application of the method to an array of human protein kinases. The sixth chapter is devoted to an overall discussion on the results and future perspectives. At the end, in an appendix, all general materials, methods and equipment are supplied.

Finally, this thesis is written in English but short summaries in French are provided in the beginning of each chapter.

Table of Contents

List of Figures

List of Tables

"*Twenty years from now you will be more disappointed by the things that you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor. Catch the trade winds in your sails.*

Explore. Dream. Discover."

--- Mark Twain

Chapter 1: Introduction

Résumé

Les cellules ont des mécanismes complexes pour la détection et la réponse aux changements environnementaux. Un élément majeur de ce dernier est un groupe de protéines appelées kinases. Ce sont des protéines bien conservées qui agissent en tant que strictement réglementée interrupteur de marche-arrêt qui négocie de la réponse à des stimuli extracellulaires. En raison de leur importance dans diverses maladies parmi lesquelles la plus visible - le cancer, ils sont des cible désirés pour le développement de médicaments. Important pour ses avancements est d'information structurel, mais malgré touts les efforts, un certain nombre de protéines kinases avec un rôle central restent non caractérisés. Généralement, le problème réside dans l'expression des protéines insolubles. Deux voies principales d'évaluer la solubilité des protéines sont, par évaluation directe du phénotype par une séparation physique ou par une mesure indirecte d'un système du rapporteur. Les derniers sont généralement utilisés dans le domaine de l'ingénierie des protéines surnommé l'évolution dirigée. Alors que l'approche à l'évolution est nouvelle dans le domaine de la biologie structurale, de nombreux précédents démontrent leur applicabilité lorsque les approches traditionnelles ont échoué. Ceux-ci ont introduit le concept de criblage d'une diversité génétique dans le domaine où, traditionnellement, le procès itératif de la conception et essai des constructions est appliqué. Dans cette thèse, une approche de l'évolution dirigée est prise pour la caractérisation structurale d'un ensemble de protéines kinases.

Cell Signalling and Disease

Cells have intricate mechanisms of sensing and responding to environmental changes. Upon a stimulus, the complex system of signal transduction is activated and this results in a change in gene expression patterns as a response to ensure cellular homeostasis. The vast diversity in the array of signal transduction mechanisms has led to the concept of the signalsome – a unique repertoire of cell signalling components (Berridge, 2008). Since the first scientific publication referring to signal transduction 37 years ago (Rensing, 1972), major efforts have been put into understanding its mechanisms. It was established early that there exist different signal transduction pathways each composed of multiple components. The classical pathway comprises a receptor which when stimulated relays the signal through protein kinases to transcription factors that then regulate gene expression. It is probably an oversimplification to think of these as linear pathways since significant crosstalk between the various pathways is required for a precise and specific response to the initiating stimulus.

Diseases Related to Signal Transduction Aberrancies

Many diseases are linked to cell signalling gone awry. Cancer, arthritis, asthma and many others have been associated with aberrant gene regulation - the consequence of abnormal cellular signalling. A major cause of mortality worldwide is attributed to cancer accounting for around 13% of all deaths (WHO reports). Cancer arises from uncontrolled cell division resulting from expression of oncogenes – mutated forms of normal cellular genes capable of transforming a cell when activated. A proto-oncogene is the unmutated counterpart of the oncogene that, upon mutation, becomes an oncogene and triggers or sustains uncontrolled cell division and growth (Nelson, 2003).

In 1960, Peter Nowell and co-workers discovered a chromosomal abnormality causing chronic myelogenous leukemia. This abnormality is caused by translocation between chromosomes 9 and 22 (Philadelphia chromosome) and results in fusion gene between Breakpoint Cluster Region (*BCR*) and *Abl1* (Nowell and Hungerford, 1960,Rowley, 1973,Westbrook, 1988) where *Abl1* encodes Abelson protein kinase, part of the Src family of protein kinases (Westbrook, 1988). This fusion results in the loss of the regulatory mechanism for the kinase and results in cancer development - an example of a genetic disorder that leads to the aberrant regulation of a protein kinase component of a signal transduction network.

In 1979, Nobel Prize winners, Michael Bishop and Harold Varmus found that the causative agent for fibrosarcomas in chickens, after infection with Rous sarcoma virus, was an oncogene that encoded a mutated protein kinase termed v-Src (viral Sarcoma). This mutation altered the regulation of its activity rendering it constitutively active, leading to cellular proliferation the resulting cancer (Oppermann*, et al.*, 1979,Stehelin*, et al.*, 1977). This was the first example of a virus induced cancer, and it again demonstrated how changes in the activity regulation of a signal transduction component could lead to a disease.

The current hypothesis about Alzheimer's disease links the disease phenotype to the appearance of senile plaques and tangles composed of extracellular deposits of amyloid β (Aβ) and overphosphorylated tau protein. These plaques are composed of 40 and 42 Aβ peptides derived from cleavage of an amyloid precursor protein (APP). It is believed that the glycogen synthase kinase-3 (GSK-3) is involved in the formation of the plaques and tangles in two ways: phosphorylation of the protein tau, the main component of the tangles, and activation of an enzyme cleaving the amyloid-β peptides from the precursor protein (Berridge, 2008,De Strooper and Woodgett, 2003). Alzheimer's is difficult to study due to

the aggregation phenotype of the tangles and plaques, and the current hypothesis has yet to stand the test of time. Nevertheless, once again a plausible cause of the onset of the disease lies in the misregulation of a protein kinase.

Protein kinases are, thus, shown to be key players in the regulation of signalling pathways that relay signals from receptors to the nucleus. Studying these enzymes is essential to further develop our understanding of signalsome complexity. In an important application of this fundamental knowledge, it is clear that modulation of kinase activity can provide successful therapeutic strategies targeting various diseases. Structural information of kinases is therefore not only important for understanding biological mechanisms, but essential for the design of small molecule inhibitors for use as drugs.

Protein Kinases – Structural Features and Regulation

Reversible phosphorylation has long been appreciated for its role in signal transduction. Small post-translational modifications like phosphorylation can result in significant conformational changes in the target protein, a property widely used for activation and inactivation of proteins (Alberts, 2002). This effect is commonly used in cellular signalling and therefore, protein kinases, the enzymes catalyzing the reaction, play a central role.

Protein kinases were found to be the second most abundant group of proteins in eukaryotes comprising 1.7% of the human genome (Manning, *et al.*, 2002). They transfer the γ-phosphate from ATP to the alcohol group of an amino acid side chain of a target protein and are thus classified as tyrosine, serine/threonine or dual specificity (Hanks and Hunter, 1995). Of high importance is another group of kinases that phosphorylates the lipid molecule phosphatidylinositol which also has an important role in cell signalling. The so-called phosphoinositide3-kinases are much less studied than protein kinases but exhibit some

structural similarities. There also exist other kinase enzymes that phosphorylate histidine and arginine residues, but these are evolutionarily distinct and are structurally and functionally dissimilar to the other kinase classes (and are not discussed further here).

Structural Features of the Kinase Catalytic Domain

Protein kinases are commonly multidomain proteins composed of a catalytic domain of a relatively conserved size and fold, together with various regulatory domains that are often highly protein-specific. The kinase catalytic domain is generally 250-300 amino acids long and organised into two lobes: an N-terminal lobe (N lobe) that is β-strand rich and a C-terminal lobe (C lobe) composed largely of α-helices (Fig. 2A). The N lobe has a topology typical for nucleotide binding proteins. It is connected to the C lobe via a loop referred to as the hinge region that allows certain flexibility between the two regions. The C lobe is responsible for the substrate binding, orientation, and catalysis with some almost invariant amino acids. Nevertheless, kinase sequences diverge extensively at the end of the C lobe and it is often difficult to define the C-terminal domain boundaries from sequence analyses, especially for ser/thr kinases. In the case of tyrosine kinases the C-terminus is commonly a hydrophobic amino acid ten positions after an invariant Arg (Hardie, 1995). All following conserved amino acids described in this section will be given with reference to cAMP-dependent protein kinase catalytic subunit alpha, PKA-Cα (Swiss-Prot Accession Number: P17612).

Kinase Subdomains

The conserved topology of the catalytic domain has distinguished secondary structure elements referred to as the twelve subdomains. Throughout, there are some invariant amino acids in regions important for ATP binding, catalysis and structural integrity:

- Subdomain I Gly50 and Gly52 - compose the G-rich loop

- Subdomain II Lys72 - β3 in the N lobe
- Subdomain III Glu91 - αC the only α-helix in the N lobe
- Subdomain VIB Asp166 and Asn171 - catalytic loop
- Subdomain VII Asp184 and Gly86 - DFG motif/loop
- Subdomain VIII Glu208 and Asp220 – activation loop
- Subdomain IX Gly225 - αF scaffold
- Subdomain XI Arg280 - reference point for the terminus of the domain

These are illustrated in Figure 2B & C (Hanks and Hunter, 1995,Kornev*, et al.*, 2008).

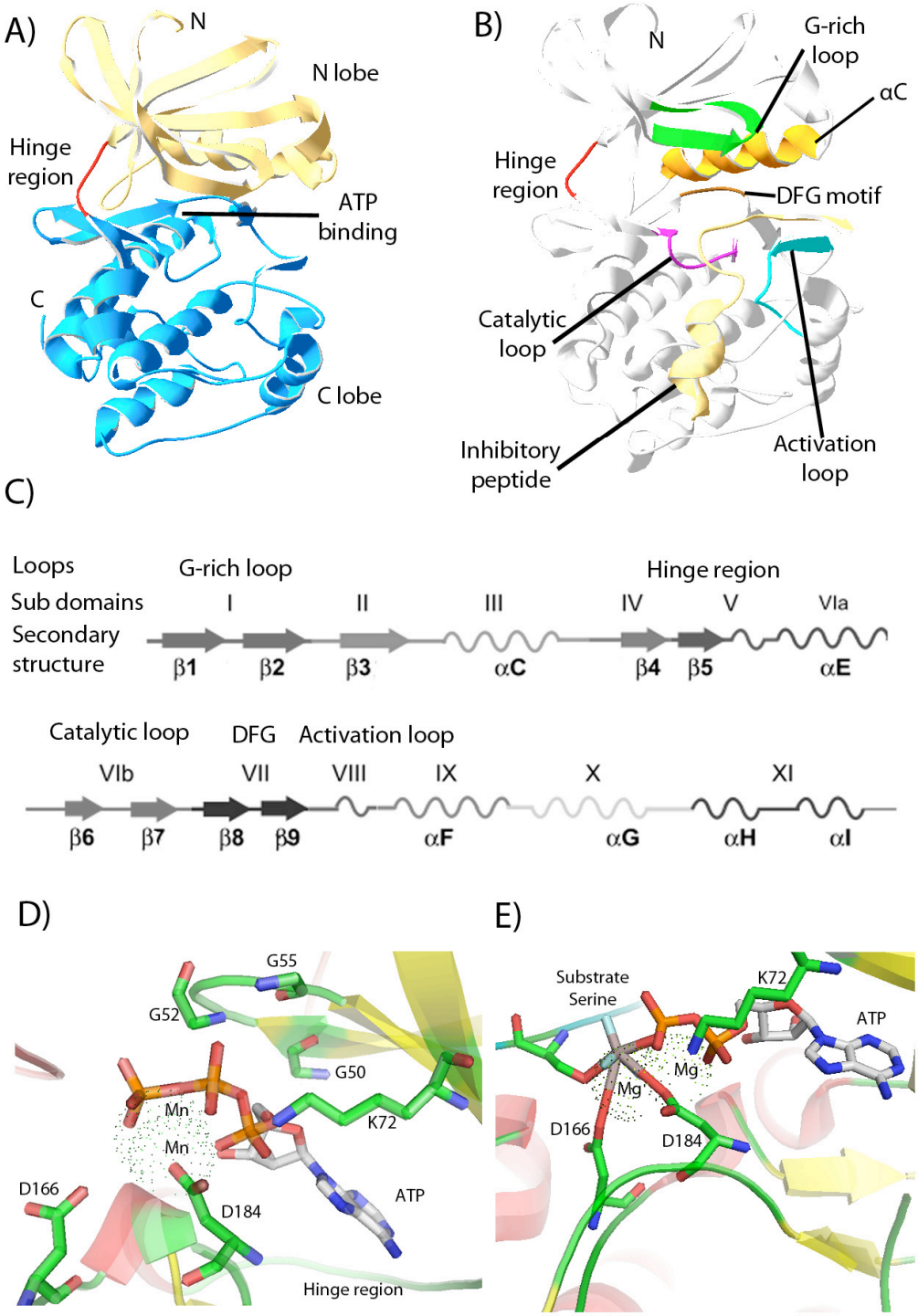


FIGURE 1.1: STRUCTURAL FEATURES OF A KINASE CATALYTIC DOMAIN ILLUSTRATED ON PKA Cα. (A) two lobe structure joined by a hinge (PDB code: 2GU8); (B) Secondary structure and major structural features (PDB code: 2GU8); (C) The twelve conserved secondary structure subdomains (Kannan, et al., 2007); (D) ATP binding illustrated on (PDB code: 1ATP); (E) The active site illustrating the transition state of the γ-

PO4 of ATP being transferred to a serine substrate (PDB code: 1L3E); (amino acids are presented with their one letter codes)

Kinase enzymatic activity relies on precise spatial orientation of the residues important for ATP binding and catalysis. The major structural elements responsible for these conformational changes are the DFG loop, the αC helix, the catalytic loop and the activation loop. Together, they provide the flexibility and capacity to create an "on/off' activity switch.

ATP Binding

ATP binding is mediated by the invariant glycines on the G-rich loop, Lys72, Glu91, hinge region and the DFG loop (Fig. 2D). The glycines, located as on a lid on top of the ligand, make contacts with its phosphate groups. So does, the invariable Lys72 in subdomain β3 making contacts with the α- and β-phosphates of the nucleotide in order to coordinate it correctly for enzymatic activity. The latter is possible because of a salt bridge formed between the invariant Glu91 in the αC. In an inactive kinase catalytic conformation, the αC changes orientation causing the disruption of this interaction. The hinge region does not contain invariable amino acids but it conserves hydrophobic ones that interact with the purine base of the ligand. The residue Asp184 of the DFG motif plays a significant role in binding the ligand in its enzymatically favourable coordination by chelating the Mg^{2+} cofactor interacting with the ATP α- and β-phosphates (Figure 2D).

DFG Motif

The DFG motif has been recognized as an on/off switch due to the conformational change that the small loop exhibits known as "in/out." The Gly186 in the triad assures high flexibility because of the conformational changes it can assume. On the other hand, Phe185 participates in a hydrophobic cluster with amino acids from the αC helix and the catalytic loop, which makes it responsive to their conformational changes. For example, upon correct orientation of the αC helix and the catalytic loop, the phenylalanine triggers the DFG "in" conformation,

consistent with the ATP coordination mode required for catalytic activity (Huse and Kuriyan, 2002,Kornev*, et al.*, 2006).

Activation Loop

The activation loop is the trigger for the conformational changes that the αC, DFG and catalytic loop exhibit in order to activate the kinase catalytic domain. Classically, in closed conformation, it prevents ATP and/or substrate from binding whilst in the open, extended orientation it coordinates ATP and mediates substrate recognition. In its extended conformation, upon phosphorylation on Thr197 (PKA-Cα numbering) the phosphoryl group forms hydrogen bonds with Arg165 in the catalytic loop and Lys189. The latter is responsible for β-sheet formation (β6 and β9) and triggers rotation of the peptide bond at DFG +2 position, which results in additional stabilizing hydrogen bonds with Arg165. Finally, within the DFG triad, the peptide bond between Phe185 and Gly186 rotates and allows formation of a hydrogen bond between the Gly186 and the Asp184 resulting in correct orientation of the aspartate (Kornev*, et al.*, 2006).

Catalytic Mechanism

Kinases use a general acid-base mechanism described in Fig. 3. Upon substrate binding, the aspartate from the catalytic loop acts as a base and accepts the proton from the alcohol group. That triggers the nucleophilic attack of the negatively charged oxygen onto the γ-phosphate forming a transient complex (Fig. 2E; Fig. 3). The alcohol group of the substrate amino acid then detaches the γ-phosphate stabilised by Mg^{2+} or Mn^{2+} and the intermediate dissociates into the reaction products: phosphorylated side chain and ADP. The latter has a reduced affinity for the binding pocket and is displaced by a new ATP molecule. (Cheng*, et al.*, 2005,Szarek*, et al.*, 2008).

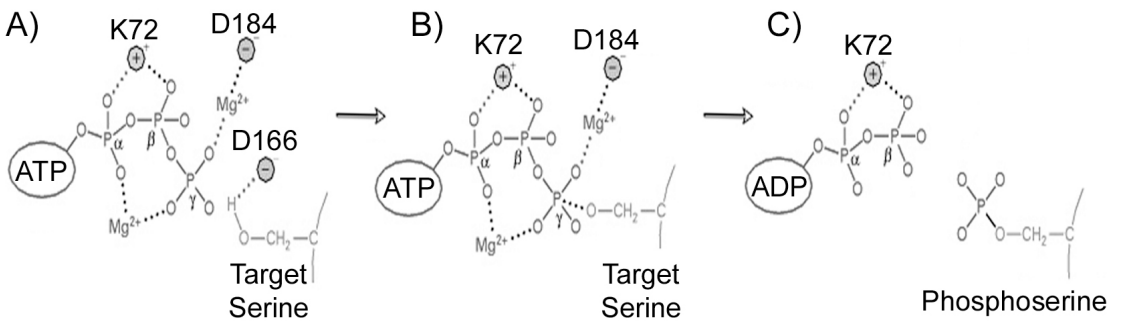


FIGURE 1.2: KINASE CATALYTIC MECHANISM.

(A) Nucleophilic attack of the negatively charged oxygen of the alcohol group; (B) Transition state; (C) Completion of the reaction with γPO4 detachment. (One letter amino acid code with numbering as in PKA Cα);

Activity Regulation

It is remarkable how protein kinases achieve high specificity given the high homology between their catalytic domains. It is essential that their activity is highly regulated given the roles they play in cellular homeostasis. Structural studies have revealed notable similarities in their active forms but significant differences in their inactive forms, the state in which they are commonly present. Their activation is subject to multiple layers of control that ultimately lead to structural changes and adoption of the active conformation of the catalytic domain (Petsko, 2004). The basic conformational changes that render a protein kinase active have been described above but these changes are often complemented by another level of regulation outside of the catalytic domain. On that level of control, the mechanisms involved often rely on regulatory domains, inhibitory autophosphorylation, binding of other proteins, small molecules or lipids. Often the regulation is specific to the protein kinase but there is often a certain similarity within the classes and families.

Regulation via Additional Domains

Src is a family of tyrosine kinases that represent an intriguing example of activity regulation. In its inactive form, the enzyme has a neighbouring SH2 domain bound to a phosphorylated

C-terminal tyrosine and a SH3 domain attached to a polyproline helix. Upon dephosphorylation of the tyrosine or alternative attachment of SH2 to the phosphorylated tyrosine of another active kinase the polyproline helix melts and the SH3 domain is free. Consequently, the two domains unlock the conformation of the kinase catalytic domain. Full activation of the kinase is achieved after a change in the orientation of the αC helix in the N lobe and subsequent phosphorylation of tyrosine in the active loop that opens it up (Fig 1.3 A). The Src family of protein kinases is an excellent example of how additional domains can regulate the activity of the protein kinase (Xu*, et al.*, 1999,Young*, et al.*, 2001).

Regulation via Binding of an Additional Protein

Cyclin-dependent kinases drive the cell cycle by periodic activation upon binding a regulatory protein called cyclin. In the inactive state, its activation loop occupies the entry to the active site thus preventing association of ATP and limiting enzyme activity to only 0.3% of the activated form. When cyclin binds to the kinase it leads to change of the αC conformation and opening up of the activation loop. A threonine residue located on the latter is specifically targeted for phosphorylation by a Cdk-activating kinase (Figure 1.3 B). Finally, the outcome of cyclin-binding in combination with the phosphorylation of the amino acid in the activation loop is the precise coordination of the ligand and the catalytic residues compatible with full activation of the enzyme (Brown*, et al.*, 1999,Cook*, et al.*, 2002).

Occasionally, catalytic domains have long C-terminal extensions that fold back onto the domain and sit in the active site inhibiting the enzyme (e.g. Ca^{2+} /calmodulin kinase II). Upon appropriate Ca^{2+} concentration, calmodulin competes for binding to the extension with nanomolar affinity. When the two subdomains of calmodulin are saturated with Ca^{2+}, it wraps around the C-terminal helix of the kinase and pulls it away from the catalytic domain thereby providing access to the substrate and its turnover (Crivici and Ikura, 1995). Intriguingly, in its

inhibited state, CaMKII forms homodimers with its helical carboxyl extensions rather than occupying the active site. Moreover, the holoenzyme forms a dodecameric assembly of inactive proteins providing an exquisite sensitivity and response to different concentrations of Ca^{2+} (Rosenberg, *et al.*, 2005).

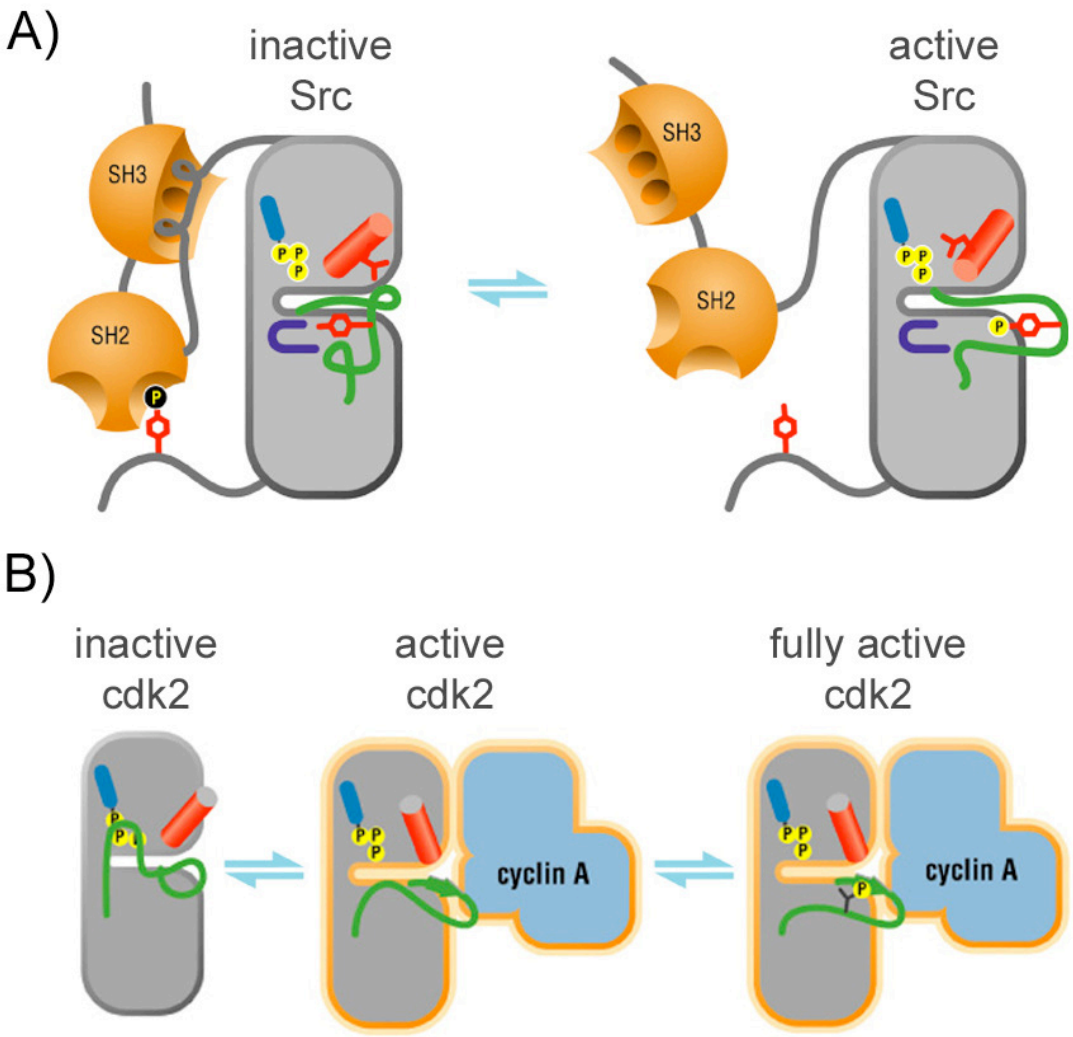


FIGURE 1.3: KINASE ACTIVITY REGULATION.

(A) Src activation mechanism depends on additional regulatory domains; (B) Cdk2 activation depends on protein binding and autophosphorylation; *adapted from* (Petsko, 2004)

Druggability of Protein Kinases

Due to their high structural similarity and common mode of enzymatic action, combined with the multitude of important processes they regulate, protein kinases were once considered unfeasible drug targets. Moreover, the considerably high concentration of intracellular ATP (2-10mM) implies successful inhibitors should bind with nanomolar affinity. Unlike proteases, kinases do not have prominent substrate binding pocket to allow higher inhibitor selectivity (Margutti and Laufer, 2007). Therefore, kinase inhibitors are commonly unhydrolysable ATP analogs like AMP-PNP or diverse compounds such as the natural alkaloid staurosporin originally isolated from *Streptomyces staurosporeus*. Their mode of action relies on displacing ATP in order to block the enzymatic activity.

Drug Types

Over the past decade, the promise of targeting kinases specifically has started to be realised. At present there are ten approved small molecule drugs and several humanized monoclonal antibodies (Table 1) (Johnson, 2009). Noteworthy is that almost all address various types of cancer where side effects are more acceptable and that all small molecule inhibitors target the ATP binding pocket. Some of the drugs target the active form of the kinase catalytic domain, so called Type I drugs based on direct ATP competition. Others bind in the ATP binding site in a manner that is selective for the inactive form of the kinase hence leading to higher specificity. The last are referred to Type II drugs and considered as allosteric inhibitors (Fabian*, et al.*, 2005,Margutti and Laufer, 2007,Noble*, et al.*, 2004). It is perhaps surprising how pharmacological and pharmacokinetic requirements for specificity and tight binding have been met by targeting an active site that is so common in so many proteins. Achieving specificity towards a particular kinase is often difficult and structure-based approaches can help by allowing exploration of ways to expand structure of the potential inhibitor to create more target-specific molecular contacts.

Generic Name	Brand Name	Type	Approved in	Targets	Disease
Fasudil		I	1999	ROCK	Cerebral vasospasm
Imatinib	Gleevec	II	2001	ABL, ARG, PDGFR, KIT	CML
Nilotinib	Tasigna	II	2007	ABL, ARG, PDGFR, KIT	CML with resistance to imatinib
Dasatinib	Sprycel	I	2007	ABL, ARG, PDGFR, Src, etc.	CML with resistance to imatinib
Gefitinib	Iressa	I	2004	EGFR	Non-small cell lung cancer (NSCLC)
Erlotinib	Tarceva	I	2004	EGFR	NSCLC
Lapatinib	Tykerb	II	2007	EGFR (Erb1, Erb2)	Breast cancer
Sorafenib	Nexavar	II	2006	VEGFR, PDGFR, etc.	Renal cell carcinoma
Sunitinib	Sutent	I	2006	VEGFR, PDGFR, etc.	Renal cell carcinoma
Temsirolimus	Torisel	Alosteric	2007	mTOR	Renal cell carcinoma
Trastuzumab	Herceptin	Fab	1998	EGFR (Her2)	Breast cancer
Cetuximab	Erbitux	Fab	2006	EGFR	Colorectal cancer
Bevacizumab	Avastin	Fab	2004	VEGFR	Colorectal cancer

TABLE 1.1: CURRENT DRUGS TARGETING KINASES ON THE MARKET.

adapted from (Johnson, 2009)

Active Conformation Inhibitors

Active catalytic domain conformations can be effectively assayed *in vitro*. This may in part explain why the majority of kinase inhibitors target its active conformation. Fasudil was the

first drug targeting a kinase and was developed in Japan using structural information from PKA (Johnson, 2009). However, the target protein kinase is not PKA, but Rho-associated coiled coil kinase (ROCK) which has an ATP binding pocket highly similar to that of PKA. As expected, this drug is poorly selective and binds to a number of other protein kinases but it is successfully administered for cerebral vasospasm and more recently Alzheimer's. This example illustrates the utility of prior structural information in the design of successful drugs against protein kinases.

Another example of a drug targeting the active site is Dasatanib (Fig 1.4 A). It is a common inhibitor for Src kinases. Intriguingly, it was developed to surpass mutations limiting the functioning of another, highly specific inhibitor, Imatinib (Gleevec, below).

Inhibitors of Inactive Conformations

An example of highly selective, successful drug in this category is Gleevec, the inhibitor of Bcr-Abl kinase (Figure 1.4 B). Essential for the action of the drug is the adoption of an inactive conformation of the kinase with activation loop mimicking substrate and DFG loop adopting the "out" position. Imatinib stabilizes this state of the kinase by making contacts with the activation and the DFG loop and keeping them tightly in an inactive conformation. It is therefore highly specific and shows low cross reaction with other kinases (Huse and Kuriyan, 2002).

Allosteric Inhibitors

Drugs targeting the mammalian target of rapamycin (mTOR) like Torisel deserve special attention as they represent an entirely different concept for small molecule kinase inhibitors. The action of Torisel is similar to that of rapamycin: both bind to an effector protein called FKBP12 and thus inhibit its interaction with the FKBP12 rapamycin binding (FRB) domain located amino terminal from the catalytic domain of mTOR (Figure 1.4 C). Due to the

inhibited effector, mTOR remains inactive and so the drug indirectly interferes with its regulation of cell protein synthesis and proliferation (Hudes, 2007).

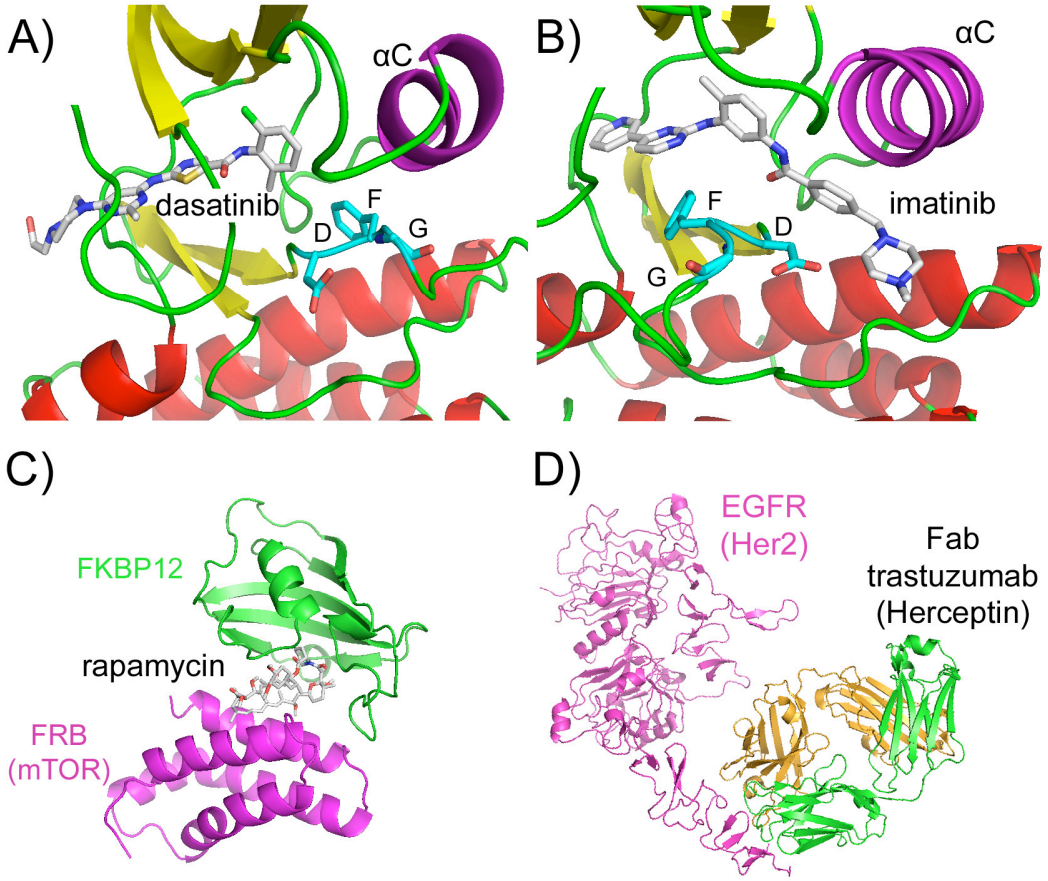


FIGURE 1.4: FOUR MODES OF KINASE CATALYTIC ACTIVITY INHIBITION

(A) Dasatinib binds to "DFG in" the kinase active conformation (PDB code: 2QGQ); (B) Imatinib recognizes the "DFG out" inactive conformation (PDB code: 2HYY); (C) Rapamycin interferes with mTOR activity by allosteric inhibition (PDB code: 1FAP); (D) Fab Trastuzumab (Herceptin) binds to the EGFR Her2 leading to its internalisation (PDB code: 1N8Z);

Humanized Antibodies

Another inhibitor strategy is based on humanized monoclonal antibodies. An example for such inhibitor is Fab Trastuzumab that recognizes the extracellular portion of the Her2 variant of the EGFR common in breast cancer (Figure 1.4 D). Upon antibody binding, the receptor is internalised and proteolysed. This drug has proven efficient in 15% of the patients

overexpressing the receptor in metastatic breast cancer and its action is further improved when used in combination with chemotherapy (Cho*, et al.*, 2003, Kute*, et al.*, 2004). An interesting aspect of this treatment is that patients can be prescreened to verify that they will be responsive to therapy, simply by testing for antibody reactivity with samples from the patient.

Inhibitor Profiling

Although Imatinib (Gleevec) is highly selective for the specific inactive form of the Bcr-Abl kinase it also binds three other non-target proteins. Inhibitor profiling of Dasatinib, another drug prescribed for chronic myelogenous leukemia (CML), targeting the same protein, shows a lower selectivity and much broader range of 35 other kinases (Hantschel*, et al.*, 2008). The difference between the two drugs is in the mode of inhibition. Imatinib is a Type II inhibitor targeting the inactive conformation of Abl kinase and Dasatinib is Type I, binding to its active conformation. Interestingly, Dasatinib is a 300 times higher affinity binder for Bcr-Abl than Imatinib and therefore exhibits higher resistance to common mutants of the protein. This study demonstrated how two drugs designed against the same target could have completely different inhibition profiles depending solely on the form of the protein. While both compounds are proven successful clinical inhibitors, they demonstrate how a good structural understanding of the protein kinase active and inactive conformation is a prerequisite for successful drug design. Furthermore, establishing *in vivo* roles for a specific kinase commonly exploits inhibitors as research tools to neutralize the kinase and study the effect on the pathways involved.

Challenges in Structural Biology and Protein Kinases

Structural biology is the field concerned with the molecular structure of biological macromolecules, especially proteins. The objective commonly is to understand the

relationship between their structure and how alterations affect their function. As demonstrated in the previous sections, structural knowledge on protein kinases is important for structure-oriented drug design towards treating diseases and studying various cellular signalling pathways.

Stuctural Genomics

One of the early products of the various genome sequencing projects was a large catalogue of open reading frames encoding both known and unknown proteins. A branch of structural biology arose to attempt to process this data into 3D protein structures. Protein Structure Initiatives (PSI) were created with the ambitious aim to express and determine structures covering a large sequence space so that computational tools can aid a more thorough coverage at a next step (Burley, 2000). These efforts resulted in "structure pipelines" capable of processing a multitude of targets in high-throughput fashion. At the front end of such pipelines, stands target selection aiming to establish a list of proteins conforming to the aim of the project e.g. novel structures/folds, or targets with significance to human health. The next steps are automated high-throughput cloning, expression, purification, crystallization, data collection, and processing. Finally, methods of semi-automatic publication are intended to annotate the new structures and deposit them in the PDB (Service, 2008).

Structural genomics has not been without criticism, generally related to the choice of targets and, in the case of the PSI, its focus on fold space coverage. Some commentators suggest it has overlooked difficult proteins and has not deposited enough novel or valuable protein structures. However, an undebatable fact is that it has driven methodological progress in most areas of structure determination. Beyond this, protein bioinformatic tools have developed significantly with benefits in tools ranging from structure solution to prediction of function (Watson*, et al.*, 2007). Moreover, the critical assessment of structure prediction (CASP)

project has shown significant improvements in structure prediction accuracy since its start in 1994 (Cozzetto, *et al.*, 2009). It has even been shown that structure prediction models can sometimes be used for structure determination by molecular replacement (Qian, *et al.*, 2007). Finally and importantly, along with the numerous new structures a consequence of the protein structure initiatives is the generation of accurate metrics describing the difficulties in expressing, purifying and crystallising proteins.

Soluble Protein Expression Bottleneck

Large amounts of soluble purified protein are required for structural characterisation and inhibitor screening (Blundell, *et al.*, 2002). For all the recent advancements in structural biology, a major bottleneck remains - the obtention of soluble protein. Protein expression in *Escherichia coli* remains the standard for heterologous protein production due to well-established recombinant DNA manipulation techniques, low cost, and eases of isotope and heavy atom labelling. However, these advantages are accompanied by problems of aggregation and degradation (Dobson, 2004) which are relatively common in *E. coli*. Sometimes targets can be expressed more successfully in eukaryotic expression hosts such as mammalian, yeast or insect cells, but these are far from universal solutions and insufficient expression still accounts for the lack of structural characterization of a large number of proteins.

Isolation of Protein Domains

It is not always clear why viral and mammalian targets are more difficult to express than those from bacteria, but one common observation there is that they have a higher complexity of domain organization. In contrast to prokaryotic proteins, they are often larger and comprise multiple domains connected by longer linkers or low-complexity regions (Ward, *et al.*, 2004). Frequently they may be subunits of multi-component complexes, or at least

require interaction of partners for stability, either via binding or post-translational modification. Expression of such proteins in full-length form in *E. coli* frequently results in aggregation or degradation. Even when expressible in alternative hosts, the protein may not be in a crystallisable form due to high flexibility.

When an entire protein cannot be expressed, it is common practice to isolate separate domains. Classically, the identification of domains within proteins relies on conservation of regions between homologues present in the sequence databases. Therefore, the successful outcome of this approach is totally dependent on the presence of sequenced homologues and when none are present in the databases, studies can be blocked. In the case of many viral genes that mutate rapidly and have poorly understood evolutionary histories, this is often the case. Furthermore, sequence analysis-based approaches sometimes fail to yield large amounts of soluble material even when homologous sequences with useful sequence conservation are available, as in the case of human protein kinases. The conclusion is that expression of stable domains is an intricate process dependent on many unpredictable underlying factors beyond those identifiable from sequence analyses including folding efficiency, requirement for chaperones, ligands and protein partners, toxicity, mRNA secondary structure, codon usage and intramolecular stabilisation from other parts of the protein.

Kinases as a Challenge in Structural Biology

Since the first protein kinase structure of PKA-Cα in 1991 our knowledge of this class of proteins has increased significantly canonizing the various structural elements of the catalytic domain and elucidating some spectacular mechanisms of activity regulation. While structural genomics initiatives have contributed significantly to the improved structural coverage of the human kinome (Marsden and Knapp, 2008) alternative approaches are still required for kinases resisting overexpression.

It is perhaps surprising that many protein kinases have resisted protein expression since the structure of their catalytic domains appears well-defined and highly conserved. Some difficulties may be linked to the loose definition of the catalytic domain C-terminal limits (as discussed earlier, it sometimes folds back to the domain and/or stabilizing). Other likely problems are poor folding efficiency and necessity/heterogeneity of post-translational modifications. Being an enzyme, overexpression may result in an unspecific, toxic activity in the heterologous expression system.

Some of the factors that complicate expression of kinases can start to be addressed using synthetic genes that reduce mRNA secondary structure or codon limitations that reduce translation efficiency in *E. coli*. However, this technology is still in development and the real factors that enhance expression are still being defined (Welch*, et al.*, 2009). Inactive mutants are often used to reduce the toxicity effects, as is coexpression of phosphatases. Even with these tricks, there is still a need for precise domain definition since a good construct is a basic prerequisite for further expression optimisation. Protein expression-compatible boundaries do not always coincide with those predicted from sequence conservation in alignments. One approach that can address this is the generation of random construct diversity coupled to a screen or selection for soluble protein fragments, borrowing strategies from directed evolution methods of protein engineering.

Directed Evolution Approaches in Structural Biology

Obtaining well-expressing protein domains is traditionally a time-consuming process involving iterations of trial-and-error, and guided by prior knowledge of the target for construct design. One way to shortcut this lengthy route, but obtain the same solution, is to make all genetic constructs at once and then screen them for soluble expression (Hart and Tarendeau, 2006). This is the rationale behind directed evolution experiments and several

groups have tried to develop such strategies towards soluble expression of challenging targets (Dyson, *et al.*, 2008,Prodromou, *et al.*, 2007,Savva, *et al.*, 2007).

Mutagenesis Strategies

The mutation strategies that best resemble the conventional methods of cloning subconstructs are gene truncation and fragmentation. Such libraries can be prepared by the enzymatic degradation of the target gene (Anderson, 1981,Lutz, *et al.*, 2001,Reich, *et al.*, 2006), by unspecific random priming PCR (Grothues, *et al.*, 1993,Kawasaki and Inagaki, 2001) and by physical breakage usually employing sonication or point-sink shearing (Thorstenson, *et al.*, 1998) (Table1.5). These various methods are relatively sequence-independent and result in a multitude of different sized gene fragments that can be size-fractionated within desired ranges by agarose gel electrophoresis.

Point mutagenesis protocols employing DNA and family shuffling have also been used to improve poorly soluble proteins (Aharoni, *et al.*, 2004), but do not address the definition of domain constructs. It is probable that a bad starting construct ruins such an approach. Two studies on GFP have revealed different mechanisms by which point mutations can increase expression (Crameri, *et al.*, 1998,Crameri, *et al.*, 1996).

In a different application, library screening can identify variants of a target with altered crystallisation properties. Both point mutagenesis and gene truncation/fragmentation may generate constructs that are less prone to aggregation and change the crystal packing contacts (Aharoni, *et al.*, 2004,Keenan, *et al.*, 2005) in the same way that crystallising homologues, or alternative constructs does in the conventional approach. The application to crystallisability is suitable mainly for proteins for which the functionality can be followed in a high-throughput assay thus allowing monitoring of the variants to identify the subset that maintain the solubility and activity of the native target, whilst containing amino acid changes (Aharoni, *et*

al., 2004). Therefore, it is more appropriate for enzymes compatible with screening for their activity. Additionally, often the levels of catalytic activity are improved along with the solubility profile.

DNA Fragmentation and Truncation			**Randomisation Methods**
Enzymatic-based	***PCR-based***	***Physical breakage***	
Exonuclease III	Random priming	Sonication	ePCR
DNaseI		Point sink shearing	DNA shuffling
Uracyl glycosylase			Family shuffling

TABLE 1.2: SUMMARY OF THE GENETIC DIVERSITY-GENERATING PROTOCOLS SUITABLE FOR LIBRARY GENERATION RELEVANT TO STRUCTURAL BIOLOGY.

One important and unwanted feature of random construct libraries is the presence of non-coding diversity – the subset of gene fragments not producing protein due to out-of-frame constructs. These clones dilute the desired ones and result in decreased screening efficiency and reduced coverage. To address this problem, reading-frame selection systems, or ORF selectors, have been developed employing life-or-death selection. Prominent are developments where the target is cloned between an export signal peptide and β-lactamase; if a colony is resistant to ampicillin, the β-lactamase has been localised in the periplasm and the target is in frame permitting complete read-through of the construct (Lutz*, et al.*, 2002). In a second generation of this system, a split intein has been inserted flanking the target. After translation of an in-frame fusion product, the intein self-splices, complements the β-lactamase with a functional signal peptide and ejects the target protein (either folded or unfolded) into the cytoplasm (Gerth*, et al.*, 2004). The result is that the reading frame of resistant clones is known permitting recloning into protein expression vectors for solubility screening. The advantage of using ORF selectors is strong, particularly for fragment libraries where only 1/9 to 1/18 of the theoretical diversity is in correct reading frame.

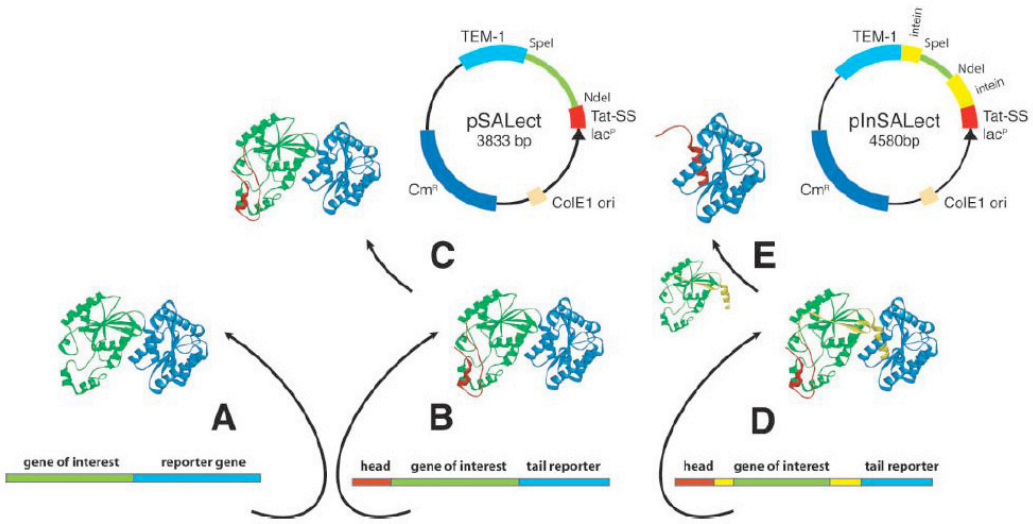


FIGURE 1.5: READING FRAME SELECTION SYSTEMS BASED ON β-LACTAMASE REPORTER GENE.
(A) C-terminal fusion of the reporter (blue) to a gene of interest (green); (B and C) Dual selection system pSALect inserts the verifies the integrity of the gene of interest by its integration between periplasmic export sequence and the reporter gene; (D and E) Development of the latter incorporates intein-mediated splicing thus allowing the protein of interest to remain in the cytoplasm; *source* (Gerth, *et al.*, 2004)

Screening Strategies

The diversity in the DNA libraries needs then to be selected or screened for the desired phenotype, which in here is high yields of soluble and stable protein. There are two major groups of screens 1) assay solubility directly by physical separation of the protein 2) assay indirectly by (i) solubility reporters (ii) small peptides (iii) phage display and (iv) endogenous cellular mechanisms. In all approaches where native activity or function is not the readout, an assumption is made that if a protein is soluble it is folded correctly and therefore functional. An effective screening assay or selection for improved solubility should meet multiple criteria such as ease to work in high-throughput fashion, applicability *in vivo* and *in vitro*, sensitivity to different levels of solubility and finally good correlation with expression scale up.

Direct Assessment of Solubility

Classically protein solubility is assessed by physical separation of soluble from insoluble protein by the means of centrifugation or filters that evaluate directly the phenotype. While centrifugation is not readily adaptable for high-throughput methods, filtration coupled with protein tag detection has been used successfully for assessing solubility of proteins from small libraries of protein fragments and was first demonstrated in a plate format (Knaust and Nordlund, 2001).

More recently, Cornvik *et al* described a screen for protein solubility directly from colonies grown on filters. Colonies induced for protein expression were lysed enzymatically and the proteins allowed to filter through a porous Durapore membrane to a nitrocellulose capture membrane below. This was then probed for the presence of tagged proteins using specific antibodies and thus the soluble constructs identified (Cornvik, *et al.*, 2006) (FigXXXX). This method has been applied to small truncation libraries generated using the Erase-a-Base kit (Promega) and shows significant promise. Due to the plating density employed to get clear signals, colony filtration is a method best suited to small libraries (several thousand variants maximum) and is perhaps less appropriate for large, high diversity libraries.

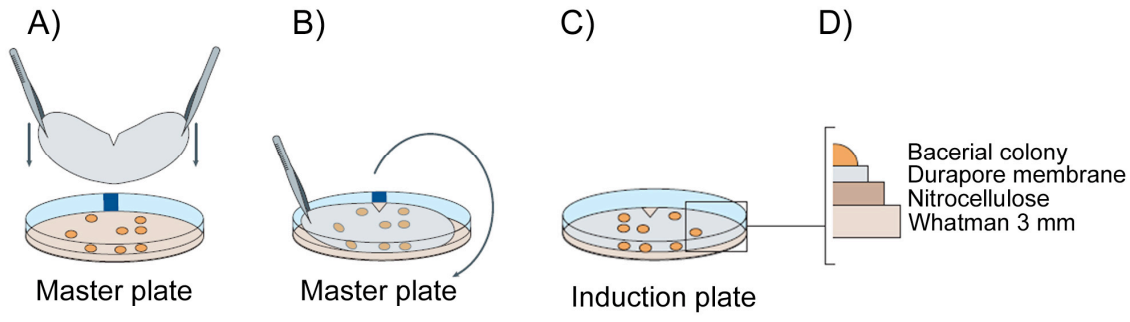


FIGURE 1.6: COLONY FILTRATION BLOT SCHEMATIC.

(A) A Durapore membrane is lowered onto growing colonies on a master platel; (B and C) After a brief contact with the colonies the membrane is transferred to an induction plate; (D) the composition of the "sandwitch"

allowing of the colonies to filter through the porous Durapore membrane to a nitrocellulose membrane below *adapted from* (Dahlroth, et al., 2006)

Along with a novel enzymatic based method for generating fragment libraries, using the random incorporation of uracil and its excision creating double strand breaks, Reich *et al* reported a two step screen for identifying soluble proteins. Firstly, detection of the presence of a protein tag on colony blots suggests the expression of a protein (soluble or insoluble) and secondly, filtration of lysates using a hydrophobic filter identifies for the soluble protein subset (Reich*, et al.*, 2006). In common with the CoFi, method, the constructs tested bore a simple hexahistidine tag. Accordingly, the constructs identified as soluble should be readily scalable with no need for subcloning. This method, though, suffers from suboptimal library generation: it requires gene optimization of the target to recode the gene so that uracils are incorporated in a less biased manner than would be with the native sequence and there are still regions over- or underrepresented as possible cleavage sites. Inserts are cloned into a pool of three vectors with cloning sites in different frames to increase the probability of catching the limited set of inserts in frame. The lack of oriented cloning leads to only 1/18 of the clones being in frame, resulting in assessment mismatch between the diversity generated and the screening power of a few thousand clones. This could potentially be reduced to 1 in 6 if 3-frame stop codon arrangements were used, but this improvement has not been described.

Indirect Assessment of Solubility

Using an indirect assessment of solubility permits high throughput assays to tackle large libraries generated by fragmentation or point-mutagenesis. These assays provide enrichment of positive hits but often require their confirmation with complementary experiments to assess their true soluble phenotype. This is due to higher likelihood of false positives coming out of the assay – in part because of the larger number of clones being tested.

Solubility reporters for library screening, usually C-terminal fusions to target proteins, have been used successfully, but suffer from a "passenger solubilisation" effect (Reich*, et al.*, 2006); constructs identified are often soluble only in the context of the fusion, but not once the reporter protein is removed upon recloning since it is essential to remove larger tags prior to crystallisation.

Among the first solubility reporter was the homotrimer chloramphenicol acetyltransferase (CAT). It was observed that when CAT was fused to soluble proteins it conferred higher resistance to the antibiotic than when fused to insoluble proteins (Maxwell*, et al.*, 1999). It provided a high screening potential due to the nature of its life-or-death selection. Although few applications of this screen have been reported, Maxwell *et al* demonstrated that point mutants of the HIV integrase domain (50-212) were well discriminated in a concentration dependent manner. This selection strategy was also used to identify domains from fragment libraries of yeast Sir3 and Sir4 (King*, et al.*, 2006) and led to the discovery that dimerisation and higher order complexes of these fragments were important in transcriptional regulation.

Another life-or-death selection for soluble protein expression uses dihydrofolate reductase (DHFR) – an essential enzyme for the synthesis of purines, thymidylic acid and certain amino acids (Chen*, et al.*, 1984). The selection is based on inhibition of the *E. coli* endogenous DHFR by trimethoprim (TMP) and complementation of the missing activity by murine DHFR that remains active in context of a solubly expressed fusion protein. Using this selection criterion extra- and intracellular soluble expression constructs of Pecam1 were identified and antibodies produced (Dyson*, et al.*, 2008). Murine DHFR tolerates fusions to its N-terminus and is monomeric, ideal characteristics of a solubility reporter (Dyson*, et al.*, 2004).

The best-known screens for protein solubility employ C-terminal fluorescent protein fusions such as green fluorescent protein (GFP) (Pedelacq*, et al.*, 2002) and ZsGreen (Heddle and Mazaleyrat, 2007). Maturation of the GFP fluorophore requires correct folding and this is sensitive to the solubility status of the GFP-fusion protein as a whole (Waldo*, et al.*, 1999). Several examples of successful variant identification have been reported from fragment libraries (Jacobs*, et al.*, 2005,Kawasaki and Inagaki, 2001) and point mutation libraries (Waldo, 2003). Fluorescent protein fusions are monomeric and non toxic. A potential advantage over life-or-death selections is that they can report small differences in solubility and so are compatible with iterative rounds of mutagenesis where each round only yields small improvements. However, remaining are potential problems with strong passenger solubilisation effects of the GFP and a high rate of false positives due to internal ribosome binding sites in the target sequence, or proteolysis.

Small Peptide Tag Solubility Reporters

One way to avoid the problems of solubility enhancement by C-terminal reporter proteins is to use a small peptide tag. An elegant solution to circumvent the intrinsic solubilization problems of GFP was generated by the original inventor of the GFP screen using a split-GFP strategy where the terminal beta strand (strand 11) is fused to the target (Cabantous*, et al.*, 2005). Upon its soluble expression as a fusion partner, it complements a coexpressed, non-fluorescent truncated form (strands 1-10) of the full GFP resulting in fluorophore maturation. This assay combines the high throughput possibilities of a fluorescence-based screen with the benefits of a small peptide tag (avoidance of passenger solubilisation).

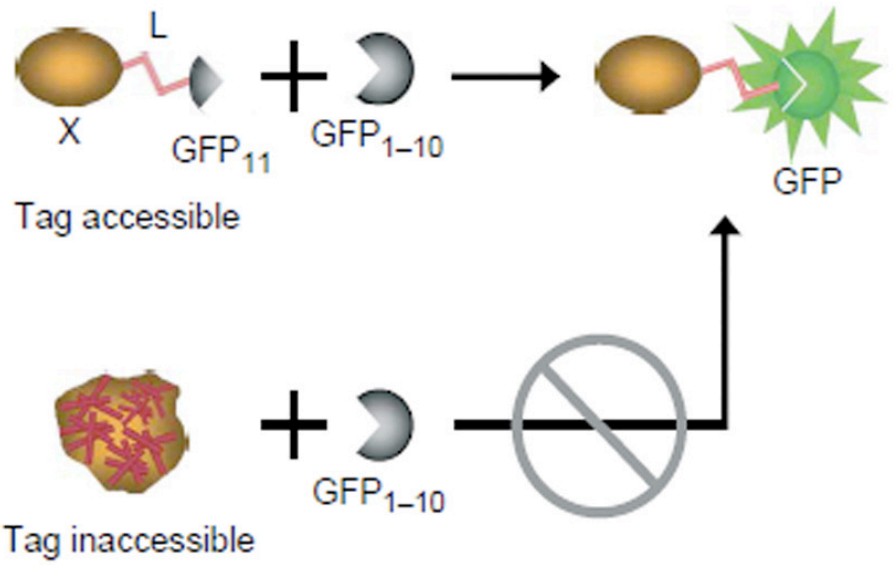


FIGURE 1.7: ILLUSTRATION OF THE SPLIT GFP METHOD.
When the protein X is soluble the GFP β-strand11 tag is accessible and can complement with the truncated counterpart GFP1-10 leading to the maturation of the fluorophore; *source* (Cabantous, et al., 2005)

Another complementation assay is the β-galactosidase solubility assay. A 54 amino acid α-fragment of β-gal was fused to the target protein and in the case of soluble expression, it was shown to complement the coexpressed ω-fragment resulting in blue colonies on X-gal agar (Wigley, *et al.*, 2001). In a variation, this assay was used as a life-or-death selection where complementation of the β-galactosidase conferred survival on minimal media with lactose as the sole carbon source. The benefits of this assay are not only in the small peptide but also in its flexibility to be used either as color-based screen or as life-death selection.

Another method that relies on a small peptide tag rather than whole protein reporter is ESPRIT, developed in our laboratory. This uses efficiency of *in vivo* post-translational biotinylation by the *E. coli* endogenous biotin ligase of a 14 amino acids linear peptide to report on solubility (Tarendeau, *et al.*, 2007). It is also unique in employing high-throughput automation to manipulate libraries to achieve coverage of 30,000 constructs per screening

round and has been used to screen over 100,000 constructs of a single protein, the influenza polymerase PB2 subunit. After many years of failure to overexpress this protein, a series of domains were identified by ESPRIT screening leading to the first structures explaining nuclear translocation domain (Tarendeau*, et al.*, 2007), host adaptation (Tarendeau*, et al.*, 2008) and mRNA cap-snatching (Guilligay*, et al.*, 2008). Those studies demonstrated the application of library-based construct screening for novel proteins with no homologues, where traditional methods of rational construct design had been unable to find a solution. Similar results were obtained on the CagA pathogenicity factor of Helicobacter pylori (Angelini*, et al.*, 2009).

Phage Display

Proteolysis is a classical tool in structural biology. It has been successfully combined with phage display in a method known as PROSIDE to identify soluble variants from libraries of a target. The protein of interest is displayed between the C-terminal domain and the N-terminal N1 and N2 domains of the g3p coat protein of filamentous phage. The other scaffold proteins M13 and fd are resistant to many proteases which allows treatment of the immobilized phage with an array of proteases. It was observed that phages with insertions of a stable target protein remained infectious thus providing a selection for stable protein domains (Kristensen and Winter, 1998,Sieber*, et al.*, 1998). This method was subsequently used on Bs-CspB from *Bacillus subtilis* (Wunderlich*, et al.*, 2005). The obvious limitation here is the fact that proteases have recognition sequences and target proteins may escape cleavage simply by not containing them. The “loop entropy reduction” phage display method avoids this problem and is based on insertions of randomised sequences in a loop of the SH2 domain from lck kinase. In this study, the aim was to exploit that unfolded proteins would exhibit higher entropy and disturb the intact host protein (SH2 domain) and inversely, folded ones would not inhibit its function. Initially, tests on insertions in the IgG domain of protein L showed

that it was not well discriminating between folded and unfolded inserts. Then, insertions in SH2 domain were selected on magnetic beads with immobilized phosphotyrosine peptide, ligand for intact SH2 domain, proving that insertions in this domain were discerning folded phenotype (Minard*, et al.*, 2001). This work demonstrated that the choice of the host protein is essential as well as that it relies on the spatial proximity of the two termini of the inserted target protein.

Endogenous Cellular Mechanisms

A method that exploits native bacterial twin-arginine translocation (Tat) pathway is a another indirect assay for protein solubility. Similar to the ORF selector mechanism described above (Gerth*, et al.*, 2004), it requires the fusion of β-lactamase reporter on the C-terminus of the target protein and a Tat signaling peptide on its N-terminus. Upon soluble expression and, the authors claim, "correct folding" of the protein of interest, ampicilin resistance is conferred, qualifying this as a potentially very powerful life-or-death selection (Fisher*, et al.*, 2006). How sensitive this system really is to the partner being correctly folded will become apparent with repeated use. It is also not clear how proteins selected in the context of low level periplasmic secretion will behave in scale-up expression trials following subcloning of the selected inserts and cytoplasmic expression.

A method that does not require any protein or tag fusions thus, permitting the expression of native proteins, is reported to use *E. coli* stress response promoters known to be upregulated upon heterologous insoluble protein expression. The promoter for a shock protein ibpA was used to control the expression of β-galactosidase followed by assessment of its activity from cell lysates and detection of 6xHis-tagged targets (Lesley*, et al.*, 2002). A downside of this method is simply that it detects insoluble and misfolded protein opposing to correctly folded and soluble one.

Prospects

Several permutations of library construction method and solubility screen have been described and of primary importance is simply to use and benchmark them on a large number of different targets. Only in this way can we understand their true advantages and limitations. One thing that is clear is that many published methods do not appear to be used beyond the primary citation, however this may be due in part to the resistance of the community to adopt an approach that is radically different to the conventional. Collaborations between groups pursuing targets and method developers are crucial in this respect.

It is also clear that eukaryotic expression hosts, e.g. insect cells, are often able to fold and express solubly constructs that are impossible to express in *E. coli* (Hofinger*, et al.*, 2007). Therefore, exploitation of other expression hosts such as insect cells for library screening could offer new possibilities. However, most higher organisms do not grow as colonies and cannot be clonally transfected with single DNA molecules as is the case of *E. coli.* The latter is an absolute prerequisite for efficient random library generation. At some point it may be possible to miniaturise insect cell expression cultures to the point that large numbers of parallel transfections, in individual wells of plates, with small focused libraries could be achieved. In such a format, direct screens for solubility and some indirect screens including fluorescent fusion reporters and reporters that confer resistance could be applicable. Currently, no such method has been reported due to the difficulties underlying the manipulation of insect cells and their use in high-throughput automation. Current advances in automation of virus generation and recombinant protein and protein complexes production in insect cells suggest that this could soon be possible (Bieniossek*, et al.*, 2009,Karkkainen*, et al.*, 2009).

Within 50 years of the birth of structural biology (Kendrew, *et al.*, 1958) we have reached a point where our understanding of protein structure has even enabled *de novo* design of unnatural proteins with novel folds to very high level of accuracy (Kuhlman, *et al.*, 2003). In an interesting variation, protein design efforts can be coupled to directed evolution (Rothlisberger, *et al.*, 2008) and result in the creation of enzymes catalyzing new reactions relevant to human needs. It is human nature to observe and imitate Nature's mechanisms for his own applications!

Thesis objectives

The work presented here commenced at the point at which proof-of-principle experiments in the lab had indicated that screening could identify soluble constructs from a random library of a target (in 2005). The first objective was to improve the system, notably the vectors needed reconstructing so that library construction, screening and purification testing could be performed without recloning steps i.e. in the same plasmid. Secondly, and the principle objective of this work, was to apply it to a set of difficult-to-express targets. These should fulfil the following criteria:

1. to be recognised as challenging and difficult to work on
2. to be of high intrinsic value and interest to add impact to the study
3. to contain novel aspects such that any new domains isolated could be studied and contribute new biological knowledge

We settled on the theme of multidomain human protein kinases due to their high value and the assumption that since many groups had worked on them previously, they could be classed uncontroversially as "difficult." We reasoned that a more quantitative estimate of how difficult-to-express a target gene could be provided by considering the amount of efforts put

into its functional characterisation versus the availability of overexpression and/or structural data (Table 3). The number of primary publications citing the genes was taken, certainly under-estimating the total number of publications related to each target. Publications such as reviews and opinions were omitted.

	DAPK1	IKKβ	TBK1	PI3Kβ	mTOR
Discovery date	1995 (Deiss, *et al.*, 1995)	1996 (Chen, *et al.*, 1996)	1995 (Chou and Hanafusa, 1995)	1993 (Hu, *et al.*, 1993)	1994 (Brown, *et al.*, 1994)
Over expressed	yes (Tereshko, *et al.*, 2001)	no	no	no	no
Structure	yes (Tereshko, *et al.*, 2001)	no	no	no	no
PubMed entries*	75	192	52	60	329

TABLE 1.3: Summary on the human protein kinase targets

*PubMed entries describing characterisation of the corresponding genes counted as of 10/09/09.

In parallel, several targets completely lacking homologues were analysed during this period (by colleagues in the laboratory) where the aim was to identify novel domains in known but unannotated ORFs. When one such protein (influenza polymerase PB2 subunit) was compared under the same criteria it was found that relatively little characterisation existed prior to its overexpression and structural studies. Therefore the kinases selected for this study can be considered a different type of target since they have been well-studied and have well-annotated primary sequences meaning that the approximate location of some of the domains is generally known, albeit in an unexpressable form.

The hypothesis formulated for this study was that *the problem of soluble overexpression in E. coli for these previously unexpressed kinase targets could be resolved by random construct screening*.

In order to test the hypothesis initially, a more characterised system was included in the study. Death associated protein kinase 1 (DAPK1), a Ca^{2+}/calmodulin regulated kinase, was selected since the kinase domain had already been expressed in *E. coli* and crystallised, but the mechanisms of its regulation have still not been fully elucidated meaning that some new biological knowledge might be gained from its analysis. The common biological themes among the rest of the multidomain protein kinases in the panel are that they are linked to cell proliferation, growth and cancer, and are essential regulators within two pathways: NF-κB and PI3K/Akt/mTOR. These cell signalling pathways have been studied heavily with close to 30,000 NF-κB related publications during the past 20 years and almost 9,000 on PI3K. A comprehensive review of these targets is clearly impossible, therefore, in the following chapters, I attempt to provide an up-to-date overview of the pathways and mechanisms of their regulation, with focus on the most recent literature.

"*We can't solve problems by using the same kind of thinking*

as we used when we created them." --- Albert Einstein

Chapter 2: ESPRIT Method

Résumé

D'Expression des Protéines Solubles par Troncation Aléatoire Incrémentielle (ESPRIT) est une méthode d'évolution dirigée pour l'isolation des domaines de protéines solubles dans *E. coli*, conçue dans notre laboratoire. Il se fie sur la troncation enzymatique d'un gène d'intérêt pour la génération de sa diversité génétique. Le criblage de la solubilité est fait à deux niveaux: d'évaluation indirecte de la solubilité par biotinylation en format haute débit et l'évaluation directe de la solubilité par purification typique par Ni^{2+}NTA sur un nombre réduit de l'criblage précédent. La méthode est adaptée pour du travaille automatisé, ce qui contribue à l'isolation du grand nombre de colonies à la première étape et de purification à la seconde. Développements précédents ont testé le concept de la méthode et éprouvé qu'elle puisse être appliquée avec succès pour l'identification des domaines solubles des protéines insolubles. Une contrainte évidente était un fractionnement de taille d'ADN imprécis. Combiné avec la nouvelle, bibliothèques de troncation bidirectionnelle, ce qui augmente considérablement la diversité, l'inconvénient est devenu apparent. Deux approches principales ont été testées pour diminuer la diversité de la bibliothèque pour améliorer sa qualité: diminution de la taille du plasmide d'accueil et de la conception et vérification d'un tag peptidique N-terminal qui tiendrait compte des capacités la purification et la solubilité ensemble. Bien que les expériences suggèrent que ce dernier n'est pas adapté à notre enquête, la stratégie simple de diminuer la taille de ce vecteur a entraîné une amélioration significative de la qualité de la bibliothèque. Le concept de co-expression de criblage est également introduite à la fin du chapitre, et, enfin, le dernier protocole d'ESPRIT peut être trouvé à la section Matériel et Méthodes.

Introduction

ESPRIT (Expression of Soluble Proteins by Random Incremental Truncations) is a directed evolution-based method for finding soluble protein domains in *E. coli* conceived in our laboratory (Hart and Tarendeau, 2006). The concept is to perform a fast, reliable and comprehensive screen encompassing all possible constructs of a gene of interest thereby removing the variable of construct design. In common with other directed evolution approaches, ESPRIT consists of a mutagenesis step and a screen.

Random DNA truncation of the target gene

Since the enzyme-mediated incremental truncation mutagenesis steps are performed on a target gene in the final expression vector, it results in oriented and efficient cloning with no observed bias in insert size following the one-part vector ligation (recircularisation) reaction, increasing the effectiveness of library construction. The protocol in use has evolved from principles defined over 20 years ago (Henikoff, 1990, 1984), and more recently adapted for ITCHY directed evolution experiments (Lutz*, et al.*, 2001). As a first step, a gene of interest is cloned in a plasmid designed for incremental truncation where one unique restriction site blocks and the other opens exonuclease truncation (Figure 2.1). Upon linearising the target plasmid with the corresponding restriction enzymes, is subjected to a time course exonuclease III digestion to generate incremental truncations in the 3′ to 5′ direction. In a second step, mung bean nuclease removes the remaining single strand and blunt-ends the extremities. The blunt end is further polished with a polymerase with proofreading activity such as *Pfu*. After ligation, a plasmid library of the truncated gene is obtained.

There are three possible truncation strategies: unidirectional truncations from either 5′ or 3′, and a combined protocol resulting in a bidirectional truncation library. In the case of 3′ or 5′

truncation, one end of the target gene is fixed as the native end or any other limit for which there is some level of confidence. In contrast, the bidirectional truncation approach does not rely on fixing termini and could explore the definition internally located gene fragments with the advantage over fragmentation protocols of maintaining the correct orientation of the insert.

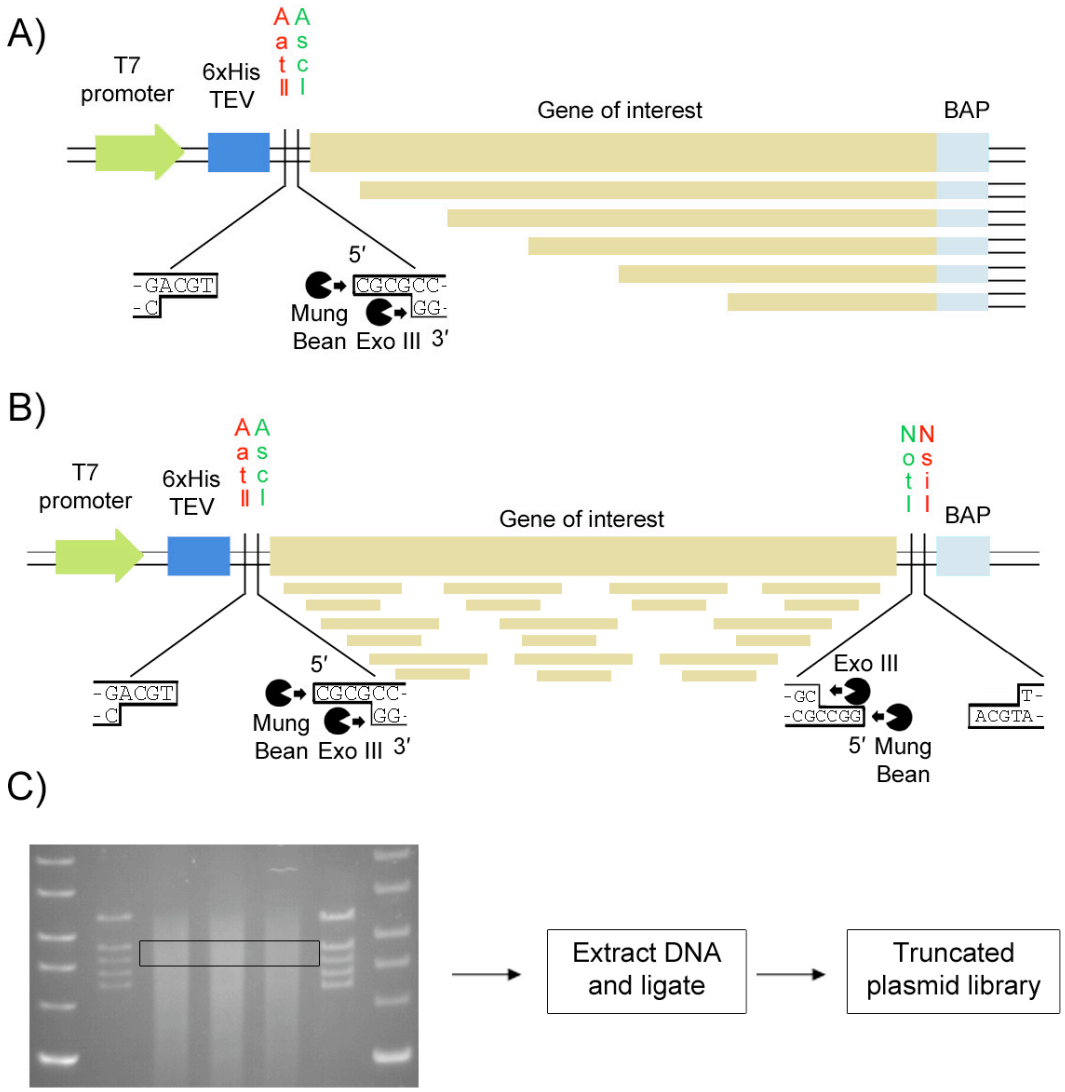


FIGURE 2.1: ESPRIT MUTAGENESIS STRATEGIES.

(A) unidirectional truncation is a good approach when the fixed end does not hinder solubility (e.g. 3′ truncation); (B) a bidirectional truncation approach consists of two sequential unidirectional truncations and allows identification of internally located soluble domains; (C) the truncated product is size fractionated by

agarose gel electrophoresis and the excised band of DNA is recircularised by ligation generating a plasmid library of the truncated gene;

Library choice

The choice of library type at the beginning of a project is an important first step due to two factors: fixing a terminus and efficient sampling of the genetic diversity. In the case of a unidirectional truncation strategy, there should be some confidence of the compatibility with expression of the fixed terminus. If incompatible, it is likely that the whole library will not yield a positive result.

Unidirectional truncation libraries have diversity equal to the length of the gene in nucleotides (N). One third of the constructs generated is in correct reading frame. During screening, usually more than the number of theoretical variants is taken, so called "oversampling." Regularly oversampling of ≥3N is tested to assure the full coverage for a given library.

The bidirectional truncation library approach permits sampling of both termini simultaneously, but the theoretical diversity of the library is quite large. It can be constrained by size fractionation by excision from agarose gel. The theoretical construct diversity for a bidirectional truncation can be calculated with the formula:

$$\sum_{m_1}^{m_2} (N - m + 1)$$

where, N is the length of the gene and m – the length of the fragments fractionated, both in nucleotides. The bidirectional library approach combines sequentially two unidirectional truncation reactions and therefore one in nine clones are in the correct reading frame with the two flanking peptide tags. This results in a level of diversity that cannot be covered comprehensively with our screening limit of 28,000 constructs. Typically up to five percent

of the theoretical construct diversity is tested; however, these libraries are still significantly smaller in size and better sampled than the point mutagenesis libraries of classical directed evolution.

The process of agarose gel size-fractionation of a subset of truncated constructs is useful since it allows higher resolution analysis of a gene region and thereby reduces the size of the library. In the early version of ESPRIT however, vectors with target inserts were large (7-8 kb) and suffered from imprecise size fractionation due to the poor resolving power of agarose gels for higher molecular weight DNA. Library quality is measured by colony PCR screening with flanking vector-specific primers; the number of amplified fragments within a particular size range is expressed as a fraction of the number of colonies tested.

Development of a high throughput screen for soluble expression

Direct screening of tens of thousands of variants for soluble expression by centrifugation or filtration would be difficult, expensive, time-consuming and probably not very effective. The strategy chosen for ESPRIT is that of indirect assessment of solubility by *in vivo* posttranslational modification of a short peptide tag (Fig. 2.2) filtering the huge numbers and enriching for putatively soluble hits to be confirmed in a second, direct assay by affinity purification; a two-step, orthogonal screen.

Orthogonal screen

Protein expression is induced in a high-density colony array format of up to 56,000 colonies, usually 28,000 in duplicate (Fig. 2.x), each colony containing a unique target gene insert. Colony blots are prepared by *in situ* lysis on nitrocellulose membranes and blots are hybridised simultaneously with a monoclonal antibody against the N-ter hexahistidine tag, with detection by fluorescent secondary antibody, and fluorescent streptavidin to detect the biotinylation status of the C-ter biotin acceptor peptide. Constructs showing no hexahistidine

tag signal are discarded immediately, and the remainder ranked on the biotinylation signal. One consequence of using probes against both ends is that degraded constructs can be eliminated at this early stage.

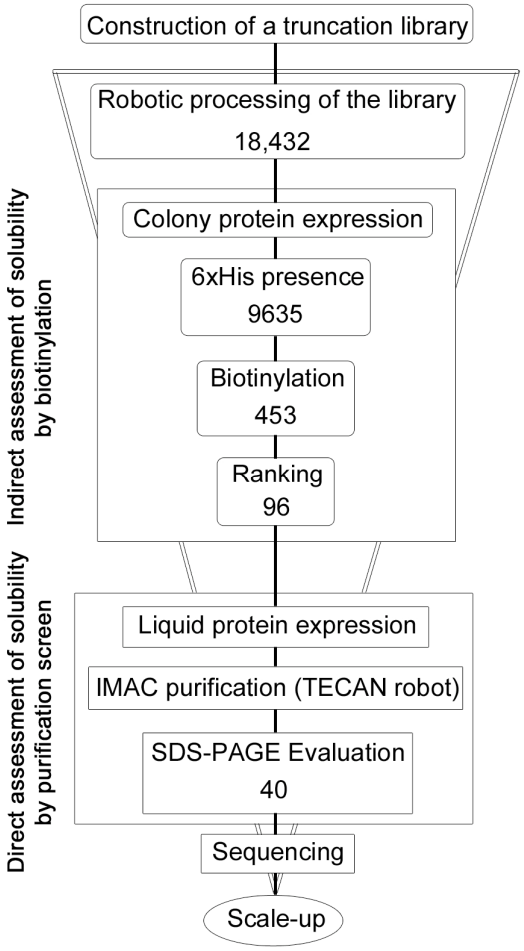


FIGURE 2.2: TWO STEP ORTHOGONAL SCREEN WORKFLOW ADOPTED BY ESPRIT.
After truncation library is constructed, it is initially assessed indirectly for solubility by biotinylation. At the next step, direct assessment of solubility results in the selection of clones for sequence identification and scale-up. The figures are taken from a project in the lab and represent the decrease of the number of constructs tested after each step.

Thus, the first step of the screen relies strongly on an *E. coli* post-translational modification where, in the natural situation, the cytoplasmic BirA enzyme ligates biotin (vitamin H) covalently to the biotin carboxyl carrier protein (BCCP) with exceptional specificity

(Chapman-Smith and Cronan, 1999). BCCP is a small folded protein that could be used as a reporter protein but with the potential issue of enhancing solubility due to the passenger solubilisation effect commonly observed with fusion proteins (Reich*, et al.*, 2006). There is however an artificial 14 aa peptide substrate for BirA that mimics the BCCP recognition site despite having no sequence similarity (Beckett*, et al.*, 1999,Schatz, 1993). This biotin acceptor peptide (BAP; commercialised as the Avitag) is a well-established tool for protein immobilisation onto streptavidin-coated surfaces (beads, plates etc.). It is efficiently labelled by the endogenous biotin ligase in both prokaryotes and eukaryotes (Duffy*, et al.*, 1998). For our screening strategy, a major advantage is that no obvious perturbation of solubility is observed on its fused partner (Waugh, 2005) meaning that the tag can be removed from library constructs prior to crystallisation trials without affecting the protein solubility status. Crystallisation with tags still in place is usually undesirable since it can perturb crystal packing.

Isolation of the most efficiently biotinylated constructs leads to an enrichment of putatively soluble clones that, in the second level of screening, are expressed in small-scale liquid format, lysed and purified on nickel affinity resin. Well-behaving soluble constructs are identified as visible by Coomassie stained SDS-PAGE and their sequence boundaries determined by DNA sequencing.

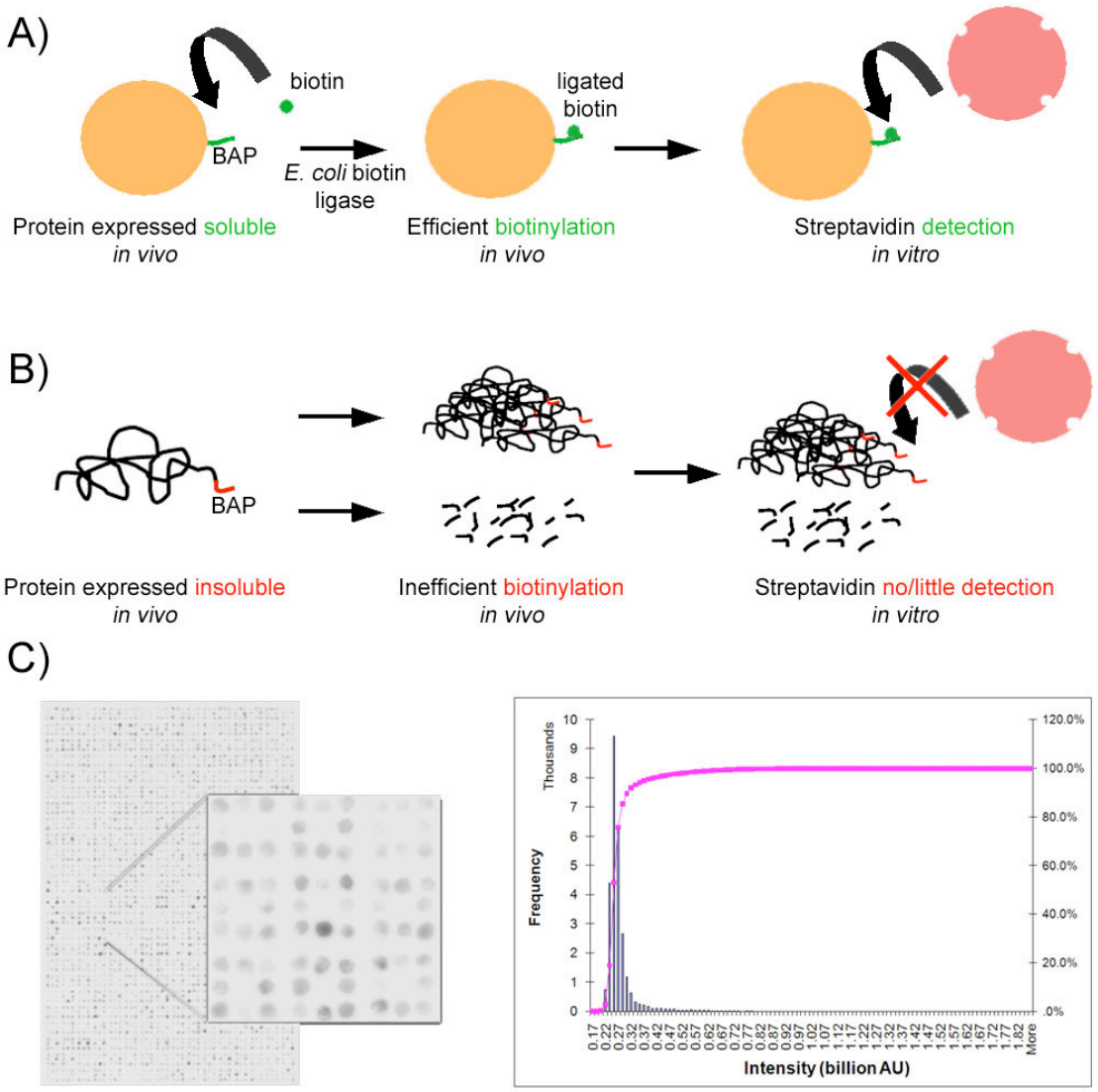


FIGURE 2.3: INDIRECT ASSESSMENT OF SOLUBILITY USING IN VIVO BIOTINYLATION EFFICIENCY AS A REPORTER.

(A) when soluble protein is expressed in vivo the endogenous biotin ligase attaches biotin to its biotin acceptor peptide (BAP), then, detection is done by streptavidin in vitro; (B) when insoluble protein is expressed in vivo the protein aggregates or is degraded resulting in less efficient biotinylation; (C) analysis of biotinylation levels on a colony blot macroarray (left) represented as a frequency histogram (right) where the majority of library colonies exhibit a signal close to background levels due to the endogenous BCCP. A minority of putatively soluble hits are more highly biotinylated;

The truncation strategy and the solubility screens are combined in a robotic process to handle efficiently the large number of clones. The automation is important at three steps: i) colony picking enables clonal separation after transformation of *E. coli* with the plasmid library, ii) gridding or arraying of inocula onto membranes to produce colony arrays for expression analysis, and iii) facilitating the affinity purification tests of the putatively soluble clones identified in the colony analysis (Fig. 2.3).

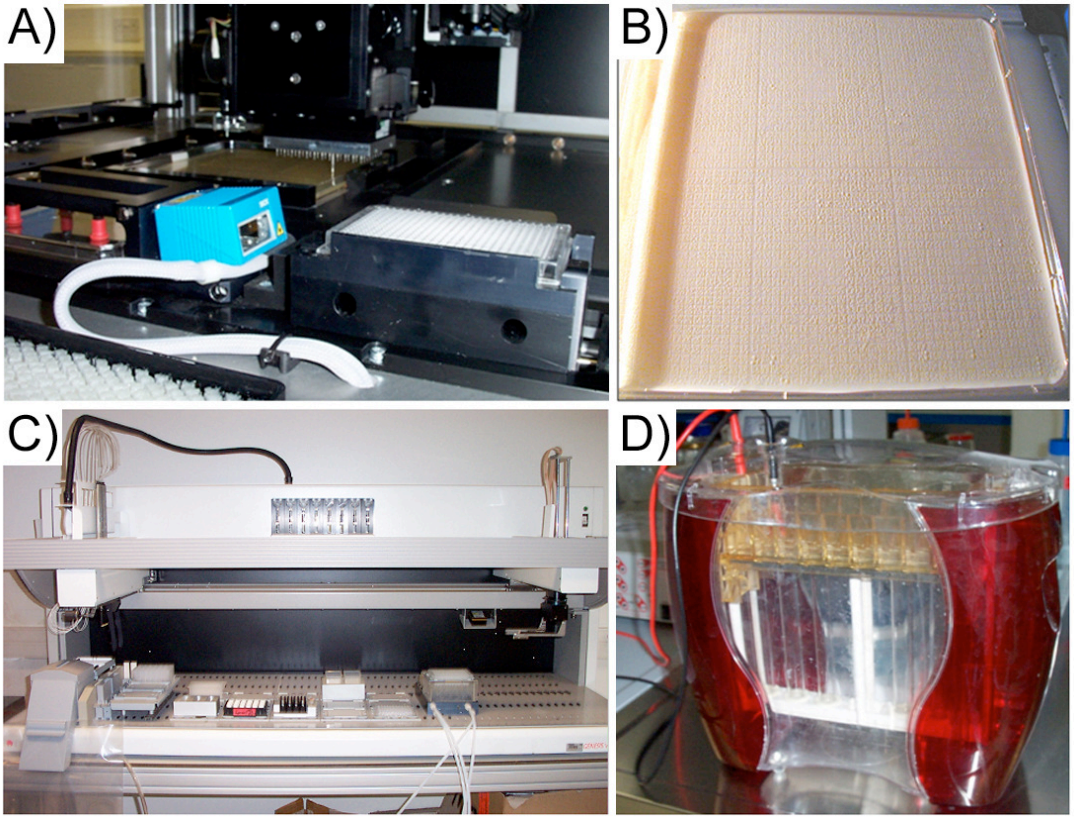


FIGURE 2.4: ESPRIT HIGH THROUGHPUT AUTOMATION

(A) picker gridder robot used for clone isolation and printing of colony macroarrays; (B) colony macroarray on Q-tray; (C) Tecan liquid handling robot used for 96 well Ni2+NTA purification screens; (D) Ruby electrophoresis system for high throughput SDS-PAGE and western blot;

Protein expression in *E. coli* can be affected by the expression plasmid. Desirable are a strong but tightly repressed promoter and an origin of replication resulting in high copy number

(Baneyx, 2004). In an early version, ESPRIT relied on a C-ter biotin acceptor peptide, no hexahistidine tag, and a medium strength Ptac promoter (pMAL-c2g derived vector, NEB). In this initial format, the concept was tested and soluble protein fragments obtained (Tarendeau, *et al.*, 2007). However, the expression was leaky and the absence of purification tag meant subcloning was required before the Ni^{2+}NTA purification screen. The next generation of vector (pTAR007) contained the strong and tightly regulated T7 promoter (from pET-9a) and a 6xHis-TEV tag allowing immediate purification testing. During that phase, the bidirectional truncation approach was conceived (Dr. F. Tarendeau) to screen for internal protein domains, leading to a significant increase of the diversity of the libraries. This introduced the requirement for size fractionation of libraries (*see* earlier discussion), but problems were then encountered due to the large vector size and poor resolving power of agarose gels at high molecular weight.

When I arrived in the lab in 2005, ESPRIT had been only applied twice with promising results (Tarendeau, *et al.*, 2007). My initial objective was the optimisation of the protocols including vector modifications to improve the efficiency of insert size fractionation and diversity coverage of bidirectional libraries. Results are presented and the current, optimised protocol for ESPRIT library generation and testing is in the methods section at the end of the chapter.

Results

ESPRIT optimisation

A general principle in directed evolution is that the higher the coverage of the diversity, the better the chance of success. When a solution is rare, the probability of success is lower. When multiple solutions can be found, comparisons can provide valuable information. If

screening capacity is limited (here to 30,000 constructs), library quality becomes an important parameter to optimise.

ESPRIT relies on enzymatic truncation of a vector and its subsequent size fractionation by agarose gel electrophoresis. As with chromatographic methods, agarose gels do not result in perfect separation but a distribution of fragment sizes at any given size point with a narrower range observed for smaller molecules. We reasoned that decreasing the size of the plasmid would improve size fractionation efficiency.

In an alternative approach addressing the library diversity, the combination of both solubility detection and purification peptides into a single N-terminal tag was explored. Such a vector could reduce the fraction of out-of-frame constructs in randomly truncated gene inserts. The 3′ truncation would be performed with no inefficiency because a 3-frame stop codon arrangement would ensure all clones were in frame. 5′ unidirectional and bidirectional truncation libraries would result in only 3-fold inefficiency due to out-of-frame clones. Finally, the entire protein tag was designed to be TEV cleavable, thus avoiding any subcloning or mutagenesis to remove the biotin acceptor peptide.

Vector Optimization

Essential elements of a vector for *E. coli* protein expression are the origin of replication, antibiotic resistance marker, and protein expression cassette with promoter and terminator (Brown, 2006). The original plasmid pYUM1001 is derived from pET-9a (Novagen) and is under the tight expression control of T7 promoter. Within the expression cassette there is a TEV-cleavable hexahistidine tag on the N-terminus (MGHHHHHHDYDIPTTENLYFQG; TEV site underlined) and C-terminal biotin acceptor peptide with linker (SNNGSGGGLNDIFEAQ**K**IEWHE; recognition sequence underlined, biotinylated lysine in

bold). Restriction site pairs enabling directional exonuclease III truncation 5′ *Aat*II / *Asc*I and 3′ *Nsi*I / *Not*I flank the gene of interest (Figure1c & 3).

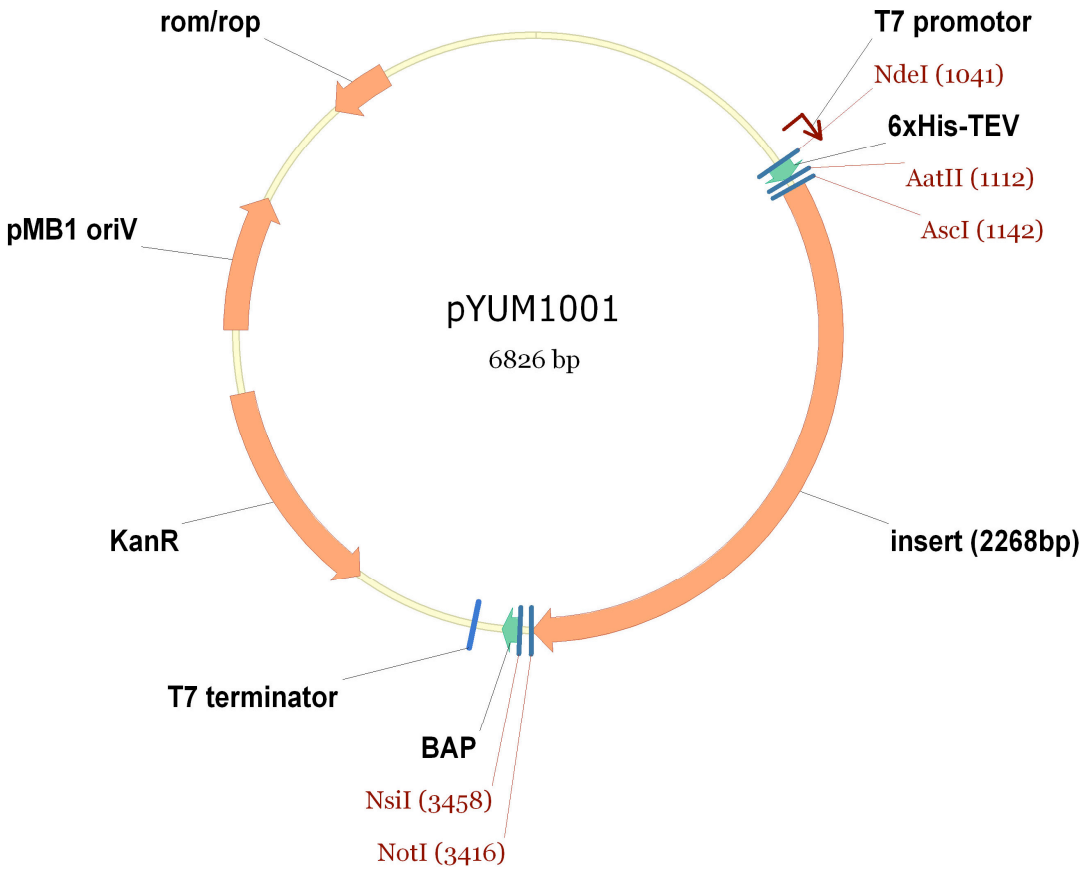


FIGURE 2.5: SECOND GENERATION ESPRIT VECTOR (PYUM1001) DERIVED FROM PET-9A.

(BAP-biotin acceptor peptide; KanR- kanamycin resistance marker)

The plasmid is derived from pBR322 (Bolivar, *et al.*, 1977) and contains the pMB1 origin of replication regulated to 20-30 copies/cell by the downregulator rom/rop protein (Swiss-Prot entry: A5PHA5) (Banner, *et al.*, 1987). Absence of rom/rop protein in the classical expression vector pUC18/19 in combination with mutation in the origin of replication results in a much higher copy number - 500-700 copies/cell (Yanisch-Perron, *et al.*, 1985). While this is one candidate for deletion, it would not affect much the size of the final vector. Therefore, BLAST-N was used to understand the unannotated regions of the vector. It was

found that there are two other reading frames, probably inherited from pBR322, encoding mob-like protein (Swiss-Prot entry: A5PHA4) and an incomplete tetracycline resistance protein (Swiss-Prot entry: P02981) (Figure4).

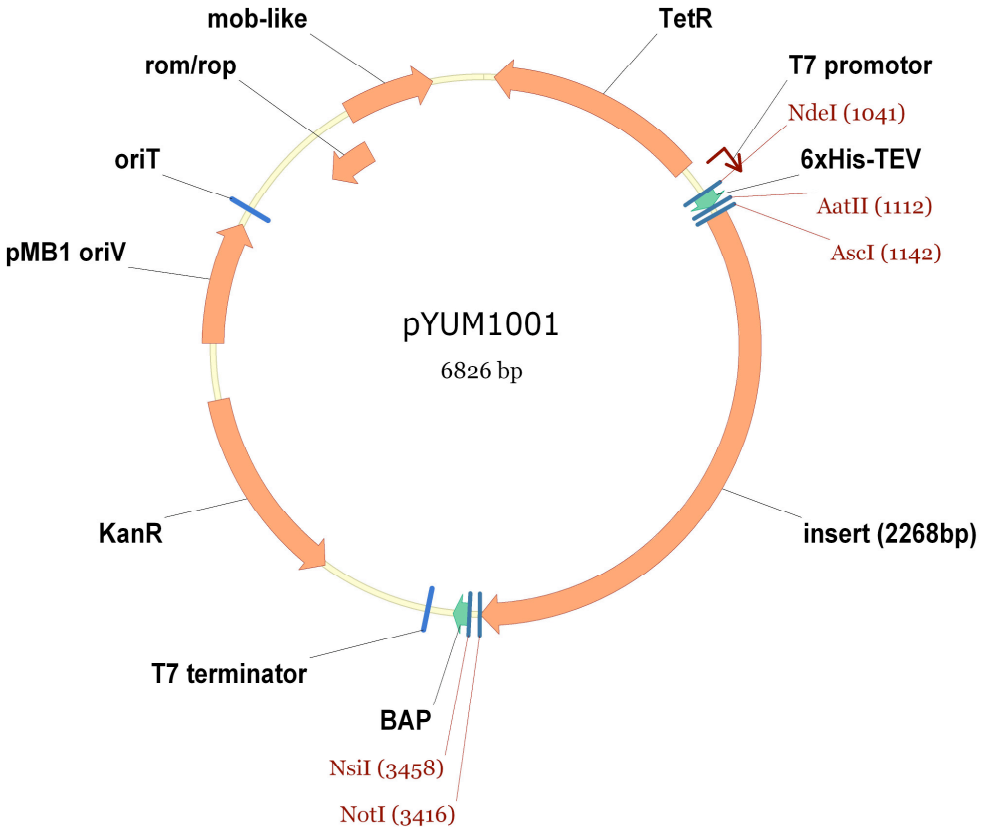


FIGURE 2.6: SECOND GENERATION ESPRIT VECTOR (PYUM1001) DERIVED FROM PET-9A WITH ASSIGNED GENES ENCODING MOB-LIKE PROTEIN AND INCOMPLETE TETRACYCLINE RESISTANCE IDENTIFIED BY BLAST

(BAP-biotin acceptor peptide; KanR- kanamycin resistance marker; TetR- tetracycline resistance marker)

The mob-like protein is a poorly characterized mobility protein that plays a role in plasmid conjugation and appears to be inherited along with origin of transfer (*oriT*) from pBR322. In a homogeneous plasmid sample there would not be any concerns, but the presence of these elements could theoretically lead to some shuffling of samples in a library by conjugation. Additionally, *oriT* is recognized by mobility proteins resulting in plasmid nicks and

triggering a relaxed state (Byrd and Matson, 1997). Nicked DNA is a substrate for exonuclease III and incremental truncation protocols require additional steps to remove nicked plasmids (Ostermeier and Lutz, 2003). Finally, the tetracycline resistance protein is inactive due to a 105 amino acid N-terminal deletion. We thus reasoned that removal of these plasmid regions would reduce the vector size and possibly provide additional advantages.

As a first step, the region directly after the origin of replication and preceding the T7 promoter was deleted removing 2105 bp. The appearance of colonies was a sign that the deletion was successful; the origin of replication undamaged and the plasmid successfully propagated. In a second step, a smaller sequence (296 bp) between the kanamycin resistance protein (aminoglycoside phosphotransferase) and the T7 terminator was removed. The result was a minimal plasmid, less than half the size of the parent vector (often even smaller than the inserts) capable of overexpressing proteins (Fig. 6). Plasmid prep yields are roughly two times higher and correlate with an increase in copy number.

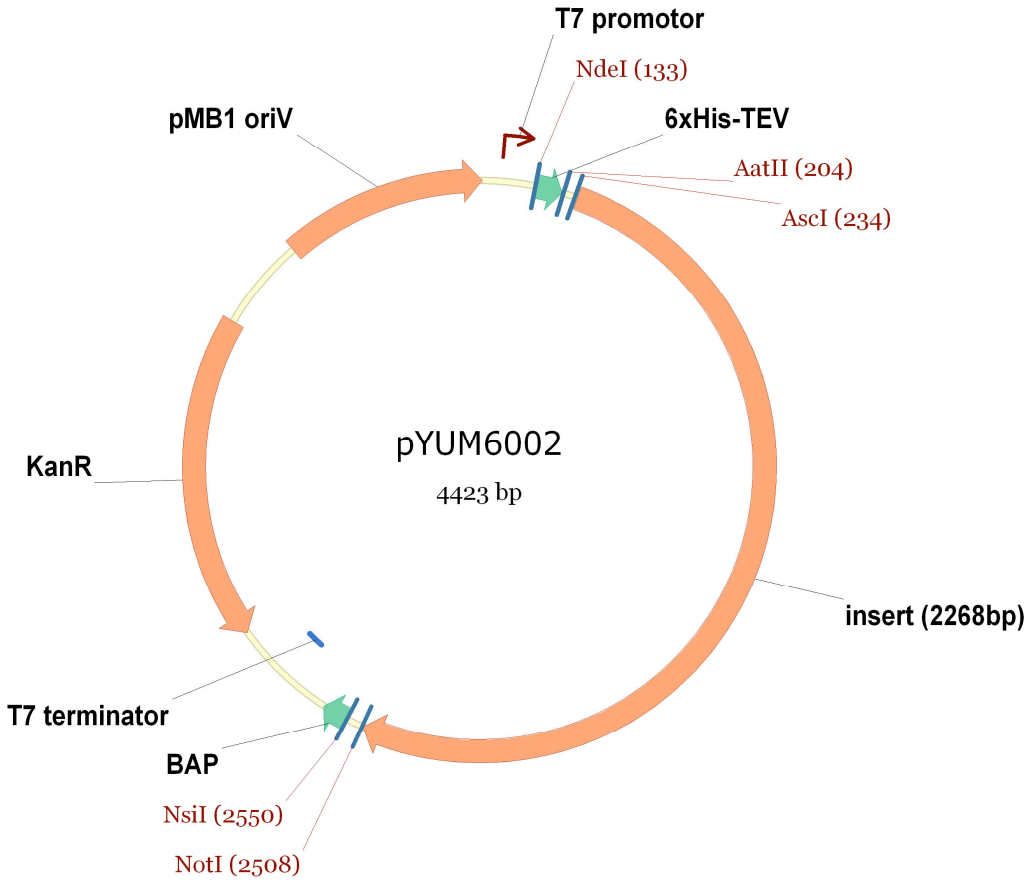


FIGURE 2.7: OPTIMISED ESPRIT EXPRESSION VECTOR (PYUM6002)
lacking regions between the pMB1 origin of replication and T7 promoter and between kanamycin resistance marker and T7 terminator.

The improvements in library quality were demonstrated by a comparison of the insert size distribution after size fractionation of two libraries: one generated with the larger vector (pYUM1001) and another with the smaller (pYUM6002). The first library was generated and run on agarose gel (0.5%) under optimal migration velocity (5V/cm). Subsequently, the library was fractionated into three size ranges corresponding to inserts encoding 0-230, 230-500, and 500-756 amino acid fragments. Many overtruncated fragments were observed in the subset of small fragments as well as absences of PCR products attributed to the deletion of

one of the primer-binding sites. Moreover, there was no clear separation for between the medium and large sub-libraries (Fig. 7).

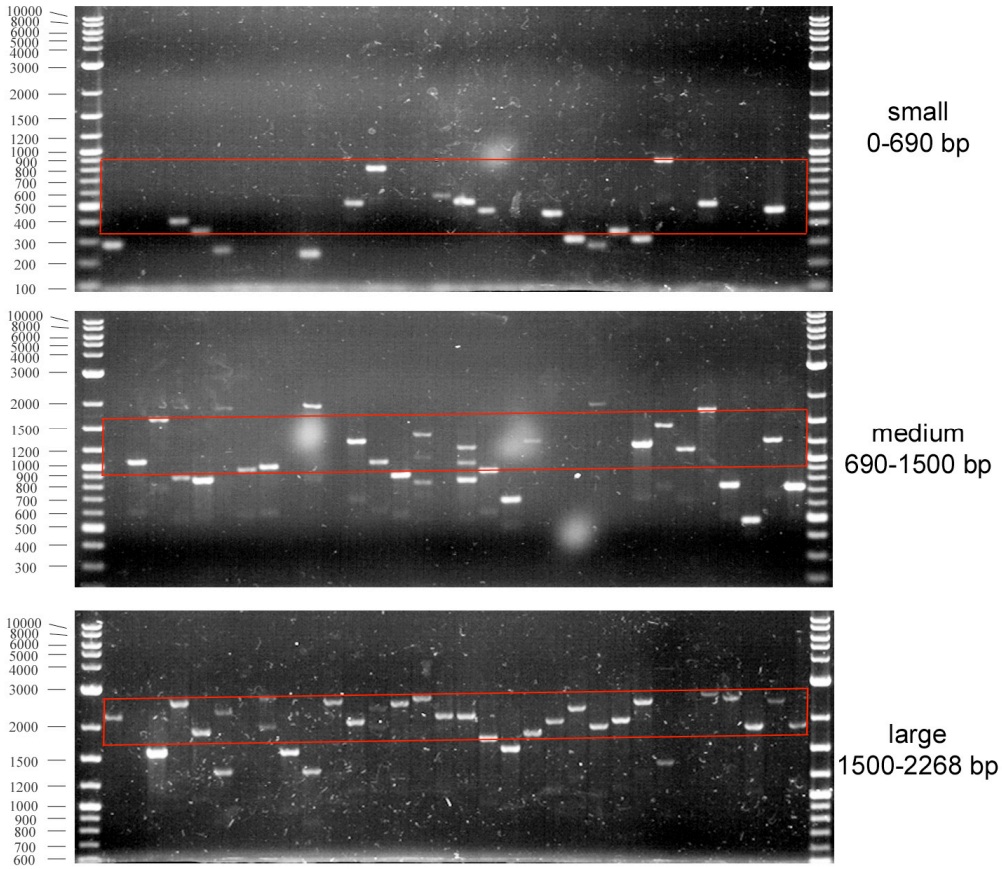


FIGURE 2.8: COLONY PCR SCREEN FOR THE EVALUATION OF IKKB 3′ TRUNCATION LIBRARY GENERATED WITH PYUM1001.

Target gene insert size is calculated as PCR product size minus 320 bp of flanking sequence.

A second library was generated and size fractionated under the same conditions into two sub-libraries encoding fragments 50-250 and 250-599 amino acids. Here, few clones appeared overtruncated and the size selection resulted in precise separation of the fragments with almost no overlap between sub-libraries (Fig. 8). Library statistics are presented in Table 1 with quality measured by determining the fraction of PCR products in the desired size range and contamination as the fraction of PCR fragments occurring outside of the desired range.

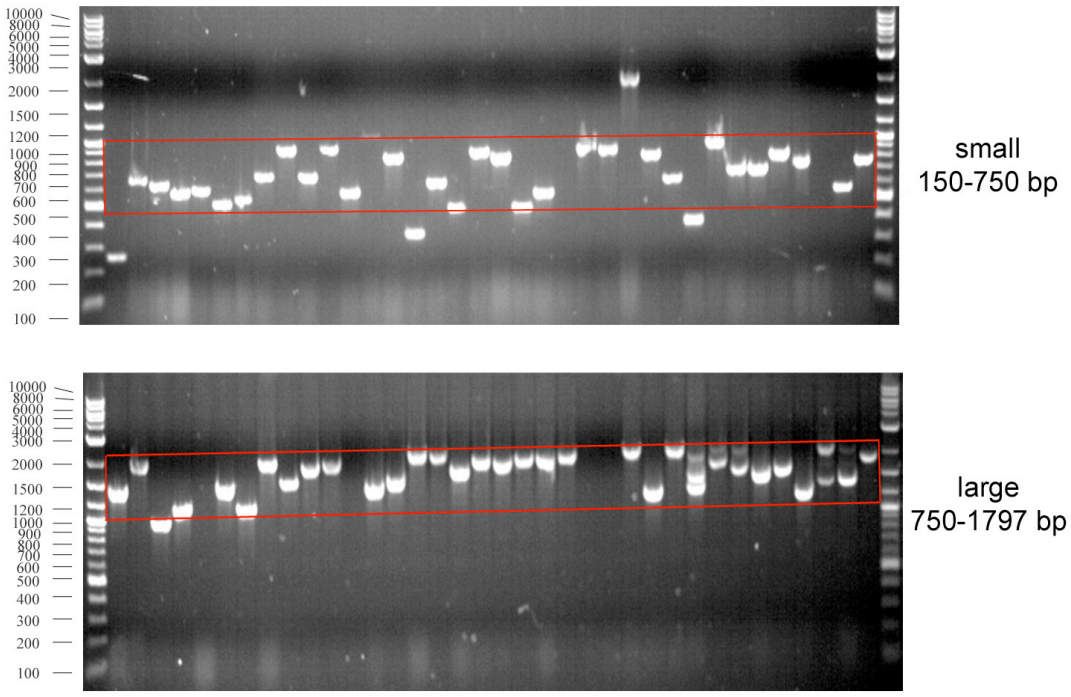


FIGURE 2.9: COLONY PCR SCREEN FOR THE EVALUATION OF mTOR 5′ TRUNCATION LIBRARY GENERATED WITH PYUM6002.

Target gene insert size is calculated as PCR product size minus 320 bp of flanking sequence.

	IKKβ short	IKKβ medium	IKKβ large	mTOR small	mTOR large
Total	56.3%	78.1%	90.6%	94.4%	88.9%
Quality	37.5%	40.6%	71.9%	83.3%	86.1%
Contamination	18.8%	37.5%	18.7%	8.3%	2.8%

TABLE 2.1: SUMMARY OF THE COLONY PCR SCREENS FROM FIG. 7 AND FIG. 8. TOTAL: NUMBER OF PCR PRODUCTS;

Quality: PCR products in the correct size range; Contamination: PCR products outside of the range.

A significant improvement in the size fractionation of truncation sub-libraries was observed due to the better resolving power of agarose gel on the smaller plasmid DNA molecules. A

reduced number of overtruncated genes were observed due to the negative selection from damage of the origin of replication or resistance marker.

In summary, the minimised plasmid was well adapted for high levels of protein expression and synthesis of truncation libraries. It replicated at a higher copy number, disabled plasmid conjugation mechanisms and a exhibited a negative selection for overtruncated variants. This vector has been used in all subsequent experiments in the laboratory.

N-terminal tag

Two studies have described N-terminal protein fusions that combine hexahistidine affinity purification and *in vivo* biotinylation. Tagwerker *et al* have reported the use of a hexahistidine-BCCP fusion (Tagwerker*, et al.*, 2006). Their tag was TEV cleavable and designed for TAP purifications under denaturing conditions. While this tag was efficiently biotinylated with reported rates close to 100%, it is less compatible with a screen for solubility such as ESPRIT due to its globular form and possible passenger solubilisation effect. Another report describes a more compact tag, using the 15-amino acid AviTag with long linkers separating it from hexahistidine and TEV site sequences (Scholle*, et al.*, 2004). They reported inefficient, protein dependent biotinylation ranging from 30-90%.

Here a minimal N-ter tag was synthesised and tested 6xHis-BAP-TEV (MKHHHHHHGLNDIFEAQKIEWH**ENLYFQG**; biotinylation sequence underlined, TEV site in bold) for its compatibility with ESPRIT. The desired phenotype was efficient biotinylation that correlated with protein solubility, with no signal when expressed out of frame with the fusion partner.

An experiment was performed to test the fusion of the tag with maltose binding protein, known for its high solubility. The construct in- and out-of-frame was expressed from the Ptac

promoter in pMAL-c2g in a first generation ESPRIT vector (data not shown). As expected, no biotinylation was observed on the tag when unfused, probably due to poor stability of the peptide the absence of a globular partner. When fused to MBP, the fusion was well expressed and soluble, but only weakly biotinylated. In order to confirm the result and compare the biotinylation under the stronger T7 promoter, the inserts were subcloned in pYUM1001 and tested for their biotinylation (Fig. 9).

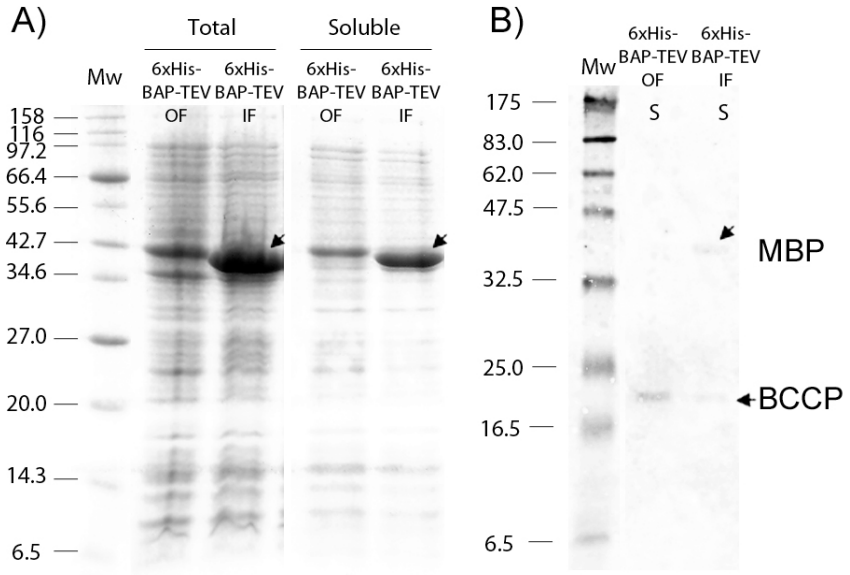


FIGURE 2.10: EVALUATION OF N-TERMINAL 6XHIS-BAP-TEV FUSED TO MBP IN PYUM1001

(A) SDS-PAGE of soluble fractions for the tagged MBP normalized for protein loading (B) western blot on the soluble fractions of the in- and out-of-frame fusions. Endogenous biotinylated BCCP is visible (OF: out-of-frame with mbp gene; IF: in-frame with mbp gene ; S: soluble fraction from panel A.

Again, MBP was clearly expressed and soluble as assessed by Coomassie stained SDS-PAGE. A streptavidin blot revealed low biotinylation levels: Only 70 ng of the 1.4 µg MBP loaded was biotinylated (*see* western blot quantification section in general methods and protocols, appendix) corresponding to 5% biotinylation efficiency. Therefore the tag developed here would not be useful for quantitative ranking of soluble proteins so the

potential advantages afforded by an N-ter double tag combined with a 3 frame stop codon sequence were not realised.

Coexpression vector compatible with library screening

Adding a second coexpression plasmid into the ESPRIT system would allow library screening to be applied to certain protein complexes. The use of a coexpressed bait would potentially allow trapping of complexes that require binding, or co-folding, for stability. Secondly, and used later in this work when analysing the DAPK1-Ca^{2+}/calmodulin work, deletion of one component of a complex allows mapping of the minimal interacting region.

Two types of library experiment could be envisaged: i) screening of a truncation library against an unvarying bait protein, and ii) screening of one library against a second. The first is completely feasible being an extension of the current screening procedure, but with the addition of the bait vector into the competent cells prior to transformation of the library. Upon Ni^{2+}NTA purification screening for the library target, the presence or absence of the bound partner could be assessed. The second library format is more problematic since the diversity present in the first library multiplied by that in the second would mean the clone collection is too large to sample effectively given the 30,000 clone limitation of the screening robotics in its current form.

There are two important considerations for building such a system. Firstly, the plasmids should coexist stably in *E. coli*. This property is determined by the origin of replication and it has long been known that pMB1 origin of replication is fully compatible with p15A (Chang and Cohen, 1978). A common plasmid with such origin is pACYC184. A derivative of this plasmid expressing rare tRNAs is commonly used for the expression of human genes in *E. coli*. It is a low copy plasmid (10-15 copies/cell) and the version used here had been further modified by SGC Toronto to introduce a T7 expression cassette allowing protein expression

along with supplementation of the rare tRNAs (Fig. 10). The plasmid was modified further to add an N-ter FLAG tag (and cloning sites) for identification of the co-expressed binding partner following hexahistidine purification of the target. Use of this plasmid in coexpression experiments with the pET-derived ESPRIT vector is presented in Chapter 3 using with a known interaction system DAPK1 and Ca^{2+}/calmodulin.

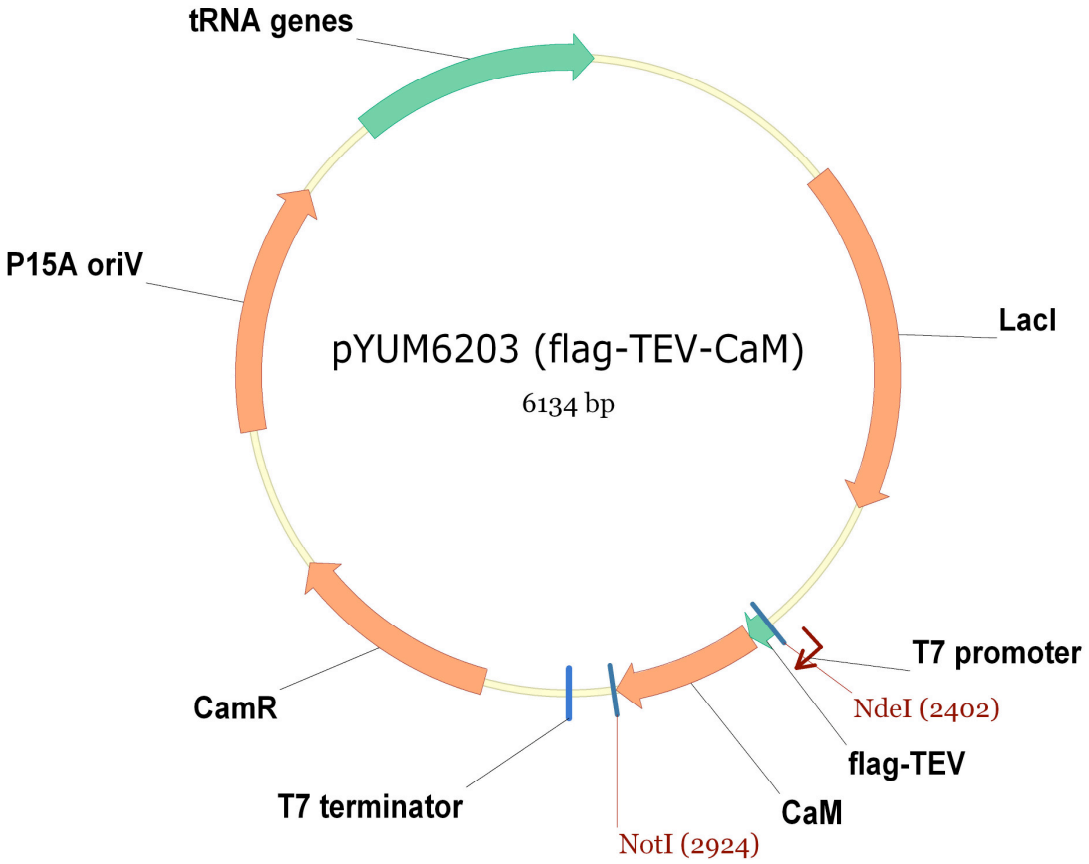


FIGURE 2.11: COEXPRESSION PLASMID PYUM6002 DERIVED FROM PACYC184 WITH ENCODING RARE TRNAS AND T7 EXPRESSION CASSETTE WITH SYNTHETIC, CODON-OPTIMISED CALMODULIN (CAM) GENE.

(LacI-lac repressor; CamR-chloramphenicol resistance marker)

Discussion & Conclusions

Initial experiments were underway using a basic form of the ESPRIT screen at the start of this work, from which limitations had been observed. Specifically, in order to assure efficient screening of the diversity present in our libraries, a better quality library was required, with fewer overtruncated fragments and more accurate size-fractionation of inserts. Two separate approaches were explored with this aim; firstly, minimizing the plasmid size made significant improvements to the accuracy of agarose gel size selection of the truncation libraries since DNA was migrating in a region of the gel with better resolving power. This plasmid has formed the basis for all further ESPRIT work. Secondly, the 3-fold efficiency improvement afforded by combining an N-terminal tag permitting both purification and solubility assay with a C-ter three frame stop codon arrangement was less successful since biotinylation of the acceptor peptide was inefficient in the context of flanking hexahistidine tag and TEV cleavage site. This idea was discontinued once the advantage of separate N-ter and C-ter tags for eliminating degraded constructs at the colony analysis step was discovered.

Finally, a coexpression vector was added to the system expanding the scope of possible experiments in the direction of protein complexes. Relative to kinases, it would also be possible to provide a substrate protein, or MBP-conjugated substrate peptide, to screen directly in *E. coli* for the activity of the kinase constructs using an anti-phosphoserine/tyrosine antibody to probe the colony blot. Experiments have been started, but results are preliminary and will not be presented here.

Materials & Methods

Open Reading Frames

The open reading frame corresponding to the following proteins was used: full-length wt MALE_ECOLI (Swiss-Prot Accession Number: P0AEX9) as in pMAL-c2g (NEB).

Construction of Minimized Vector

At a first step 10 ng pYUM1001 was PCR amplified by *Pwo* polymerase with 400 nM of mESPRITfor2 (5′ GCTAG TCATG GGTAC CAGAT CTCGA TCCCG CGAAA TTAAT ACG 3′) and mESPRITrev2 (5′ GCGAT CGATC GGTAC CAAAG GCCAG CAAAA GGCCA GGAAC 3′) in 50 µl reaction for 20 cycles of T_m=57 °C and 4 min of elongation at 72 °C. The linear vector was digested at terminal *Kpn*I sites recircularised by using a rapid ligation kit (Roche) and used to transform Mach1 *E. coli* cells. Plasmid DNA was extracted and PCR amplified with *Pwo* with 400 nM of mESPRITfor1 (5′ GTACT AGACC GATAT CCCTG ACTGC GTTAG CAATT TAACT GTGAT AAAC 3′) and mESPRITrev1 (5′ GTCCT GTGGA TATCC GGATA TAGTT CCTCC 3′) in 50 µl reaction as above. The PCR product was digested, religated and eventually the plasmid verified by sequencing.

N-Terminal Tag Cloning

Two oligos comprise the tag: H6BTfor1 (5′ TATGA AACAC CATCA CCACC ACGGT CTGAA CGACA TCTTC GAGGC GCAGA AAATC GAATG GCACG AAAAT CTGTA TTTTC AGGGG CCCGC CTGAT CAAGC GGCCG 3′) and H6BTrev1 (5′ TAGCG GCCGC TTGAT CAGGC CGGCC CCTGA AAATA CAGAT TT TC ATGCC ATTCG ATTGG CTGCG CCTCG AAGAT GTCGT TCAG ACCGT GGTGG TGATG ATGGT GTTTC 3′). Firstly, 2.5 µM of each was phosphorylated with 5 U of PNK for 1 h at 37 °C in 20 µl total reaction. Then, they were annealed by reducing the temperature from 95 °C to 24 °C at a rate of 3 °C per min. 3 µg of pMAL-c2g were cut with *Nde*I and dephosphorylated with shrimp alkaline phosphatase (SAP). Then, the annealed oligos and vector were ligated generating vector pHAR1301 (out-of-frame N-terminal tag). PCR on the ligation product generated the in-frame vector pHAR1302. Primers used were scrtagrev1 (5′ CCCCT GAAAA TACAG ATTTT CGTGC CATTC GATTT TCT 3′) and mbpfor1 (5′ AAAAC TGAAG AAGGT AAACT GGTAA TCTGG ATTAA CGGCG A 3′) in 30 µl final reaction carried out with annealing temperature of 66 °C and 30 cycles of amplification with 20 min elongation time at 72 °C. As above, the vector was ligated, recovered in *E. coli* and verified by sequencing.

Protein Expression Test

C-terminal and N-terminal in and out of frame tag fusions to MBP constructs were transformed in BL21-RIL cells. 50 ml terrific broth (TB) cultures were supplemented with chloramphenicol (30 µg/ml) and ampicillin (50 µg/ml) and grown at 37 °C to OD_{600} 0.6. Then, they were supplemented with biotin (50 µM) and induced with IPTG (100 µM) for 3 h, 37 °C. Pellets were resuspended in buffer at pH7 (200 mM NaCl, 20 mM Tris HCl, 10 mM β-mercaptoethanol, Complete protease inhibitors (Roche) and Benzonase (Roche), and lysed by sonication. Soluble protein was obtained by centrifugation at 15,550 g for 20 min at 4 °C. Total and soluble fractions were then boiled in Laemmli buffer prior to analysis by SDS-PAGE.

ESPRIT Protocol

Library Generation

Plasmid was extracted from 200 ml overnight *E. coli* culture by classical alkaline lysis and phenol chloroform protocols followed by ammonium acetate precipitation (Sambrook, 2001) and further extracted using a plasmid miniprep kit (Qiagen). 10 µg of supercoiled plasmid were cut to completion with library specific restriction enzymes in 200 µl total reaction volume. 4 µg of linearised plasmid was truncated with 400 U of Exo III (NEB) at 22 °C in 120 µl total volume. Truncation rate was adjusted to the DNA size using NaCl concentration calculated: N = (log(δ/47.9))/-0.0064, where N=NaCl concentration (mM), and δ is bp/min. Each min over 1 h, 2 µl of the reaction was quenched in 200 µl of 3 M NaCl at 70 °C. At the end of the truncation, DNA purified with Nucleospin Extract (MachereyNagel) and DNA eluted in 35 µl. Mung bean nuclease (MBN) treatment followed using 5 U of enzyme, 30 ºC, 30 min. Finally, blunt ends were polished with *Pfu* (Stratagene) treatment at 72 °C for 20 min with 12.5 mM each dNTP. Libraries were electrophoresed at 5 V/cm on EtBr-containing 0.5% w/v agarose gels at 4 °C gel slices excised. DNA was extracted from the gel using QiaexII (Qiagen) and Mach1 cells transformed (Invitrogen). For unidirectional libraries, ≥3 N colonies were scraped from plates and pooled, then plasmid extracted. For bidirectional truncation, after pooling, DNA was extracted by midiprep (Qiagen) and submitted to a second truncation reaction from other side. Purified DNA was prepared for all of the libraries and either arabinose or IPTG inducible BL21 (DE3) codon+ (RIL) (Invitrogen) *E. coli* strains transformed.

Qtray Preparation

For colony picking, Qtrays (X6023, Genetix) were prepared with 300 ml of LB-agar supplemented with kanamycin (50 µg/ml) and chloramphenicol (50 µg/ml) poured on a horizontal surface under a microbiological hood. Protein induction plates were prepared identically but supplemented with 50 µM biotin and inducer (either 0.1 mM IPTG or 0.2% w/v arabinose).

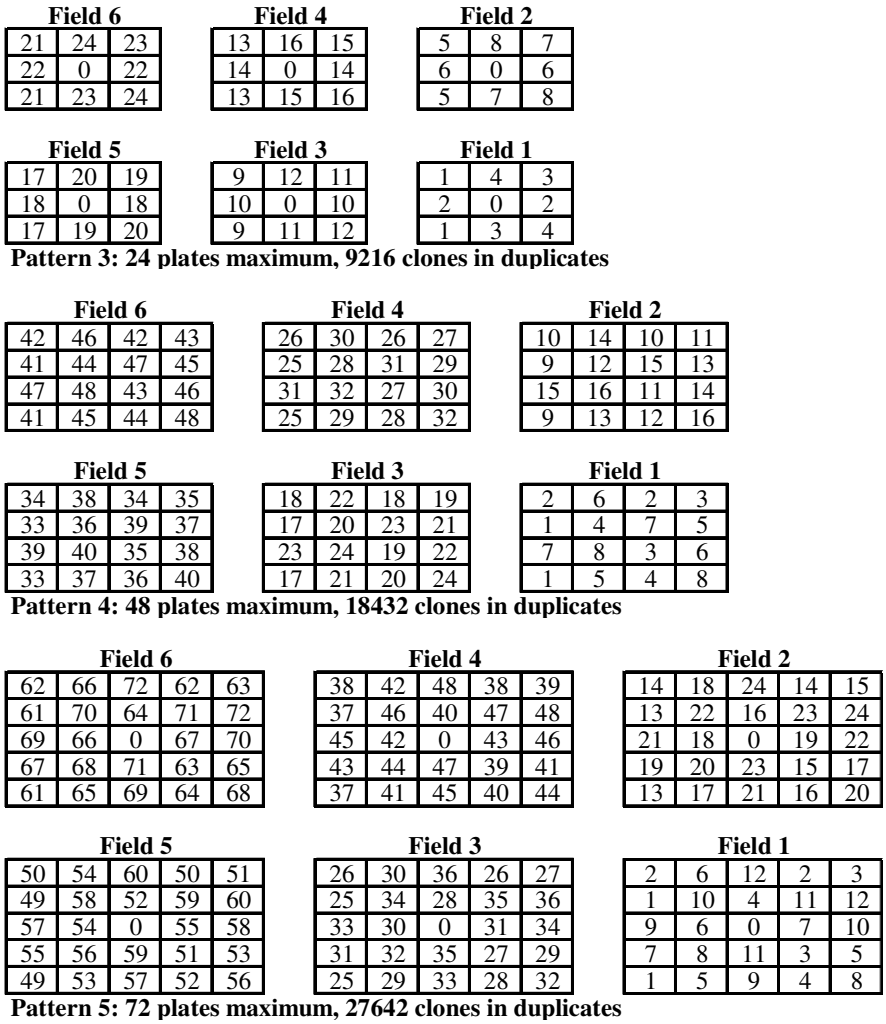
Colony Picking with Picker Gridder

Transformants were spread using glass beads on Qtrays in a dilution series to obtain a density of approximately 4000 colonies per plate. Then, a picker gridder robot (KBiosystems) was used to isolate clones into 384 well plates (X7007, Genetix). Cultures were grown overnight at 37 °C in a HiGro incubator (GeneMachines).

Gridding and Macroarray Preparation

After replication of the library in fresh TB with antibiotics minus glycerol using 384-well disposable replicators under the microbiological hood, it was printed onto nitrocellulose membrane (Amersham) in a pattern according to the number (Fig. 1.12). Colonies were grown overnight at room temperature and nitrocellulose membranes moved to prewarmed inducer-containing agar at 30 °C. Induction of protein expression was 4 h, 30 °C. *In situ* lysis was performed by moving membranes onto filter paper soaked with denaturing solution (500

mM NaOH, 1.5 M NaCl) for 10 min. Two cycles of neutralization on filter paper soaked with neutralization solution pH 7.5 (1 M Tris HCl, 1.5 M NaCl) were performed for 5 min each. Finally, the nitrocellulose arrays were immersed in 2× SSC solution pH 7 (300 mM NaCl, 35 mM $Na_3C_6H_5O_7$) for 15 min and cell debris removed by gentle scraping.



Field 6		
21	24	23
22	0	22
21	23	24

Field 4		
13	16	15
14	0	14
13	15	16

Field 2		
5	8	7
6	0	6
5	7	8

Field 5		
17	20	19
18	0	18
17	19	20

Field 3		
9	12	11
10	0	10
9	11	12

Field 1		
1	4	3
2	0	2
1	3	4

Pattern 3: 24 plates maximum, 9216 clones in duplicates

Field 6			
42	46	42	43
41	44	47	45
47	48	43	46
41	45	44	48

Field 4			
26	30	26	27
25	28	31	29
31	32	27	30
25	29	28	32

Field 2			
10	14	10	11
9	12	15	13
15	16	11	14
9	13	12	16

Field 5			
34	38	34	35
33	36	39	37
39	40	35	38
33	37	36	40

Field 3			
18	22	18	19
17	20	23	21
23	24	19	22
17	21	20	24

Field 1			
2	6	2	3
1	4	7	5
7	8	3	6
1	5	4	8

Pattern 4: 48 plates maximum, 18432 clones in duplicates

Field 6				
62	66	72	62	63
61	70	64	71	72
69	66	0	67	70
67	68	71	63	65
61	65	69	64	68

Field 4				
38	42	48	38	39
37	46	40	47	48
45	42	0	43	46
43	44	47	39	41
37	41	45	40	44

Field 2				
14	18	24	14	15
13	22	16	23	24
21	18	0	19	22
19	20	23	15	17
13	17	21	16	20

Field 5				
50	54	60	50	51
49	58	52	59	60
57	54	0	55	58
55	56	59	51	53
49	53	57	52	56

Field 3				
26	30	36	26	27
25	34	28	35	36
33	30	0	31	34
31	32	35	27	29
25	29	33	28	32

Field 1				
2	6	12	2	3
1	10	4	11	12
9	6	0	7	10
7	8	11	3	5
1	5	9	4	8

Pattern 5: 72 plates maximum, 27642 clones in duplicates

FIGURE 2.12: GRIDDING PATTERNS FOR COLONY ARRAYING.

Immunofluorescent Detection

The activated membrane was washed with water and 0.1% Tween-PBS for 5 min, 4 °C and blocked overnight with SuperBlock (Pierce) or with 0.1% Tween-PBS Milk 0.5%. It was then washed 3× 15 min with 0.1% Tween-PBS, then incubated with 1/3125 dilution in 0.1% Tween-PBS of mouse anti-his-tag primary antibody (ref, manufact) for 2 h, 4 °C in a roller blot oven (Techne). The antibody was drained off and washed 3× 15 min with 0.1% Tween-PBS The secondary antibody was rabbit anti-mouse Alexa532 (Amersham) 1/1000 dilution, and 1/5000 dilution of streptavidin conjugated with Alexa488 (Amersham) in 0.1% Tween-

PBS for 1 h, 4 °C. Membranes were then washed 3× 15 min with 0.1% Tween-PBS, then distilled water.

Typhoon Scanning and Data Analysis

A Typoon scanner (63005587, GE Healthcare) was used for scanning the membranes with 488 nm laser and 522BP emission filter (streptavidin). Then, a scan was performed with 532 nm laser and 555BP emission filter (6xHis tag). Data was saved in .gel format for subsequent analysis using VisualGrid software (GPC Biotech) to extract pixel intensity to Excel (Microsoft). Clones with no visible hexahistidine signal were eliminated and the remainder ranked upon biotin signal to generate a list of clones for expression and purification tests.

High Throughput E. coli Expression Testing

A fresh starter culture of 1 ml in a 96 deep-well block (Qiagen) was grown overnight at 37 °C 220 rpm in the HiGro. Expression cultures prepared as 4 ml TB in 6 deepwell blocks (Qiagen) and inoculated with a 1/100 dilution of inocula. Cultures were grown at 37 °C to OD_{600} 0.4 and induced by addition to IPTG at 25°C for 14-20 h. Cells were harvested by centrifugation at 3700 rpm for 10 min in a swinging rotor. Cell pellets were resuspended with 4 ml (20 mM Tris pH 8, 250 mM NaCl, 20% sucrose and 1 mg/ml lysosyme) for 15 min and sphaeroplasts pelleted at 3700 rpm, 15 min. The supernatant was removed and pellets kept at –80 °C prior to resupension in 200 µl (10 mM Tris pH 8, 0,5% Brij58, 1/1000 dilution of benzonase, 5 mM β-ME and protease inhibitor cocktail without EDTA). They were vortexed for 25 min at 4 °C and 460 µl buffer (300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 pH 7,5 mM imidazole and 5 mM β-ME) to lyse. Protein purification was performed with 60 µl Ni^{2+}NTA agarose (Qiagen) in 96 well filter plates according to a published protocol (Shih, *et al.*, 2002), modified to include a 30 min binding step. Wash buffer was 300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 pH 7,5 mM imidazole and 5 mM β-ME and elution buffer the same but supplemented with 300 mM imidazole. Total, soluble and purified protein fractions were mixed with Laemmli buffer and boiled for 5 min.

"*All you have to do is know where you're going. The answers will come to you of their own accord.*" --- Earl Nightingale

Chapter 3: DAPK1

Résumé

DAPK1 est une protéine kinase dépendante de Ca2 + / calmoduline. C'est une protéine grande, multidomaine, qui joue un rôle dans l'apoptose cellulaire. La structure de son domaine catalytique a été résolue en 2001, mais encore de nombreux aspects de sa régulation sont pas révélées. Une étude publiée sur cette protéine démontre d'identification, du premier connu, substrat peptide pour le groupe des protéines kinases. Comme un système mieux étudié, cette protéine a été ciblée pour démontrer l'applicabilité de l'ESPRIT sur la classe des protéines kinases et à tenter d'améliorer la compréhension actuelle de son règlement d'activité. Une bibliothèque de troncation 3′ a était effectués sur une construction proche de sa deuxième isoforme. Un certain nombre de variant solubles qui concentrent autour de la taille attendue pour un domaine catalytique kinase sont trouvé. Lorsque séquencés, tous les constructions sont discernées d'être plus long que celle caractérisait auparavant. La protéine est monodisperses, et un test sur sa stabilité thermique révèle des disparités significatives sur de petites variations de son C-terminal. Deux constructions, objet de nombreux essais, cristallisent et un modèle de structure préliminaire révèlent que l'extension C-terminale se replie sur le domaine précédemment prévu par une homologie de séquence de prédiction de structure. En outre, la protéine est cristallisée dans un nouvel groupe spatial facilitant nos connaissances actuelles pour les conformations de la protéine. La structure montre un réarrangement de trois régions du domaine kinase loin de la poche de liaison de l'ATP. D'autres investigations sont nécessaires pour une meilleure caractérisation de ces derniers. Une étude complémentaire sur la coexpression avec son partenaire contraignant - calmoduline, révèle une interaction minimale liaison non définies auparavant. Cette dernière est une démonstration de l'approche de coexpression décrite précédemment.

Introduction

Ca^{2+} is among the most important secondary messengers in the cell. It is present at concentrations of about 10^{-7} M and is involved in a multitude of regulatory processes ranging from muscle contraction to apoptosis. Its actions are often mediated by calcium binding proteins such as calmodulin (CaM), which has four binding sites with micromolar affinity for the ion. Upon increase of the default concentration of Ca^{2+} in the cytoplasm, calmodulin binds to it, changes conformation, and interacts with an array of proteins (Berridge, 2008).

Death-associate protein kinase 1 (DAPK1) is a 160kDa multidomain protein that responds to the change of calcium concentration mediated by calmodulin binding in a similar fashion to the canonical Ca^{2+}/calmodulin-regulated kinases (CaMK). It is part of a group of protein kinases that consists of two very closely related members: DRP-1 and ZIP (Inbal*, et al.*, 2000,Kawai*, et al.*, 1998) and two others more distinct: DRAK-1 and DRAK-2 (Sanjo*, et al.*, 1998). They are all Ser/Thr kinases but only DAPK1 and DRP-1 are Ca^{2+}/calmodulin-regulated, both sharing high homology in their kinase catalytic domains and the CaM binding region (Fig. 3.1). The other proteins are nuclear and do not have the same mechanism of regulation (Inbal*, et al.*, 2000). Upon overexpression of the aforementioned proteins, plasma membrane blebbing, cell rounding and detachment from the extracellular matrix are observed (Inbal*, et al.*, 1997) - all signs of apoptosis. While comparison of the catalytic domains of the members of the death-associate protein kinase family reveals high homology, the remaining part of the proteins differs significantly. In comparison with other family members, DAPK1 is considerably larger and contains regions involved in binding to the cytoskeleton, multiple ankyrin repeats as well as a death domain, all absent in the others (Fig.1A) (Cohen*, et al.*, 1997).

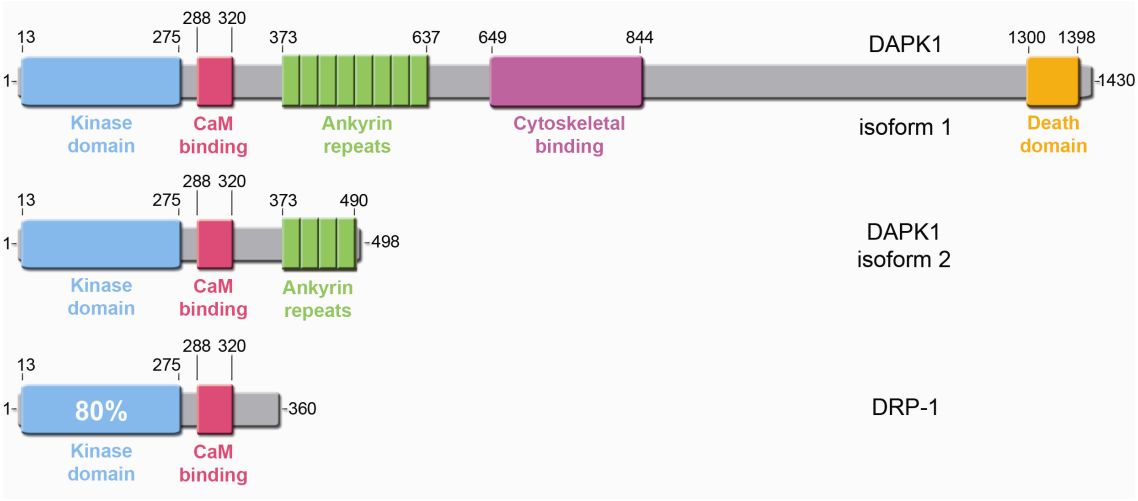


FIGURE 3.1: DEATH ASSOCIATE PROTEIN KINASES REGULATED BY CA2+/ CALMODULIN BINDING.

DAPK1 exists in two almost identical on sequence level isoforms. DRP-1 shares 80% identity on the level of the kinase catalytic domain.

The structure of the DAPK1 kinase catalytic domain has been resolved by *Tereshko et al* in 2001 as a first example of catalytic domains from its family (Tereshko, *et al.*, 2001). Structures were solved for apo- and AMPPNP forms with Mg^{2+} and another with Mn^{2+} coordinating ions. Curiously, the ligand-containing crystal forms do not reveal a second metal ion cofactor and when structures are superimposed with other similar proteins, the γ-phosphate of the nucleotide shifts into an atypical orientation. It was found that the three dimensional structure of the apo form shared high similarity with those of the twitchin, titin, CaMKI, phosphorylase and cAMP-dependent kinase. The structure appears to be in active conformation with a specific feature – a basic loop that clusters seven Arg and Lys residues invariant for the closely related kinases in the family. It appears important for substrate recognition as it acts as a lid upon a putative peptide binding region (Tereshko, *et al.*, 2001).

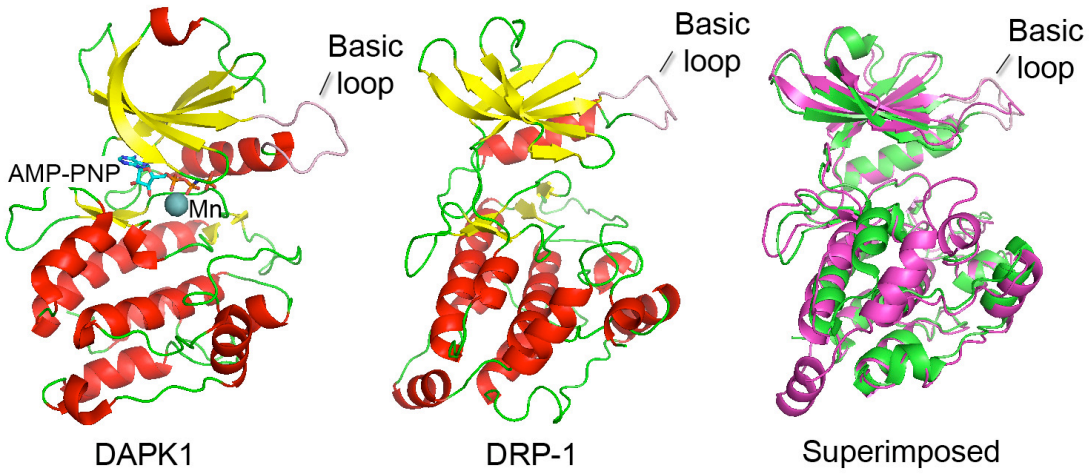


FIGURE 3.2: DAPK1 (PDB CODE: 1JKS) AND DRP-1 (PDB CODE: 1WMKE) SHARE HIGH STRUCTURAL SIMILARITY.

Both have the typical for the class basic loop and when superimposed deviate from each other only with 1.8Å RMSD. In the structures, DRP-1 has a C-terminal extension.

The cellular substrates for DAPK1 are unknown and in order to assay its enzymatic activity, studies have led to the discovery of peptide substrates for the enzyme. Initially, the consensus substrate sequence of RXXSX was established by comparison of DAPK1's structure with that of phosphorylase kinase crystallised with a peptide substrate. After screening for the variant positions, substrates with K_m of 10 μM or better were established as R(ART)(LYPMF)S(ND) - parentheses signify similar preference for amino acids at the particular position (Velentza*, et al.*, 2001). This work represents the first account of substrate discovery by positional scanning as a method for identifying protein kinase substrates and made possible measurements of the catalytic constants of DAPK1.

DAPK1 mediates IFN-γ-induced cell death by both caspase-dependent and caspase independent pathways (Inbal*, et al.*, 2002,Raveh and Kimchi, 2001). While its natural substrates are unknown it was found that p53 is essential for caspase-dependent apoptosis

(Type I) as it is responsible for Smad-mediated transcriptional upregulation of the *DAPK* gene (Raveh and Kimchi, 2001). Consequently, caspases and cysteine proteases attack mitochondria and lead to their membrane depolarization, Bcl-2 prosurvival factors are downregulated and numerous changes lead to activation of DNAses and resultant DNA degradation. In the case of caspase-independent apoptosis (Type II), also referred to as autophagic cell death, the role of DRP-1 and ZIP kinases is essential within different cellular settings. It has also been shown that intracellular autophagic vesicules consume organelles and cytoplasmic components without visible DNA loss or cytochrome *c* release (Gozuacik and Kimchi, 2004).

Due to its ability to sense apoptotic signals and counter hyperproliferation, DAPK1 is defined as a tumour suppressor (Jang*, et al.*, 2002). DAPK1 inhibits tumour development and a decrease in its potency may be one mechanism that is exploited by cancerous cells. It has therefore been suggested that DAPK1 expression level could be used as a biomarker to evaluate the disease state (Bialik and Kimchi, 2006).

The role of DAPK1 in cell death makes clear the importance of precise regulation of its activity. The protein is present in a constitutively inactive form that results from a default autophosphorylated state of the carboxyterminal region of the kinase catalytic domain on Ser308. It was suggested that a death-regulated phosphatase specific for DAPK1 family members dephosphorylates this residue permitting the binding of calmodulin under appropriate Ca^{2+} concentrations, resulting in full activation of the enzyme (Bialik and Kimchi, 2006). Structural information for the autoinhibitory conformation and the active conformation in complex with calmodulin is currently unavailable.

In this study, DAPK1 has been used as a model multidomain protein kinase for the experimental definition of soluble kinase catalytic domains by library screening. The

previous structural characterisation of a C-terminally truncated (we suspect overtruncated – see later) N-terminal catalytic domain ((Tereshko, *et al.*, 2001) PDB code: 1IG1) suggested that we would be able to identify soluble constructs encompassing its catalytic domain. Additionally, biologically interesting data might be obtained in the case of identifying a C-terminally extended catalytic domain encompassing the calmodulin binding site containing Ser308 as well as even Ser289 implicated in the kinase regulation (Anjum, *et al.*, 2005). Previously the calmodulin binding region was characterised as alpha helical when a chemically synthesised peptide was crystallised in complex with calmodulin (PDB code: 1YR5). In a further development towards defining a functional, derepressed enzyme, we screened the ESPRIT DAPK1 library against a co-expressed calmodulin bait to identify the subpopulation of the library that interacts with calmodulin to form a purifiable complex for further study.

Results

Kinase catalytic domain screen and characterisation

Initial considerations

A bioinformatic analysis of the target was performed prior to experimental work before the design of to define the construct screening approach. Whilst in classical structural biology the objective of this analysis would be to estimate at high resolution where domain boundaries could be located, here we were simply establishing the approximate location and size of the catalytic domain together with some estimation of order and disorder propensity.

Structures of DRP-1 show a carboxy terminal extension of the kinase catalytic domain not present in the current structures of DAPK1. An obvious question was whether or not a similar extension existed in DAPK1 and if it completes the kinase catalytic domain. Homology

sequence structure predictions (HSSP) have proven useful especially in the family of protein kinases due to the high conservation of their domain topography. A notable example of how accurately homology models can be in this protein class was a model's key role in the design of the first kinase drug fasudil (Johnson, 2009). Since this method explores the sequence alignments with existing structures it could confirm the identity and location of a particular protein domain as well as definition of its approximate limits.

DAPK1 sequence 1-330 was submitted to I-TASSER (Zhang, 2008) to generate a homology model using DRP-1 (PDB code: 2WMKE) as a template with 74% sequence identity. While evaluation of the accuracy of the structure prediction is a complicated process, a measure could be given by comparing the deviation between the existing structures and the model (Zaki, 2008). An HSSP can be considered potentially quite accurate if it does not deviate at the level of Cα atoms in the protein chain more than the maximum deviation among the determined structures. In this case, the use of the model was to predict the protein domain location and integrity so a high accuracy model was not necessary. The maximum deviation between the experimental structures was seen between DAPK1 1JKT and the rest with up to 1.01 Å RMSD. When the model returned was superimposed onto 1JKT as a representative deviating structure 1.16 Å RMSD was observed, decreasing slightly when compared to 1IG1 as a representative of the similar DAPK1 structure with deviation of 1.13 Å rms for the Cα atoms. Therefore, the model returned was closely matching the original structure regarding the kinase domain itself and could be considered of good accuracy. The additional C-terminal amino acids modelled to form an extension, which completes the kinase catalytic domain by folding back onto it. A long helical extension is modelled for the Ca^{2+}/calmodulin binding site. The helical extension appears to stretch up to the basic loop typical for the family (Fig. 3.3a & b). Close inspection of the modelled C-terminus reveals the presence of hydrophobic

patches formed between the domain and the extension. In the latter, six amino acids spanning the region 288 - 298 participate in the cluster (Fig. 3.3c). In this first phase of modelling using available information from homologous structures a low resolution model was produced that suggested that firstly, the earlier structures of DAPK1 may have been significantly over truncated, and secondly provided a framework to try and understand any longer constructs obtained during the following experimental phase.

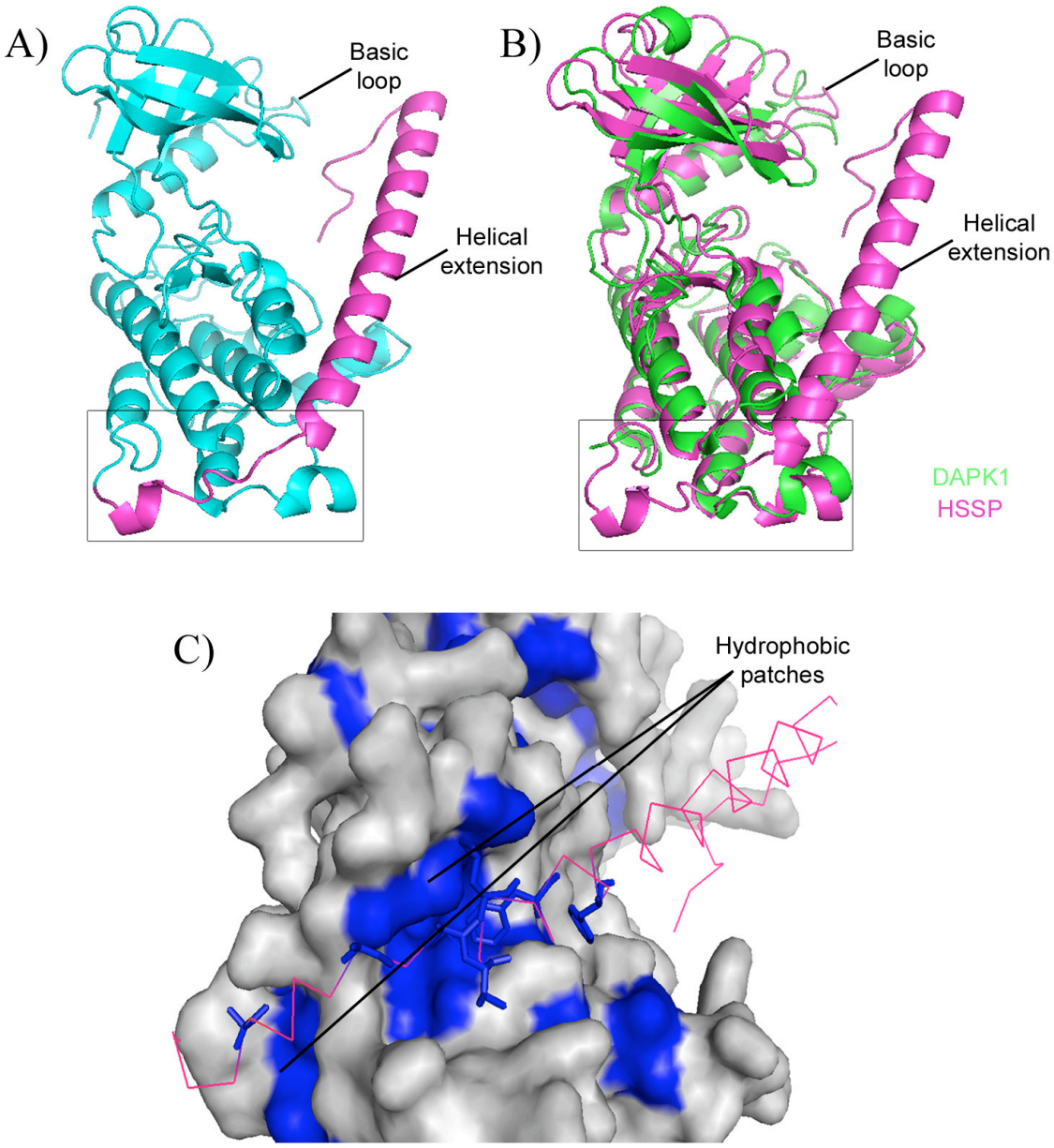


FIGURE 3.3: HOMOLOGY SEQUENCE STRUCTURE PREDICTION.

(A) the extension folds back to the kinase catalytic domain (magenta); (B) superimposition of the original DAPK1 structure (PDB code: 1IG1) and the model shows close fit; completion of the kinase domain is shown in a rectangle and in close-up; (C) surface representation of the crystallised construct shows two hydrophobic patches that are covered by the C-terminal extension;

Library synthesis and screening

DAPK1 is present in two different isoforms – a canonical sequence of 1430 amino acids and another 498 amino acid sequence that lacks the region 499-1430 (Deiss*, et al.*, 1995). At the sequence level, both of the isoforms are identical except for 21 amino acids at carboxyl terminus of the second isoform. The construct screening strategy chosen on this target was to generate a 3′ truncation library on DNA encoding a 500 amino acid construct from the canonical isoform sequence, thus fixing the native 5′ end of the gene. This construct also matches very closely the second isoform of the protein, containing four ankyrin repeats in addition to the kinase catalytic domain and calmodulin binding region. Since this protein was successfully expressed without knocking down its activity, a catalytically dead mutant was considered unnecessary. The synthetic gene construct (Geneart) was cloned in the ESPRIT library vector (pYUM6002) described in chapter 2. The truncation performed resulted in an efficient library with a distribution of fragments spanning 70 amino acids to full length with few if any excessively truncated gene inserts (Fig.3.4).

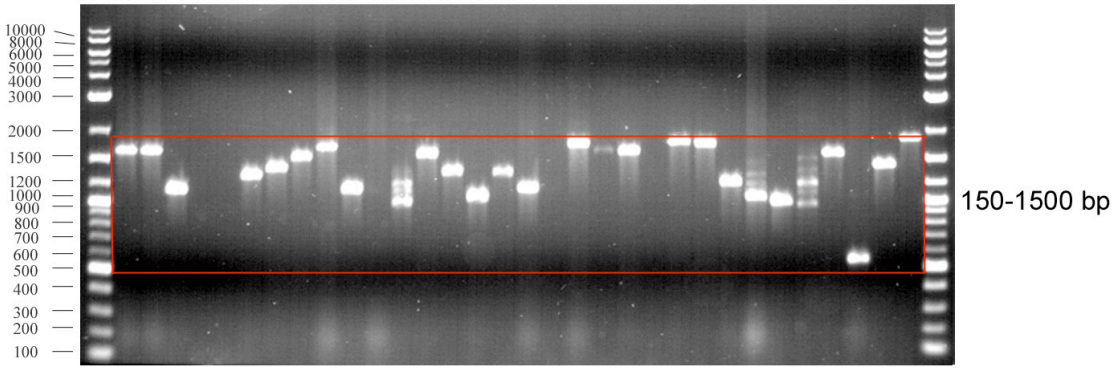


FIGURE 3.4: REPRESENTATIVE COLONY PCR SCREEN FOR THE EVALUATION OF DAPK1 3′ TRUNCATION LIBRARY QUALITY.

24 of 96 tested are shown. Insert size is equal to PCR product size minus 320bp from the flanking primer sites.

E. coli BL21 was transformed with the plasmid library and 4608 individual clones isolated by automated colony picking into 384 well plates. The oversampling for the library was calculated at 3-fold taking in consideration the library quality of 90%. Comprehensiveness of the screening is measured by calculating its completeness defined as the percentage of library covered (Patrick, *et al.*, 2003). The library statistics are presented in Table 3.1. This level of completeness is significantly higher than the majority of classic directed evolution experiments (e.g. with point mutagenesis) where only a fraction of a percent of the diversity is usually tested.

	DAPK1
Construct (aa)	1-500
Truncation Type	3′
Size Range (bp)	150-1500
Oversampling	1350
Quality (%)	3.4
Contamination (%)	90
Diversity	0
Completeness (%)*	95.4

TABLE 3.1: SUMMARY OF THE GENERATION AND SCREENING OF DAPK1 3′ TRUNCATIONAL LIBRARY.

*Completeness is calculated using server http://guinevere.otago.ac.nz/aef/STATS/index.html.

The isolated clones were arrayed onto nitrocellulose membranes and tested for their N-ter hexahistidine and C-ter biotinylation levels by hybridisation of fluor-conjugated anti-6his

mAb and streptavidin. Constructs with no hexahistidine tag signal were eliminated from further analysis and the 96 most strongly biotinylated clones were isolated for the second level of screening by purification from 4 ml expression cultures. Ni^{2+}NTA purification was carried out with using a liquid handling robot (TECAN Genesis) and purified proteins were analyzed on 15% SDS-PAGE (Fig. 3.5a). A number of proteins were clearly visible by Coomassie blue staining suggesting relatively high soluble expression. They all clustered between 35-42kDa (including tags), a size corresponding well with a typical kinase catalytic domain with 5 kDa of additional tag sequences. Purified clones were then sequenced to identify their exact boundaries.

The ESPRIT screen on this protein identified a number of soluble constructs clustering about 30 amino acids longer than the only construct crystallized previously, 1-285 (the sequence after amino acid 277 is invisible). When mapped onto the HSSP model it was apparent that this extra sequence corresponded to the region modelled as a helical extension from Ala300 to Phe322 (Fig. 3.5b). This region is expected to be flexible allowing calmodulin binding and subsequent revelation of the kinase active site. The hydrophobic patches observed from the homology sequence structure predictions appear to be integral part of the domain and truncations in this region were not observed. Therefore, it appears that the construct screening approach leads to the precise definition of the complete kinase catalytic domain.

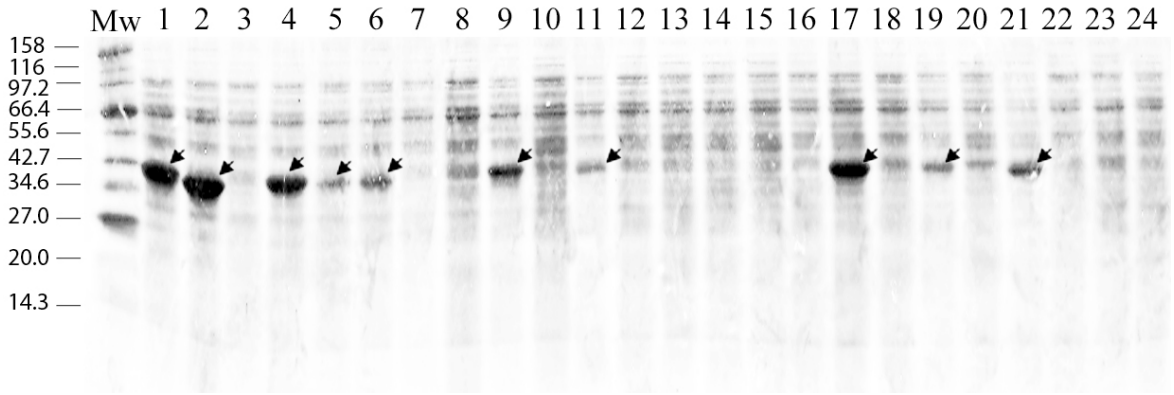


FIGURE 3.5: IDENTIFICATION OF SOLUBLE AND PURIFIABLE CONSTRUCTS BY ESPRIT.
Representative Coomasie blue stained 15% SDS-PAGE from the purification screen. 24 of 96 constructs are shown;

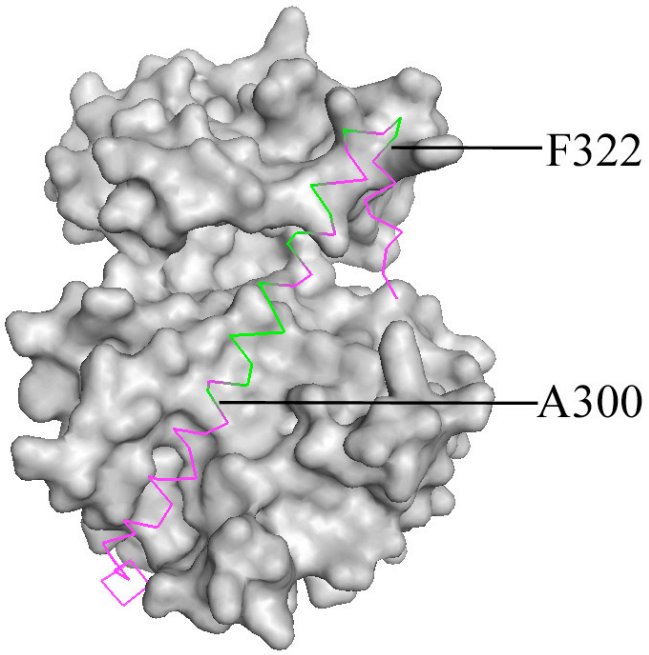


FIGURE 3.6: MAP OF THE SOLUBLE CONSTRUCTS IDENTIFIED FOR DAPK1 ONTO THE HSSP MODEL.
Sequencing reveals them to cluster in the C-terminal helical extension ranging from Ala300 to Phe322;

Three constructs were taken for further characterisation: 1-300, 1-304 and 1-314. The first two were scaled up with high yields of approximately 40 mg of purified protein from 1 l of expression culture in TB medium while the third one exhibited lower yields of 2-3 mg per litre of culture. After Ni^{2+}NTA purification, the proteins were of high purity and their N-terminal purification tags were quantitatively cleaved by TEV protease. The first two constructs appeared monomeric by size exclusion chromatography eluting in the expected fraction for their size volume and with almost no visible aggregation in the void volume (Fig. 3.6a & b). When analysed by dynamic light scattering (DLS), they exhibited one single peak

with a polydispersity index between 0.1 and 0.2 (Fig. 3.6c), suggesting low levels of aggregation. Little aggregation could be detected for the shorter construct.

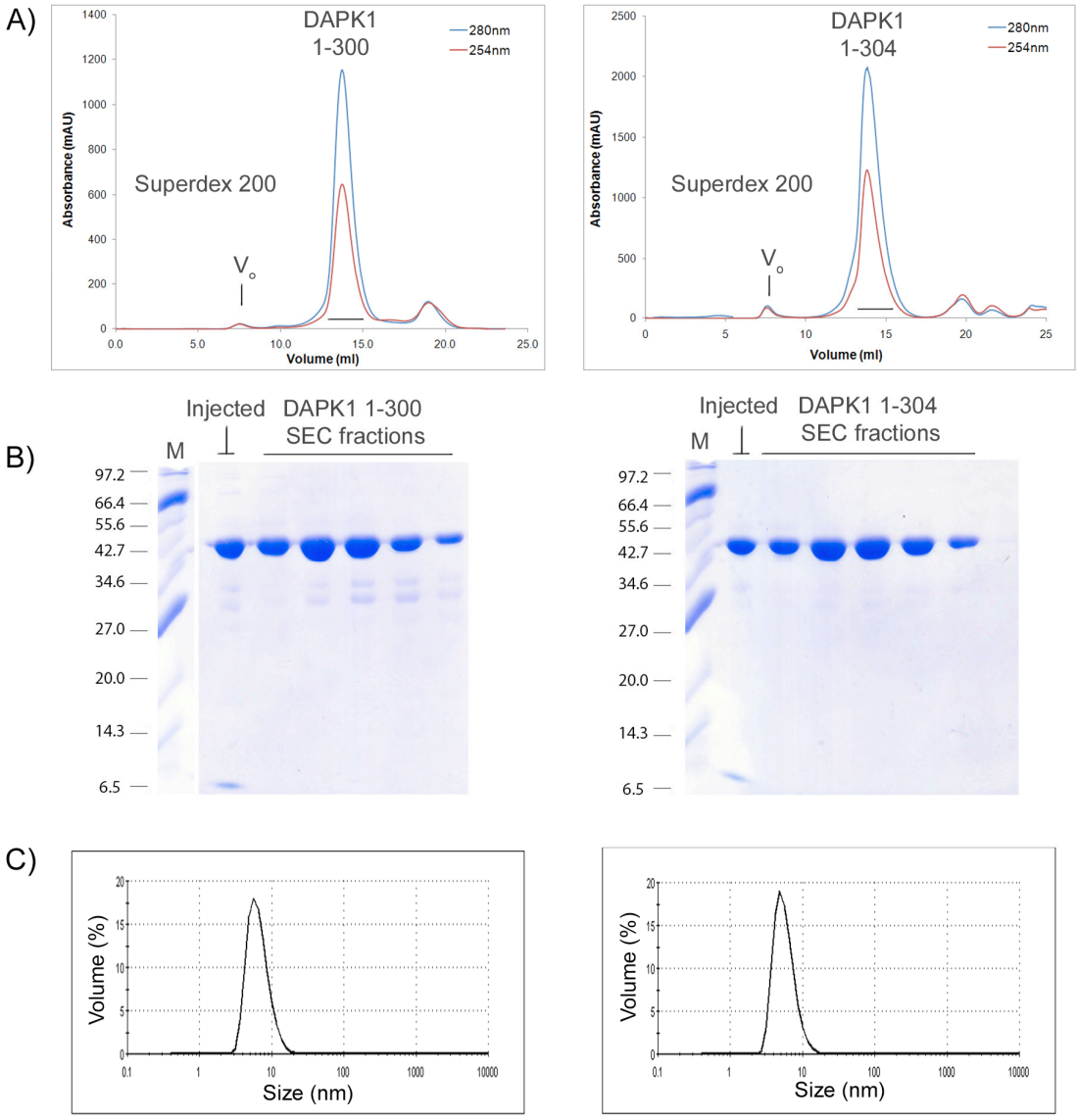


FIGURE 3.7: CHARACTERISATION OF TWO REPRESENTATIVE CONSTRUCTS FROM THE SCREEN

(A) SEC shows almost no protein in the void volume and a single peak at the expected elution volume; (B) SDS-PAGE analysis of the peak areas showing highly pure protein (the injected sample had the hexahistidine tag cleaved off with TEV and was purified with a second Ni2+NTA step); (C) DLS analysis of purified protein showing single peaks and low polydispersity;

Further validation of experimentally defined constructs

In contrast to a classical approach where only a small number of clones are tested, construct screening methods such as ESPRIT can result in multiple similar hits when a domain is identified. The additional C-ter sequence of these DAPK1 domains may be disordered or may be an ordered part of the domain that can be removed. We reasoned that measurement of melting temperature might indicate whether additional amino acids were making stabilising contacts i.e. were an integral part of the structure. Additionally, specific ligand binding might increase the T_m and if so, would indicate that the domains had the expected, functional (native) fold. Thermofluor analysis is a relatively recent technique that has the advantage of low sample consumption and easy parallelisation, permitting direct comparison of multiple constructs across multiple buffer/ligand conditions (Ericsson*, et al.*, 2006). The method follows the increase in fluorescence of SYPRO orange in the presence of the target protein upon thermal denaturation. The dye fluoresces upon binding to hydrophobic, buried parts of the protein that become more accessible when the protein unfolds. While this assumption holds true in a number of cases it is not universal thus there are cases in which this assay does not work.

The initial screen was set up in 48-well format aiming to identify optimal conditions for the stability of the protein (Table. 3.2). Two constructs were screened against two different salt concentrations and various buffers ranging from pH 5 to pH 9. All conditions screened suggested the presence of stable, folded protein with decreased stability at low pH. Another interesting tendency was that construct 1-300 exhibited lower Tm values in the various conditions with approximately 7 °C difference when compared to construct 1-304. The best

conditions for both constructs appeared to be in the presence of 100 mM NaCl and pH 8 with slightly better Bicine and similar HEPES and Tris buffer.

	NaCl 500 mM			NaCl 100 mM		
	1	2	3	4	5	6
A	Citric acid pH 5	NaH_2PO_4/ Na_2HPO_4 pH 6.6	HEPES pH 8	Citric acid pH 5	NaH_2PO_4/ Na_2HPO_4 pH 6.6	HEPES pH 8
B	$Na_3C_6H_5O_7$ pH 5.5	NaH_2PO_4/ Na_2HPO_4 pH 7	Tris pH 8	$Na_3C_6H_5O_7$ pH 5.5	NaH_2PO_4/ Na_2HPO_4 pH 7	Tris pH 8
C	MES pH 5.8	KH_2PO_4/ K_2HPO_4 pH 7	Tricine pH 8	MES pH 5.8	KH_2PO_4/ K_2HPO_4 pH 7	Tricine pH 8
D	KH_2PO_4/ K_2HPO_4 pH 6	HEPES pH 7	Bicine pH 8	KH_2PO_4/ K_2HPO_4 pH 6	HEPES pH 7	Bicine pH 8
E	$Na_3C_6H_5O_7$ pH 6	Amomium acetate pH 7.3	Bicine pH 8,5	$Na_3C_6H_5O_7$ pH 6	Amomium acetate pH 7.3	Bicine pH 8,5
F	MES pH 6	NaH_2PO_4/ Na_2HPO_4 pH 7.6	Tris pH 8,5	MES pH 6	NaH_2PO_4/ Na_2HPO_4 pH 7.6	Tris pH 8,5
G	MES pH 6.2	Tris pH 7.5	Bicine pH 9	MES pH 6.2	Tris pH 7.5	Bicine pH 9
H	MES pH 6.5	Imidazole pH 8	H_2O	MES pH 6.5	Imidazole pH 8	H_2O

TABLE 3.2: BUFFERS FOR THE THERMAL STABILITY SCREEN IDENTIFYING OPTIMAL CONDITIONS BY THERMOFLUOR.

Next, ligand/inhibitor effects on the temperature-dependent unfolding of the protein were determined. This can be of general interest for screening various ligands and inhibitors in a high throughput manner to rank with low resolution the affinity of the different binders. An initial screen identified a range of concentrations for testing the binding of different ligands (Fig. 3.2).

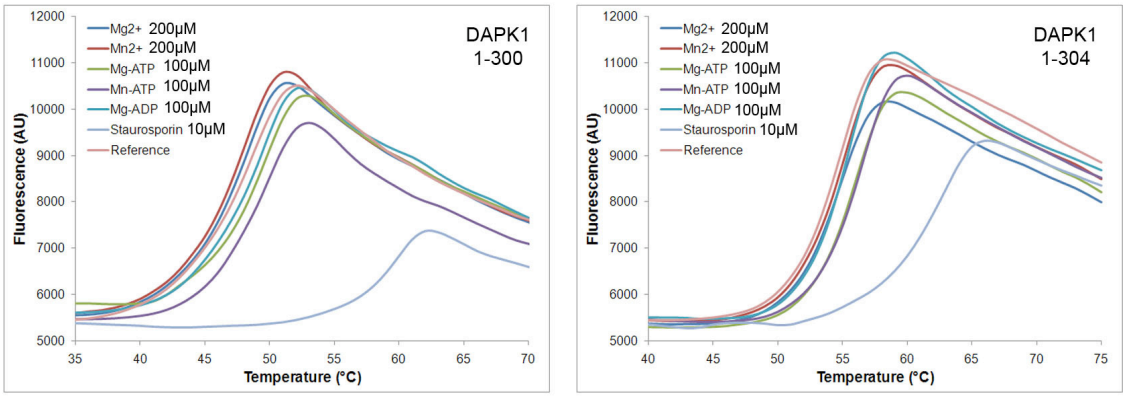


FIGURE 3.8: THERMOFLUOR THERMAL STABILITY SCREEN ON DAPK1 INDICATES LIGAND AND INHIBITOR BINDING.

ATP is a low affinity ligand and Tm values are lower than the high affinity inhibitor staurosporin.

From that screen, it was observed that ATP was successfully coordinated in the presence of either Mg^{2+} or Mn^{2+}. This correlates with the observed cocrystalization in the presence of either of the ions (Tereshko*, et al.*, 2001). In contrast ADP did not appear to influence the thermal stability of the domains. While 100µM concentrations of the cofactors led to a slight increase instability, ten times lower concentrations of staurosporin significantly shifted the curves to the right (Table 3.3).

	DAPK1 1-300 (°C)	DAPK1 1-304 (°C)
Bicine pH8	47	54
200µM Mg^{2+}	47	54
200µM Mn^{2+}	47	54
100µM Mg-ATP	48	56
100µM Mn-ATP	49	56
100µM Mg-ADP	47	54
10µM Staurosporin	59	62

TABLE 3.3: COMPARISON OF DAPK1 1-300 & 1-304 IN PRESENCE OF ATP AND STAUROSPORIN.

The data presents estimation of melting temperatures (Tm) from figure 3.8

This observation correlates well with the difference of binding affinity for the aforementioned ligands as staurosporin is known to bind tightly to many different kinases. The presence of ligands and inhibitor did not affect the trend of lower stability exhibited by the shorter, DAPK1 construct 1-300.

In an extension of this study, the effects of different concentrations of ATP and staurosporin were observed. A range from 0-1 mM ATP concentrations and 0-32 μM ATP were tested against 8 μM of final protein concentration. ATP stabilised the protein by a maximum of 10 °C at a saturating concentration between 300-400 μM exhibiting a classical profile. On the other hand staurosporin exhibited a profile ranging from a single peak at low concentrations, correlating with weaker inhibition, through two peaks and a single peak to the right at highest saturating concentration (Fig. 3.9). This observation suggests the presence of different populations of species depending on the concentration of the inhibitor. At lower concentrations of the inhibitor than the protein of interest the inhibition as expected 1:1 molecular ratio appears to be incomplete due to limiting quantities of inhibitor. Upon increase of the concentration the population of bound molecules increases, the peak shifts gradually to one at saturated concentration of binding. At saturating ligand concentrations, the highest stabilization observed accounted for a shift of stability of about 15°C. This effect was more pronounced for DAPK1 1-300 (Fig. 3.9).

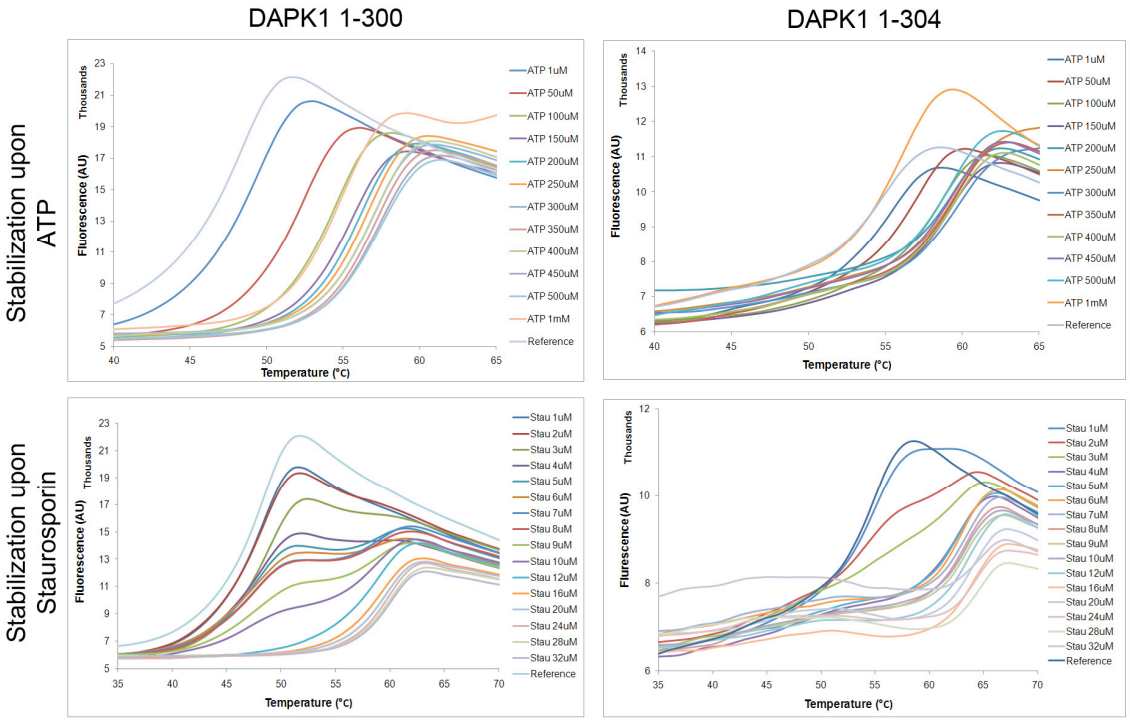


FIGURE 3.9: DAPK1 THERMAL STABILIZATION UPON TITRATION OF LIGAND AND INHIBITOR CONCENTRATIONS.

Structural studies on the DAPK1 kinase domain

The homology sequence structure prediction implies that the kinase catalytic domain, as crystallised to date, is incomplete. Moreover, DAPK1 is Ca^{2+}/calmodulin regulated kinase with studies suggesting the region 302-320 is responsible for calmodulin binding. There is a particular value in studying the extension of the kinase catalytic domain structurally in order to understand the underlying activity regulation mechanisms. DAPK1 1-300 and 1-304 as identified by construct screening with the C biotin acceptor peptide were expressed in BL21 (DE3) with 500 mL TB expression culture. The proteins were purified by single affinity step to high homogeneity and their aminoterminal 6xHis tag cleaved by TEV protease. After a final size exclusion chromatography step they were submitted for robotic crystallization trials at concentrations 2.2 and 3.1 mg/ml respectively. The screen was set with sitting drops in 384

well plates with 100 nl drops. Multiple crystallization conditions were found for both constructs with a large range of pH - 4.5 to 9.1. No particular requirements for buffer, salt or precipitant were seen indicating that DAPK1 is relatively easy to crystallise. Intriguingly, all crystals for DAPK1 1-300 formed needle like crystals while for DAPK1 1-304 there were two different forms – needle and more bi-pyramidal forms. Crystals were fished from the screening robotic plates and were initially tested on beamline ID14-4 (ESRF, Grenoble). The crystals diffracted to 2.8-3 Å resolution (Fig. 3.10) and two space groups were identified by indexing these images with MOSFLM. The needle-like crystals were monoclinic ($P2_1$) and the bi-pyramidal form were orthorhombic ($P2_12_12_1$) (Table 3.4). Among the resolved structures in PDB, almost all DAPK1 crystals crystallized in a $P2_12_12_1$ space group with the exception of an apo form, which crystallised in a P4 space group (PDB code: 1JKT). We have thus crystallised DAPK1 in a new space group whereby alternative crystal packing may reveal another conformation of the protein. It is a common observation that small variations in the protein construct can reveal new information and a major advantage of using the ESPRIT methods is that multiple similar constructs often clustering around a core domain.

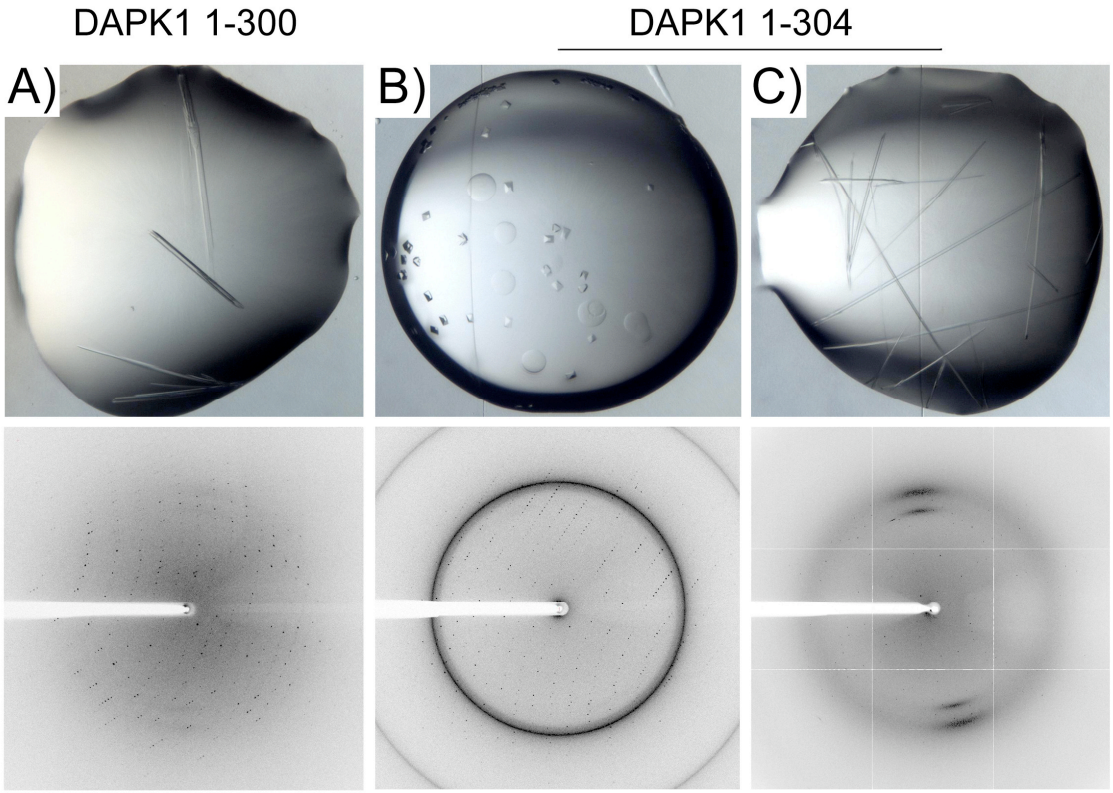


FIGURE 3.10: DAPK1 CONSTRUCTS ARE READILY CRYSTALLISABLE AND DIFFRACT. (A) Crystals grown in 0.2 M potassium fluoride and 20 %w/v PEG 3350 at pH 7.2; (B) Crystals grown in 0.1 M MES and 5% w/v PEG 6000 at pH 6; (C) Crystals grown in 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate and 10% PEG 8000 at pH 6.5.

In a second session at the micro-focus beamline ID23-2 (ESRF, Grenoble), data sets were collected for the best diffracting crystals screened for both the DAPK1 1-300 and DAPK1 1-304 constructs. The highest resolution complete dataset was obtained at 2.6 Å and it was solved by molecular replacement using PDB 1IG1 (DAPK1: 1-277). Preliminary model building on data collected for the orthorhombic crystals of DAPK1 1-304 shows rearrangements at regions 100-114, 149-156, and 169-178 (Fig. 3.11). The first part of the C-terminal extension (residues 277-300) folds back to the kinase catalytic domain as predicted in the HSSP model and suggests that previous structures may have been over truncated in this region. An extra helix at the terminus of the construct appears to lead to rearrangements of

the aforementioned regions 100-114 and 149-156 amino acids. The third rearrangement (169-178) appears at the level of the activation loop. The ATP binding pocket does not appear to be influenced with the αC almost perfectly superimposed on 1IG1. When the amino acids responsible for catalysis and ATP coordination are compared, they were also observed to preserve their original conformations.

	1-300	**1-304**	**1-304**
Crystal Form	Needles	Orthorhombic	Needles
Data Collection			
Beamline (ESRF)	ID23-2	ID23-2	ID14-4
Wavelength (Å)	0.873	0.873	~ 3.0
Space Group	$P2_1$	$P2_12_12_1$	$P2_1$
Cell Dimensions			
a, *b*, *c* (Å)	56.6, 49.9, 129.4	50.2, 78.9, 111.5	57.0, 50.0, 131.4
α, β, γ ( )	90, 96.4, 90	90, 90, 90	90, 94.2, 90
Resolution range (Å)	50.0 – 2.7 (2.8-2.7)	50.0 – 2.6 (2.7-2.6)	
Unique reflections	18,386 (1,658)	13,175 (1,060)	
Observed reflections (I/σ(I) > 0)	50,659 (4,612)	52,998 (4,309)	
Completeness (%)	91.5 (80.9)	92.9 (79.8)	
R_{merge}†	11.4 (59.6)	9.9 (64.8)	
Mean I/σ (I)	9.1 (2.3)	12.3 (2.3)	

TABLE 3.4: SUMMARY OF CRYSTAL FORMS AND DATA COLLECTION FOR THE DAPK1 CONSTRUCTS IDENTIFIED BY ESPRIT CONSTRUCT SCREENING.

† Rmerge = $\sum_h \sum_i \left| I_{h,i} - \langle I \rangle_h \right| / \sum \left| \langle I \rangle_h \right|$ calculated for the whole data set. Values in parentheses are for the outermost shell of data collected.

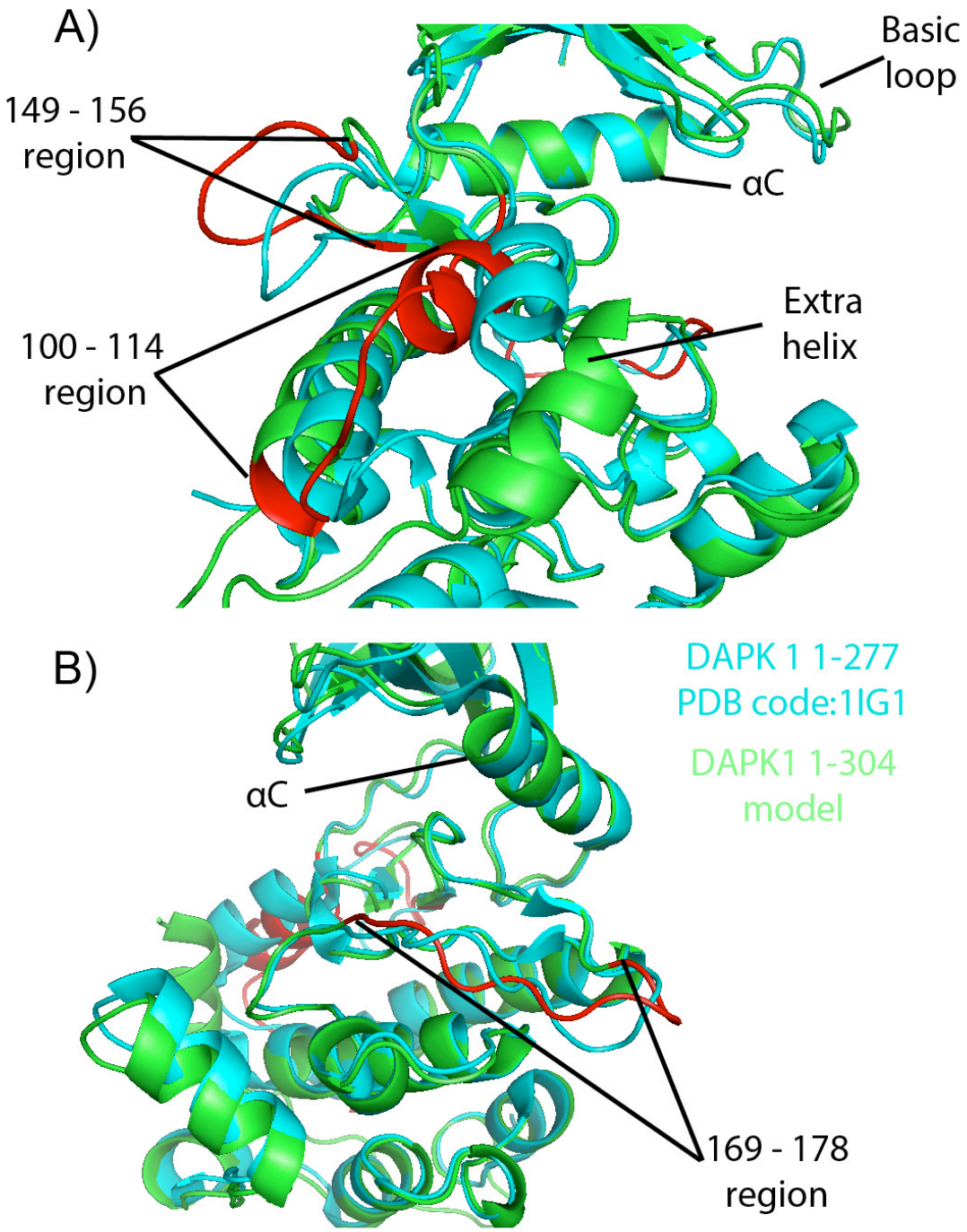


FIGURE 3.11: COMPARISON OF DAPK1 1-277 (PDB CODE: 1IG1) AND PRELIMINARY STRUCTURE OF DAPK1 1-304.

Superimposed structures deviate with 0.91 Å RMSD. Rearranged regions are coded in red. (A) Zoom up on the region at the C-terminus longer than as visible from the original structure also shows rearrangement in amino acids regions 100-114 and 149-156; (B) View at horizontal 90° rotation showing the rearrangement of the activation loop of the kinase - amino acids region 169-178;

DAPK1 calmodulin (CaM) co-expression screen

A new co-expression version of ESPRIT was tested for the direct in-cell identification of soluble complexes between DAPK1 and calmodulin. A synthetic gene encoding calmodulin with N-ter flag tag was cloned into pACYC-based co-expression vector under the control of a T7 promoter (pACYC-LIC; SGC Toronto) and competent *E. coli* BL21 cells were prepared containing this vector. The cells were transformed with the 3′ truncation library prepared previously for DAPK1 and 4608 colonies were isolated with a colony picker robot. Initially, clones were ranked for biotinylation as was originally used for identifying soluble DAPK1 constructs. Following the second level screen for protein purification of the His6-tagged DAPK1 by NiNTA, a western blot against the flag tag was performed in order to identify the subset of DAPK 1clones copurifing with calmodulin. While calmodulin was well expressed from the same pACYC vector in a control experiment, it was found that expression levels in the context of DAPK1 were significantly reduced resulting in poorly visible bands on SDS-PAGE gels of lysates from the 4 ml cultures. These were however confirmed as calmodulin by western blot and a subset of the DAPK1 clones copurified with this calmodulin during the NiNTA purification step (data not shown).

Some clones with DAPK1 that copurified with calmodulin plus some that did not were scaled up to 50 ml expression tests. A control DAPK1 construct 1-320 was made (corresponding to the catalytic domain plus calmodulin-binding domain previously studied as an isolated peptide) and expressed in the presence and absence of calmodulin. All purifications were performed with 2 mM $CaCl_2$ as a source of Ca^{2+} shown to be necessary for correct folding of calmodulin. SDS-PAGE analysis of the purifications revealed the presence of copurified calmodulin for a subset of the clones and the positive DAPK1 control (1-320) co-expressed in

with calmodulin (Fig. 3.12a). The expression and copurification of calmodulin was verified by western blot for the total protein and NiNTA purified fractions (Fig. 3.12b).

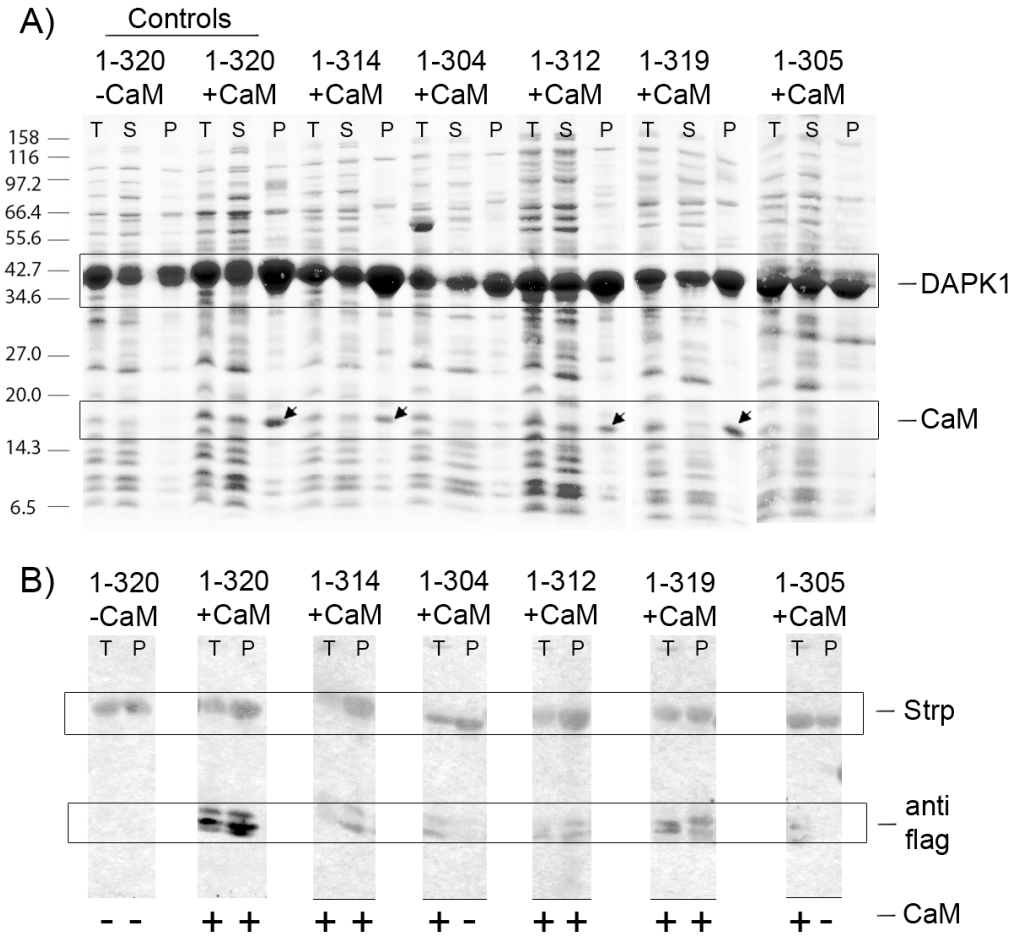


FIGURE 3.12: MINIMAL CAM BINDING TO DAPK1 REQUIRES C-TERMINAL EXTENSION UP TO ILE312 (PDB CODE: 1YR5).

(A) SDS-PAGE with rectangles highlighting DAPK1 and CaM (shown also by arrows); (B) anti flag and streptavidin detection western blot reveals the expression of calmodulin and its copurification indicated with +/- on the bottom; Legend: T-total, S-soluble and P-purified by Ni2+NTA fractions;

The identity of the DAPK1 constructs was revealed by DNA sequencing. Surprisingly, the clones that copurified calmodulin were found to be shorter than the expected calmodulin binding region spanning up to Arg320. A stable interaction was observed in truncations as

short as Ile312. For constructs shorter than 305 amino acids, no interaction at all was observed. When the interaction is compared with a reported structure of calmodulin with a peptide of the DAPK1 calmodulin binding region, it appears that N-terminal domain of calmodulin would not fully interact with the shorter constructs. There are two possible explanations for the presence of interaction between constructs shorter than 320 amino acids and calmodulin. Firstly, the interaction of calmodulin with the DAPK1 peptide as reported from the structure (PDB code: 1YR5) may not be the same as the one in context of the kinase catalytic domain. Secondly, the affinity of calmodulin for the shorter constructs decreases but is still sufficient to form a complex.

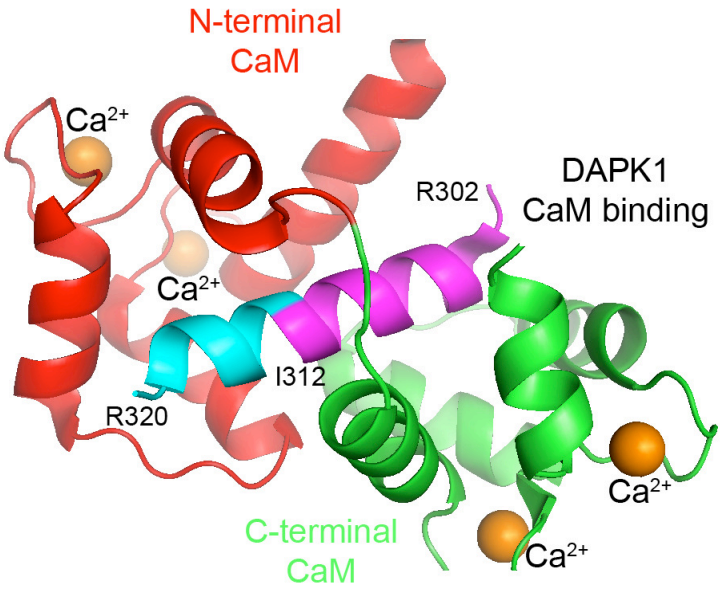


FIGURE 3.13: CA2+/CAM BINDING TO DAPK1 CONSTRUCTS IDENTIFIED BY THE CO-EXPRESSION SCREEN.

The results are mapped onto 1YR5. DAPK1 C-terminal extension up to Ile312 (pink) binds to calmodulin.

Discussion & Conclusions

In order to test ESPRIT as a screening method for the determination of kinase catalytic domains in multidomain human kinases, DAPK1 was chosen as a more characterised system. By bioinformatic analyses it was also found that DAPK1 is closely related to the main target IKKβ (*see* chapter 4). The availability of an existing structure allowed us to understand our screening data regarding domain boundaries.

Preliminary analysis by homology sequence structure prediction suggested that the kinase catalytic domain of DAPK1 was crystallized previously with a missing extension that forms part of the domain when present. Two hydrophobic patches appear to be covered by the extension and thus suggest the integrity of the kinase catalytic domain requires this extension.

A 3′ truncation library was synthesised on a starting construct comprising the first 500 amino acids from DAPK1 isoform 1 using the optimised vector pYUM6002 (*see* chapter 2). A high quality library was obtained with coverage of approximately 94%, thus assuring a highly comprehensive screen. The solubility screen successfully identified soluble protein fragments clustering between 35 and 42 kDa clearly visible on SDS-PAGE after the purification screen from 4 ml expression cultures. Subsequently, sequencing of the hits identified the borders of the constructs fell between residues 300 and 322 suggesting the isolation of longer constructs than the ones crystallised previously. Although the starting construct was significantly longer than this and included four ankyrin repeats, soluble constructs containing these regions were not identified.

Scale up and characterization of two constructs DAPK1 1-300 and 1-304 revealed they have different behaviour. While they were both highly expressed and monomeric by size exclusion chromatography with almost no visible aggregation and high monodispersity, they differentiated on the level of thermal stability. Both constructs were stable in variety of buffers with slight preference for neutral and low basic pH as well as low amount of salts.

DAPK1 1-304 was significantly more stable with average Tm difference of about 7°C to DAPK1 1-300 when tested for buffers, ligands and inhibitor. This suggests that the extra amino acids are ordered since they appear to contribute extra stability, presumably by making additional protein contacts. The thermal stability studies of these constructs show that significant enhancement of stability can occur from small construct changes. Ligand binding assays revealed that ATP can be successfully coordinated by both Mg^{2+} and Mn^{2+} and, at saturating concentrations, shifts the stability of the proteins by up to 10 °C. At much lower concentrations, the generic kinase inhibitor staurosporin, shifts the thermal stability of the constructs with about 15°C suggesting higher affinity and more protein contacts than ATP. More generally, ligand binding is strong evidence for native fold.

DAPK1 constructs 1-300 and 1-304 were readily crystallisable directly from the ESPRIT screen. They crystallized in diverse conditions with no strong preference for pH, buffer, salt or precipitants. Both constructs crystallized forming needles but the longer one formed also orthorhombic crystals. The crystals were isolated directly from the robotic screens and successfully diffracted to approximately 3 Å resolution and revealed that the proteins were crystallized in two different space groups. The orthorhombic crystals were P222 as were almost all previous published DAPK1 crystals whilst the needles were monoclinic P2 – a new space group for this protein. Alternative crystal packing arrangements can be helpful to study alternative conformations of the protein in ligand-bound or apo forms (Mozzarelli and Rossi, 1996). Construct screening methods such as ESPRIT may reveal several forms of domain (differing in termini) in contrast to the classic approach of construct design where typically only one construct is crystallised. Here we have shown in our model system how the identification of several well-expressing constructs can lead to alternative crystal forms that would not have been identified conventionally.

A preliminary structure of DAPK1 1-304 reveals a rearrangement at the C-terminus due to the presence of an extra helix that folds back against a predicted hydrophobic patch on the main domain. This new structured region appears to complete the integrity of the kinase catalytic domain. The local rearrangements observed in this new structure influence neither the ATP binding site nor the catalytic loop conformation. The significance of the changes in the orientation of the activation loop are inconclusive since variation in this region of kinase catalytic domains is common.

A co-expression screen for the identification of soluble complexes between truncated DAPK1 variants and calmodulin demonstrated the use of the co-expression system for fine mapping of an interaction interface. Unexpected, shorter variants of the C-terminal extension still competent for calmodulin binding were identified. The functional relevance of these constructs awaits the structure determination of a DAPK1-Ca^{2+}/calmodulin complex.

DAPK1, as a model system, demonstrated that soluble kinase catalytic domains could be identified by the ESPRIT construct screening approach. A cluster of soluble constructs was found all located in a flexible region not structurally important for the kinase catalytic domain. In an extension of this study, the truncation library was co-expressed in the presence of calmodulin and various soluble variants of the complex were identified. This work shows the potential of construct screening for both identifying soluble constructs but also for the investigation of kinase activity regulation mechanisms.

Materials & Methods

Open Reading Frames

Synthetic genes corresponding to the following proteins were used: DAPK1_human (Swiss-Prot Accession Number: P53355) 1-500aa and synthetic full length CALM_human (Swiss-Prot Accession Number: P62158) with N-terminal flag tag. The synthetic genes were

purchased from Geneart with optimised sequence for *E. coli* expression addressing codon bias, internal ribosome binding sites and mRNA secondary structure.

DAPK1 Library Entry Clone Cloning

The synthetic DAPK1_human gene encoding aa 2-500 was cloned in the small ESPRIT vector pYUM6002 by overlap extension PCR to add H6TEV tag with *Nde*I restriction site for cloning. The first PCR amplifying the gene was done with 400 nM gene specific primers DAPK1for1 (5′ CTACC GAGAA TTTGT ATTTT CAGGG GACCG TGTTT CGTCA GGAAA ATGTG 3′) and DAPK1rev1 (5′ CATTA CGCCG CGGCC GCACA CAGCG CTTTC GCCAC G 3′) with *Nsi*I restriction site amplified by *Pwo* polymerase in 25 µl reaction for 10 cycles with T_m=56 °C and 3 min of elongation at 72 °C on 10 ng of synthetic gene template. Consequent, amplification using 1 µl product of the first PCR reaction with 400 nM H6TEVfor (5′ CCATG ATTAT GATAT TCCAA CTACC GAGAA TTTGT ATTTT CAGGG G 3′) in place of DAPK1for1 and T_m=56 °C. The final amplification with 1 µl of the last reaction was done with 200 nM PB2synthfor3 (5′ GAGAT ATACA TATGG GCCAC CATCA TCACC ACCAT GATTA TGATA TTCCA ACTAC CGAGA 3′) (*Nde*I restriction site) as forward primer with T_m=57 °C, 5 min of elongation at 72 °C and 20 cycles of amplification in 50 µl total reaction volume. The PCR product was gel purified from 1% gel with a QIAEXII kit. Both PCR product and 2 µg of pYUM6002 were cut in buffer 2 (NEB) with 20 U *Nde*I for 1 h at 37 °C and buffer exchanged with Nucleospin kit (MachereyNagel). A second cut in buffer 3 (NEB) with 20U *Nsi*I followed for 1 h at 37 °C for both. The vector was dephosphorylated during the last restriction digest with 1 µl of SAP phosphatase for 1 h at 37 °C. After ligation and transformation, the colonies were tested by PCR screen and positives sequence verified.

DAPK1 Control Construcct Cloning

Positive control DAPK1 2-320 was cloned. The cloning strategy was the same as for the library entry clone. Firstly, 10ng synthetic DAPK1 was amplified with *Pwo* using 400 nM DAPK1for1 and DAPK1rev_pos2.2 (5′ CCCAT TGTTC GATGC ATTAC GGCTC AGACG CTGAC ACAGG CT 3′). The amplification was for 10 cycles with T_m=58 °C and 2 min of elongation at 72 °C. At the second amplification 1 µl from the previous reaction was used as template but substituting the forward primer with H6TEV. The final amplification was done with 1 µl of the previous reaction and PB2synthfor3 (5′ GAGAT ATACA TATGG GCCAC CATCA TCACC ACCAT GATTA TGATA TTCCA ACTAC CGAGA 3′) as forward primer. 1 µg of the backbone (pYUM6002) and the PCR products were digested for 1.5 h at 37 °C with tenfold excess of *Nde*I and *Nsi*I sequentially with PCR clean up in between. The backbone was further cut by *Dra*III supplemented with BSA along with *Nsi*I and dephosphorylated by addition of 6 µl of SAP, extending the digestion at 37 °C for 1 h. After ligation and transformation, the colonies were tested by PCR screen and positives sequence verified.

Calmodulin Cloning

The calmodulin construct was subcloned directly from 2 µg of the synthetic gene and pACYC vector digested in 30 µl with 10U *Nde*I and 10 U *Not*I in buffer 3 consecutively each for 1 h at 37 °C with a PCR reaction clean up step in between. The vector was dephosphorylated with 1 µl SAP for 1 h and cleaned up. After ligation and transformation, the colonies were tested by PCR screen and positives sequence verified.

Competent cells

Electrocompetent cells containing the calmodulin gene on pACYC vector were prepared as in Appendix (*see* Electrocompetent Cells Preparation, Appendix)

Library Generation

The library generation was performed largely as described in chapter 2 (*see* ESPRIT Protocol, Materials, Methods and Protocols section) with the following modifications: 4 µg of the linearised with *Nsi*I and *Not*I vector were submitted to time course exonuclease III reaction limited to 25 bp/min truncation by 44.13 mM NaCl at 22 °C quenching 2 µl of reaction every min in 3 M NaCl at 70 °C. The library was screened in an IPTG inducible BL21 (DE3) *E. coli* strain.

Identification of Library Hits

Due to the presence of a second plasmid also using a T7 promoter, DNA sequencing with T7for and T7rev was not possible on plasmid preps of the positive clones. Instead, PCR amplification was performed with DAPK1for and pET5 (5′ CTGGA AAGAA ATGCA CAAAC TTTTG CCATT CTC 3′) primers, T_m=62 °C and elongation at 72 °C for 2 min on isolated colonies. The PCR products were cleaned with Nucleospin kit (MachereyNagel) and sequenced with T7rev primers.

Protein Expression tests for DAPK1

Clones isolated from the library were expressed in 50 ml TB at 25 °C overnight induced with 100 µM IPTG. Cultures were pelleted at 8,000 rpm for 10 min and resuspended in buffer at pH 7 (50 mM NaH_2PO_4/Na_2HPO_4, 300 mM NaCl, 5 mM imidazole and 10 mM β-ME). The lysis was done by French press and soluble protein was separated by centrifugation at 15550 g for 20 min. The supernatant was incubated overnight with 50-100 µl Ni^{2+}NTA resin (Qiagen) at 4 °C. The resins were collected after centrifugation at 200 g for 1 min, resuspended in 5 ml lysis buffer, and loaded on a gravity column. One wash of 20 volumes with buffer containing 10 mM imidazole was performed and protein eluted in 2 bed volumes of buffer containing 300 mM imidazole.

Protein Expression tests for DAPK1-CaM

Clones isolated from the library were expressed in 50 ml TB and grown at 20 °C overnight after induction with 100 μM IPTG. The cultures were pelleted at 8000 rpm for 15 min and resuspended in buffer at pH 8 (50 mM HEPES, 300 mM NaCl, 2 mM $CaCl_2$, 5 mM imidazole and 10 mM β-ME). The cells were broken by sonication and soluble protein was separated by centrifugation at 15,550 g for 30 min. 100 μl Ni^{2+}NTA were incubated with the soluble fractions overnight at 4 °C. The Ni^{2+}NTA resins were collected by centrifugation at 200 g for 1 min and resuspended in 5 ml lysis buffer. They were loaded on gravity columns and washed first with 10 resin volumes of lysis buffer and second with 20 bed volumes of buffer with 15 mM imidazole. The protein was eluted with 2 bed volumes of buffer containing 300 mM imidazole.

Western Blot Immunofluorescent Detection

Western blots were blocked overnight at 4 °C with PBS 0.1% Tween (PBS-T) with milk powder 0.5% w/v. Then, they washed 3 times for 15 min with PBS-T. Firstly, incubation with 1/1000 dilution of rabbit anti-flag-tag primary antibody (50 μl in 50 ml of PBS Milk 0.5%) for 2 h at 4 °C in the roller blots was performed. After the removal of the antibody, the membranes were washed 3 times for 15 min with PBS-T. Second incubation was with 1/1000 dilution of goat anti-rabbit secondary antibody-Alexa633 conjugate and 1/5000 dilution of streptavidin-Alexa488 conjugate (50 μl of goat anti-rabbit and 10 μl of streptavidin in 50 ml of PBS-T) for 1 h at 4 °C in roller blots. Finally, the membrane was washed 3 times for 15 min with PBS-T and buffer removed by a single wash with ddH_2O for 5 min. The probed membrane was then scanned with Typhoon immunofluorescent imager.

Thermofluor Assay

A 20 μl reaction containing 8-10 μM of protein was set up in a 96-well plate PCR plate (Stratagene). SyproOrange (Invitrogen) at 5x final concentration diluted in buffer (100 mM HEPES, 150 mM NaCl, pH 7.5) was added to the reaction and the plate sealed with self-adhesive microseal (Bio-Rad). The reactions were mixed centrifuged briefly using a microplate compatible rotor A-2-DWP (Eppendorf). The fluorescence was read over the range of 25-75 °C (1 °C/min increment) with Mx3005P Q-PCR (Stratagene). The excitation and emission filters used were 490 nm and 575 nm respectively and all fluorescent changes were recorded and analysed in MS Excel.

Crystallisation Trials

DAPK1 1-300 and 1-304 with C-ter BAP were expressed at 25 °C overnight induced with 100 μM IPTG in 500 ml TB KH_2PO_4/K_2HPO_4. The cultures were pelleted at 6,000 g for 15 min and resuspended in buffer (50 mM Tris pH8, 300 mM NaCl, 5 mM imidazole 5 mM β-

ME, complete inhibitors and 1/1000 dilution Benzonase). Lysis was carried out in a microfluidiser with 10 passes of the lysate. Lysates were centrifuged at 76,500g for 30 min and the supernatant was incubated in fresh tubes with 500 µl of Ni^{2+}NTA (Qiagen) overnight at 4 °C. The Ni^{2+}NTA resins were collected by centrifugation at 200 g for 1 min and resuspended in 5 ml lysis buffer. They were loaded on gravity columns and washed first with 10 volumes of buffer (50 mM HEPES, 1 M NaCl, 10 mM imidazole and 1 mM DTT, pH 8) and then with 20 volumes of buffer (50 mM HEPES, 100 mM NaCl, 50 mM imidazole and 1 mM DTT, pH8). The protein was eluted with buffer containing imidazole in steps of 100 mM, 200 mM, 300 mM and 400 mM. The purified fractions were centrifuged for 10 min 20,817 g at 4 °C and 5 mg of DAPK1 1-300 and 1-304 were incubated with 50 µl of TEV protease for 8 h at room temperature. Protein was dialysed to remove imidazole with 1000 times dilution in buffer (50 mM Tris pH 8, 100 mM NaCl, and 1 mM DTT) for 1 h at room temperature. Protein was subjected to size exclusion chromatography with S200 (Amersham) and collected fractions of the peak were incubated for 30 min with 50 µl Ni^{2+}NTA resins in order to remove TEV protease. After centrifugation, the supernatant was recovered and concentrated to 2.21 mg/ml for DAPK1 1-300 and 3.06 mg/ml for DAPK1 1-304. The protein was submitted for crystallisation trials with sitting drops over six sets of Hampton research conditions in 384-well plates (HTX platform, EMBL, Grenoble).

Dynamic Light Scattering

Dynamic light scattering (Malvern Instruments) was performed with approximately 0.5 mg/ml protein at 25 °C with multiple iterations (usually 10) reading the counts of molecular species per volume.

"*If you try you may fail, if you don't try you're*

guaranteed to fail." --- Jesse Jackson

Chapter 4: IKKβ & TBK1

Résumé

Une caractéristique commune pour IKKβ et TBK1 est un domaine de type ubiquitin. Avec l'expansion de nos connaissances sur l'ubiquitin l'appréciation de son implication dans diverses signal augmente processus de réglementation. IKKβ est une protéine bien étudié, a trouvé une place centrale pour l'activation de NF-κB pathway. Il semble être strictement réglementé avec de nombreux mécanismes liés à son activation ou répression. Malgré tous les efforts pour sa caractérisation structurale, aucun d'autre élément ont été encore surexprimé soluble. Néanmoins, des études ont déjà découvert de nombreux inhibiteurs de l'interrompre à son activité. TBK1 est beaucoup moins étudiée. La question de sa participation dans la boite de signalisation NF-κB reste sans réponse. Troncation 3′ a été subi à la fois des protéines en raison de l'emplacement de leur domaine catalytique kinase à la N-terminus. Faible solubilité et des limites de domaine imprécis sont le résultat de criblage pour tous les deux. Grâce à sa haute valeur, IKKβ identifié des constructions correspond à un domaine catalytique kinase putative ont été poursuivis. Aucune méthode conventionnelle a tenté (diminution des niveaux d'expression, co-expression avec des chaperonnes, de repliement, la MBP-fusion et d'expression de cellules d'insectes) a abouti au sauvetage de son phénotype insoluble. Néanmoins, d'un criblage supplémentaire sur la protéine abouti à l'identification de constructions pour son ubiquitin like domain ainsi que des constructions similaires dans la région de liaison avec NEMO, précédemment décrit contraignant. Malgré, les caractérisations biophysiques démontrer que la protéine est monodisperse et monomérique, la cristallisation sur le meilleur comportement résultant des constructions dans les précipités. Une étude de protéolyse ménagé à démontré des troncatures N-terminal du domaine. Est sensiblement, une construction plus compacte a été sous-cloné et a trouvé de bien se comporter. Jusqu'à ce jour, aucuns cristaux n'ont été obtenus. Enfin, dans l'étude in vivo a été réalisée sur des concepts identifiés par ESPRIT correspondant au domaine catalytique IKKβ-kinase, ULD et NBD. Il a été constaté qu'ils influencent la voie NF-κ B ainsi, confirmant leur intégrité structurelle et fonctionnelle.

Introduction

While ubiquitin is typically recognized by the cell as a marker for degradation targeting proteins to S26 proteasome, it has been also linked to various non-degradative functions such as mediating protein interactions and regulating signal transduction (Adhikari, *et al.*, 2007,Liu, *et al.*, 2005). Intriguingly, ubiquitin-like domains (ULD) and ubiquitin binding domains (UBD) are found within a variety of different proteins although their function is unclear (Hartmann-Petersen and Gordon, 2004). Relative to this work, ULDs have been implied important for the activity regulation of kinases such as IKKα, IKKβ, IKK*ε* and TBK1 (Ikeda, *et al.*, 2007,May, *et al.*, 2004). They share a similar domain arrangement with an N-terminal catalytic domain followed by a ULD (Fig. IV.I.I). These kinases are involved in NF-κB and interferon signalling pathways.

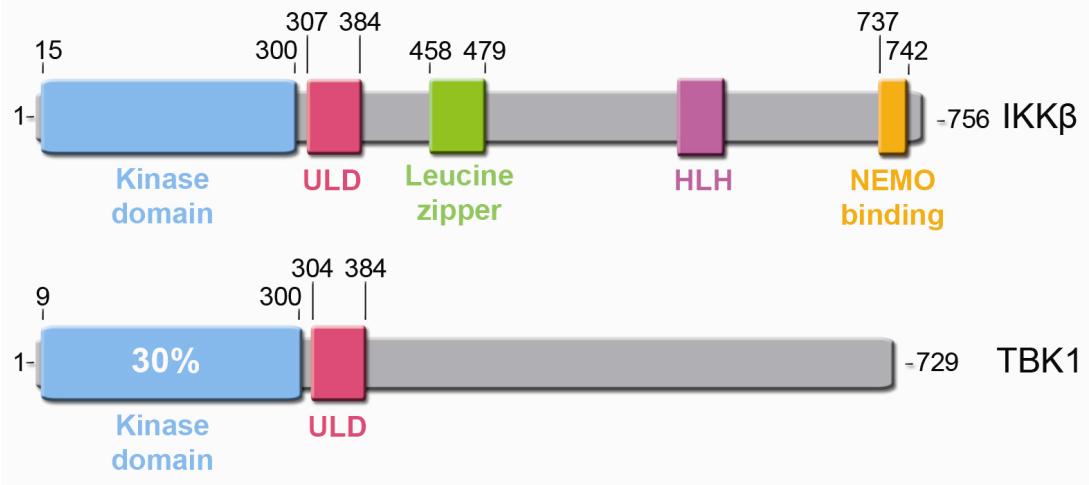


FIGURE 4.1: DOMAIN COMPOSITION OF IKKB AND TBK1. TBK1 SHARES 30% SEQUENCE IDENTITY WITH IKKB WITHIN ITS KINASE CATALYTIC DOMAIN.

Nuclear factor κB (NF-κB) signalling pathway controls cell proliferation, inflammation, apoptosis, as well as responses to different environmental exogenous and endogenous agents. It provides fast response to various stimuli like cytokines (TNF-α, IL-1), endotoxins (LPS)

and radiation damage since the transcription factor is present in the cytoplasm, ready for action, but in an inactive form sequestered in the cytoplasm through interaction with the Inhibitor κB (IκB) proteins (Karin, *et al.*, 2004). The signalling pathway is only partially understood, but significant progress has been recently made over the so-called canonical NF-κB signalling (Fig below).

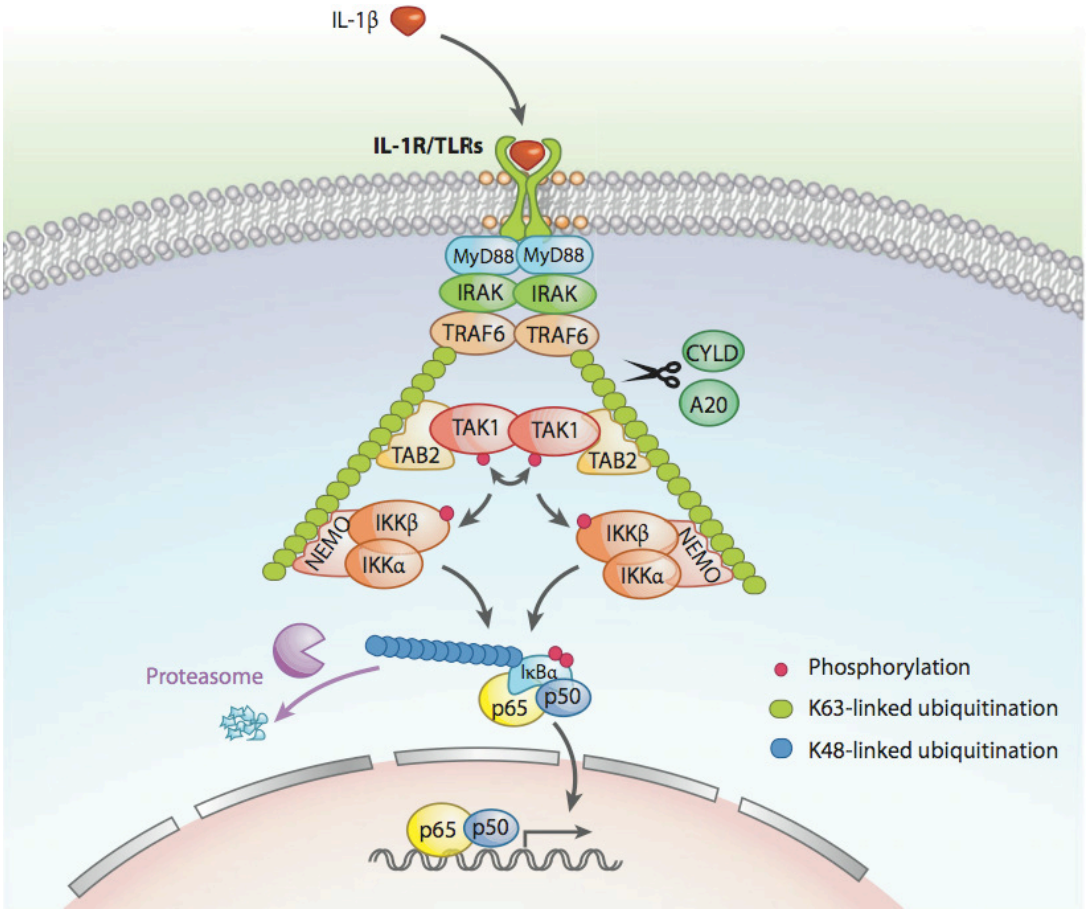


FIGURE 4.2: TLRS MEDIATED CANONICAL NF-KB SIGNALLING PATHWAY (XU, ET AL., 2009).

The latest model suggests the formation of large protein complexes differing for each of the receptors. Upon binding of a specific ligand, adaptor proteins recognize the receptor and (in the case of TLRs - TRAF6) are polyubiquitinated by K63-linked ubiquitination. Then,

TAB/TAK1 and IKK complexes are assembled into a complex. It has been suggested that the polyubiquitin network has scaffolding role and that TAK1 activates the IKK complex (Li, *et al.*, 2009,Xu, *et al.*, 2009). Recent reports suggest that the ubiquitin chains may be unanchored but otherwise confirm the order of events and modifications as previously suggested (Xia, *et al.*, 2009). Additionally, the activation of IKK appears to be receptor-dependent. Once IKK is activated, its Inhibitor κB kinase β (IKKβ/IKK2) phosphorylates IκBα at two N-terminal serines Ser32 and Ser36. Following phosphorylation, IκBα is ubiquitinated at adjacent lysines Lys20 and Lys21 by E3 ubiquitin-protein ligase. Ubiquitin-labeled IκBα is then degraded by the 26S proteasome so that the NF-κB dimer becomes free for translocation into the nucleus. There it binds to the recognition sequence GGGRNNYYCC (R-purine; Y-pyrimidine) and triggers the transcription of cytokines, adhesion molecules, and others (Karin, *et al.*, 2004).

IKKβ and its regulation

IKKβ is an 86kDa Ser/Thr kinase identified as one of the kinases that phosphorylates IκB. It was identified as part of a larger 700 kDa complex, termed Inhibitor κB Kinase (IKK), phosphorylating IκBα in stimulated HeLa cells (Chen, *et al.*, 1996). The complex consists of a regulatory protein "NF-κB essential modulator" (NEMO/IKKγ) and the highly similar Inhibitor κB kinase α (IKKα/IKK1) (with 65% identity within the catalytic domains) (Krappmann, *et al.*, 2000). Both kinases homo- and heterodimerise via their leucine zipper motifs (Kwak, *et al.*, 2000) but the major contribution in classical NF-κB signalling was attributed to IKKβ (Woronicz, *et al.*, 1997). The essential role of this kinase was illustrated by the observation that IKKβ knockout mice died at the embryonic stage between day 12.5 and 14 and exhibited sever liver degeneration from apoptosis that could not be rescued by mutating TNFα receptor (Li, *et al.*, 1999).

The region of interaction of the two IKK kinases with NEMO was identified in their C-termini (Yamaoka, *et al.*, 1998). Subsequent work showed different affinities of this interaction for the two kinases with IKKβ binding more strongly to NEMO than IKKα (May, *et al.*, 2002). In May 2008, the structure of the NEMO binding domain (NBD) of each of the kinases in complex with a fragment of NEMO was reported (Rushe, *et al.*, 2008) with the IKK part of the structure chemically synthesized and not overexpressed. The structure shows elongated coiled coils associating via hydrophobic interactions between NEMO and the IKKβ NBD, the latter mediating contacts via conserved C-terminal aromatic amino acids.

There are many modes of regulation reported for the IKK complex. Reports link IKKβ activation loop tyrosine phosphorylation with c-Src (Funakoshi-Tago, *et al.*, 2005). The presence of a classical mitogen activated protein kinase kinase (MAPKK) recognition sequence of SXXXS in the activation loop also suggests regulation by MAPKKs (Chen, *et al.*, 1999). Another level of regulation is redox based where, in the presence of reactive oxygen species (ROS), IKKβ dimerises through Cys179 of its catalytic domain. This prevents activation by upstream phosphorylation of the two serines in its activation loop because of the resulting steric clash (Ahmad, *et al.*, 2006). An inhibitor exploiting this mechanism has been reported (Harikumar, *et al.*, 2009). It was also reported that IKKβ undergoes autophosphorylation at its C-terminus (Delhase, *et al.*, 1999). The mechanisms of IKKβ activity regulation are diverse and complicated, and those presented above are just a selection. One of the recent models on IKK activation is presented (Fig.XXX).

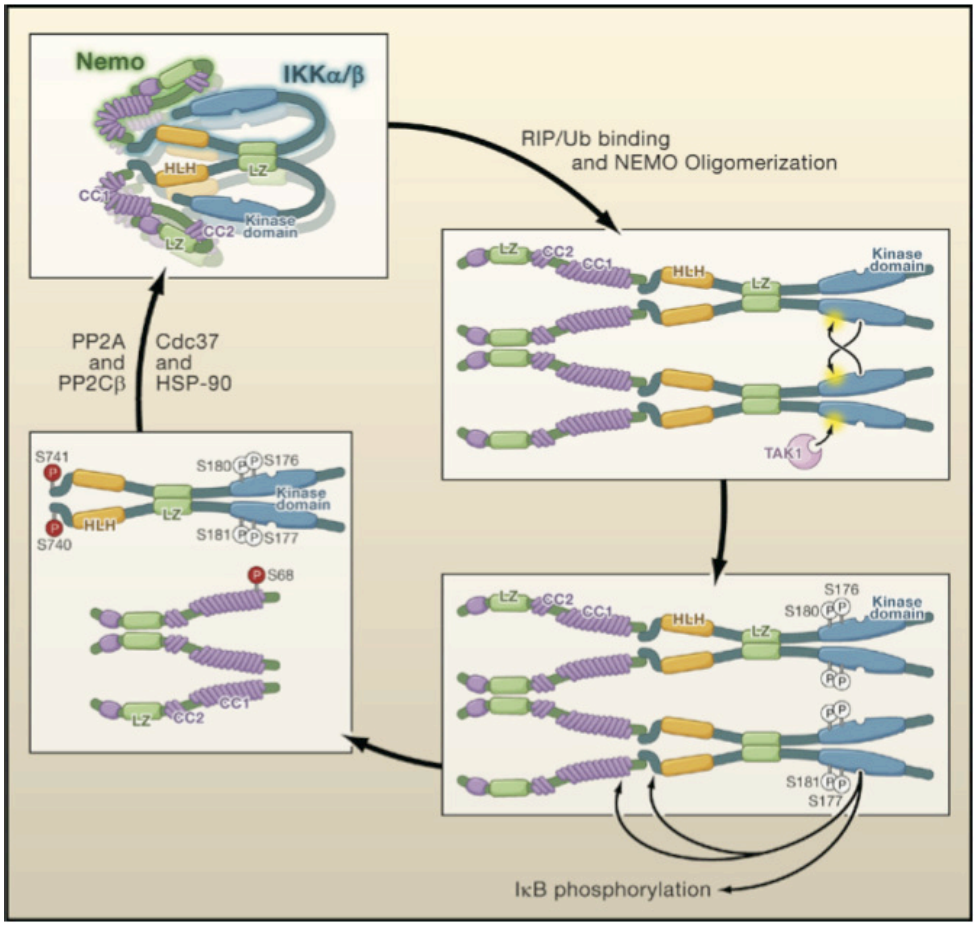


FIGURE 4.3: ACTIVATION MODEL OF IKK (HAYDEN AND GHOSH, 2008).

In active state, the complex is closed with the two kinases dimerised through their leucine zipper motifs and NEMO capping them. Upon ubiquitin binding and NEMO oligomerisation TAK1 phosphorylates the kinases at their activation loop and activates them. The active kinases autophosphorylate their C-termini resulting in disruption of the interaction with NEMO and their full activation. The inactivation and assembly in inactive complex is thought to be mediated by PP2A and other phosphatases. Another component important in the transition of the different states seems to be the heat shock protein 90 (Hsp90) which has been found to associate with the kinases (Broemer, et al., 2004).

Numerous studies have been carried out on IKKβ mostly in mammalian culture cells and a series of inhibitors have been developed to target inflammation and autoimmune disorders (Bhagwat, 2009). Nevertheless, despite much effort, structural characterization has been hindered by the low quantities of expressed protein (typically in insect cells) and the

difficulty in identifying soluble constructs for over-expression in *E.coli*. No structures of the IKK or IKK-related kinases have been published to date.

TANK-binding kinase 1 (TBK1)

TBK1 is an 83 kDa kinase similar in many respects to the IKKs and exhibits 30% identity within its catalytic domain. It is thus considered as an IKK-related kinase. It forms a complex with another, similar kinase, IKKε, and this complex resembles that formed by IKKα and IKKβ. Little is known about TBK1 but a role in interferon signalling pathways has been established although the exact mechanism remains to be elucidated. Some reports implicate it plays role also in the NF-κB signalling pathway since NF-κB dependent transcripts have been observed upon TBK1 activation (Harris*, et al.*, 2006,Tojima*, et al.*, 2000). Others argue that it exclusively plays a role in interferon production mediated by activation of interferon regulatory transcription factor 3 (IRF3) (Clark*, et al.*, 2009). Most probably, cross talk between the pathways exists. Because of the obvious similarity between the IKK and that of TBK1, it is also possible that it assembles in large receptor dependent protein complexes upon ligand binding.

In this work ESPRIT random construct screening was performed to obtain soluble protein from IKKβ and TBK1 with a focus on their kinase catalytic domains. An additional screen for regulatory domains of IKKβ was also performed in an attempt to generate constructs useful for subsequent investigations into the regulatory mechanisms of this remarkable protein.

Results

Kinase catalytic domain screen and characterization

Initial considerations

IKKβ and TBK1 share 28% sequence identity at the level of their catalytic domains. They are also similar to DAPK1 with 22% and 21% identity respectively. I-TASSER (Zhang, 2008) HSSP algorithm analysis of IKKβ and TBK1 suggested that they have highest sequence and structural similarities with the phosphorylase kinase (PDB code: 1PHK). The model returned for IKKβ is ranked with high similarity to twitchin-like protein kinase (PDB code: 1KOBA) and DRP-1 (PDB code: 1WMKE), a close relative of DAPK1. Similarly, the Swiss-Model structure prediction server finds similarity between DRP-1 and IKKβ, choosing 1WMKE as a model template. These analyses suggest strong structural similarities between IKKβ and Ca^{2+}/calmodulin dependent kinases and correlate well with its classification in the same family with CaMKI/II. The TBK1 homology model is closest to its template (phosphorylase kinase), a G coupled receptor kinase 6 (PDB code: 2ACXA) and CASK CaM kinase (PDB code: 3C0GA). HSSP for these proteins is of lower accuracy due to low homology with the best template found (phosphorylase kinase) sharing 28% identity. Thus, conclusions about structural details from such models cannot be as precise as the ones returned for DAPK1 but confirm the location of their catalytic domains.

A screen for soluble over-expressing constructs of IKKβ and TBK1 kinase catalytic domains was attempted. Due to confirmation of the N-terminally located kinase catalytic domains, similar to DAPK1, the 3′ unidirectional truncation screening approach was selected as successfully used previously.

Generation of IKKβ and TBK1 libraries

The gene encoding IKKβ was cloned into pYUM1001 resulting in a starting construct with hexahistidine tag fused in an invariable manner to the first codon and with restriction sites necessary for exonuclease truncation at the 3′ end. The truncation was performed and fragments separated into three size ranges corresponding to 0-230, 230-500 and 500-756 amino acid insert sublibraries (Fig. 4.4).

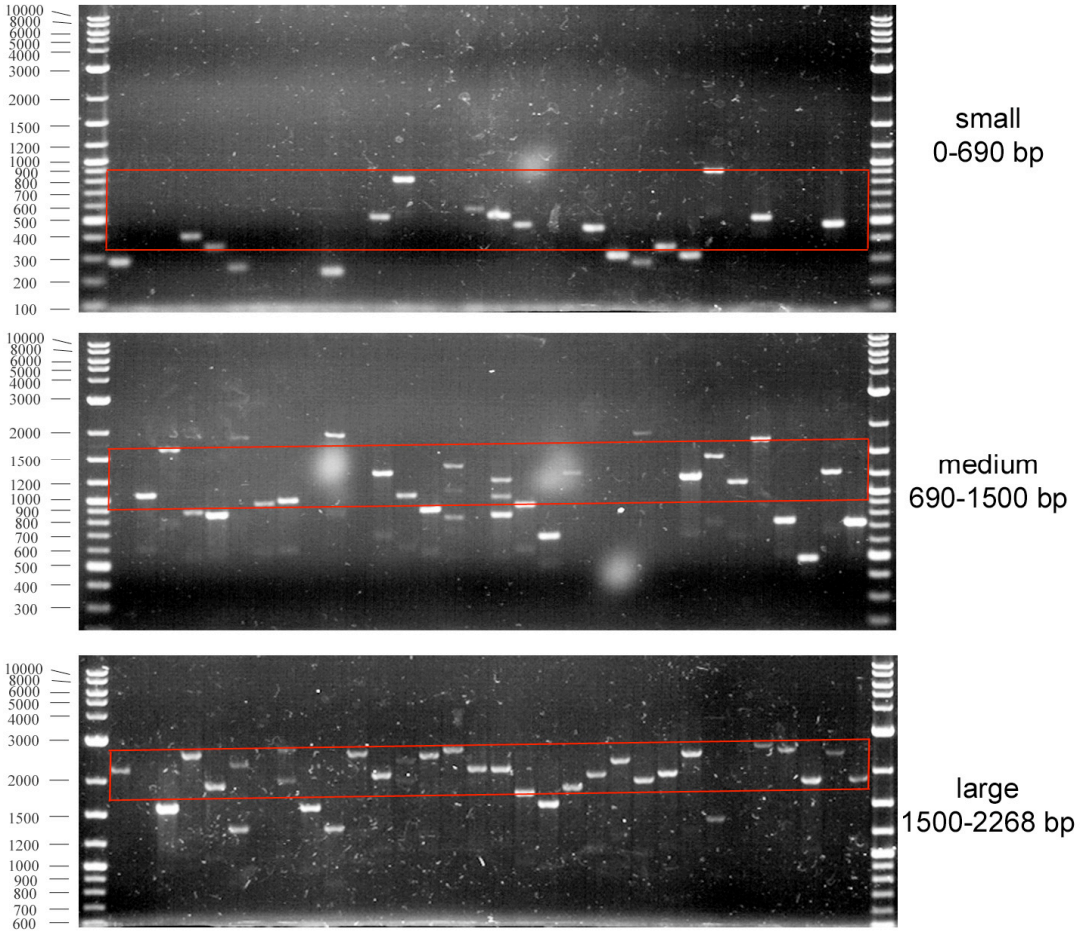


FIGURE 4.4: COLONY PCR SCREEN FOR THE EVALUATION OF IKKB 3′ TRUNCATION LIBRARY QUALITY.

32 of 96 tested per sublibrary are shown. Target gene insert size is calculated as PCR product size minus 320 bp of flanking sequence.

With TBK1 the strategy taken was similar to that with IKKβ, but performed on a sub-full-length construct comprising kinase catalytic domain, ubiquitin like domain and unassigned sequence as far as amino acid position 500. Additionally, the library was synthesized using a mix of two plasmids encoding TBK1 constructs starting at amino acids 1 and 12, the latter corresponding to deletion of a short predicted disordered region. Moreover, to eliminate potential problems of toxicity both of the TBK1 entry clones were the catalytically inactive mutant (D135N). Equal concentrations of both entry constructs were mixed for the exonuclease truncation reaction. Due to the shorter length of the starting construct, no size fractionation into sub-libraries was necessary and the entire library was tested (Figure 4.5).

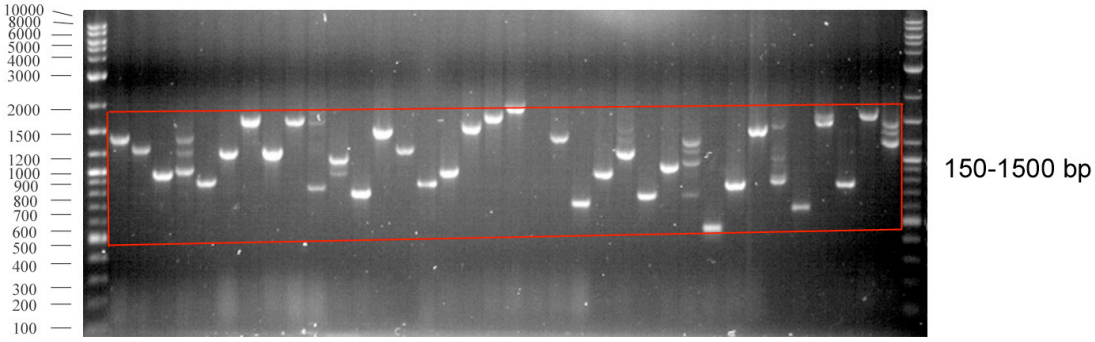


FIGURE 4.5: COLONY PCR SCREEN FOR THE EVALUATION OF TBK1 3′ TRUNCATION LIBRARY QUALITY.

36 of 96 tested per sublibrary are shown. Target gene insert size is calculated as PCR product size minus 320 bp of flanking sequence.

The plasmid libraries were plated out and 4608 colonies robotically picked for each of the IKKβ sublibraries and for the single TBK1 library. The PCR screen demonstrated differences in the quality of the libraries generated. The IKKβ library was of lower quality than TBK1 with a higher contamination of fragments outside of the desired range. Nevertheless, by picking more colonies from the IKKβ library to compensate a coverage of >90% was achieved (Table 4.1).

Protein	IKKβ			TBK1
Construct (aa)	1-756			1-500
Truncation Type	3′	3′	3′	3′
Size Range (bp)	0-690	690-1500	1500-2268	150-1500
Diversity	690	810	768	1350
Oversampling	6.7	5.7	6	3.1
Quality (%)	37.5	40.6	71.9	94.7
Contamination (%)	18.8	37.5	18.7	0
Completeness (%)	91.8	90.0	98.6	96.1

TABLE 4.1: SUMMARY OF THE GENERATION AND SCREENING OF IKKB AND TBK1 3′ TRUNCATION LIBRARIES.

*Completeness is calculated using server http://guinevere.otago.ac.nz/aef/STATS/index.html.

Solubility screens on IKKβ

The isolated clones were gridded on nitrocellulose membranes and tested for their biotinylation level. The first 96 clones ranked on biotinylation from the largest and most efficient sublibrary (1500-2268 bp corresponding to protein fragments 500 aamino acid to full-length with the catalytic domain expected at approx 280 aa) were selected for the second level of screening by purification from 4 ml expression cultures. Ni^{2+}NTA purification was carried out with the liquid handling robot and purified proteins were analyzed by 15% SDS-PAGE. No well purified proteins were observed. The protein was either not expressed or the yields were too low to observe from such small cultures. Total lysates were analyzed by SDS-PAGE and western blot was performed on the purified proteins. It was observed that the fragments were very well expressed but of low solubility (Figure 4.6a). For several clones, small amounts of soluble material was purified (Figure 4.6b) and this was confirmed as IKKβ by streptavidin western blot. The identified fragments ranged between 15 kDa to 66 kDa in contrast to the DAPK1 model system where a precise cluster of purified proteins was

apparent. That a number of the clones were of sizes smaller than would be expected from the DNA fragment range analysed reflects the limited resolving power of agarose gel electrophoresis during the excision of bands from the gel after the exonuclease truncation step. Analysis of clones from small and medium insert size sublibraries (0-690 and 690-1500 bp) resulted in similar results (not shown) although with a greater abundance of protein fragments in the lower size ranges. As apparent from Figure 4.6 c (and subsequently in Figure 4.7, below) these slightly soluble constructs correlated to the expected size of the catalytic domain, and also the N-terminal lobe only.

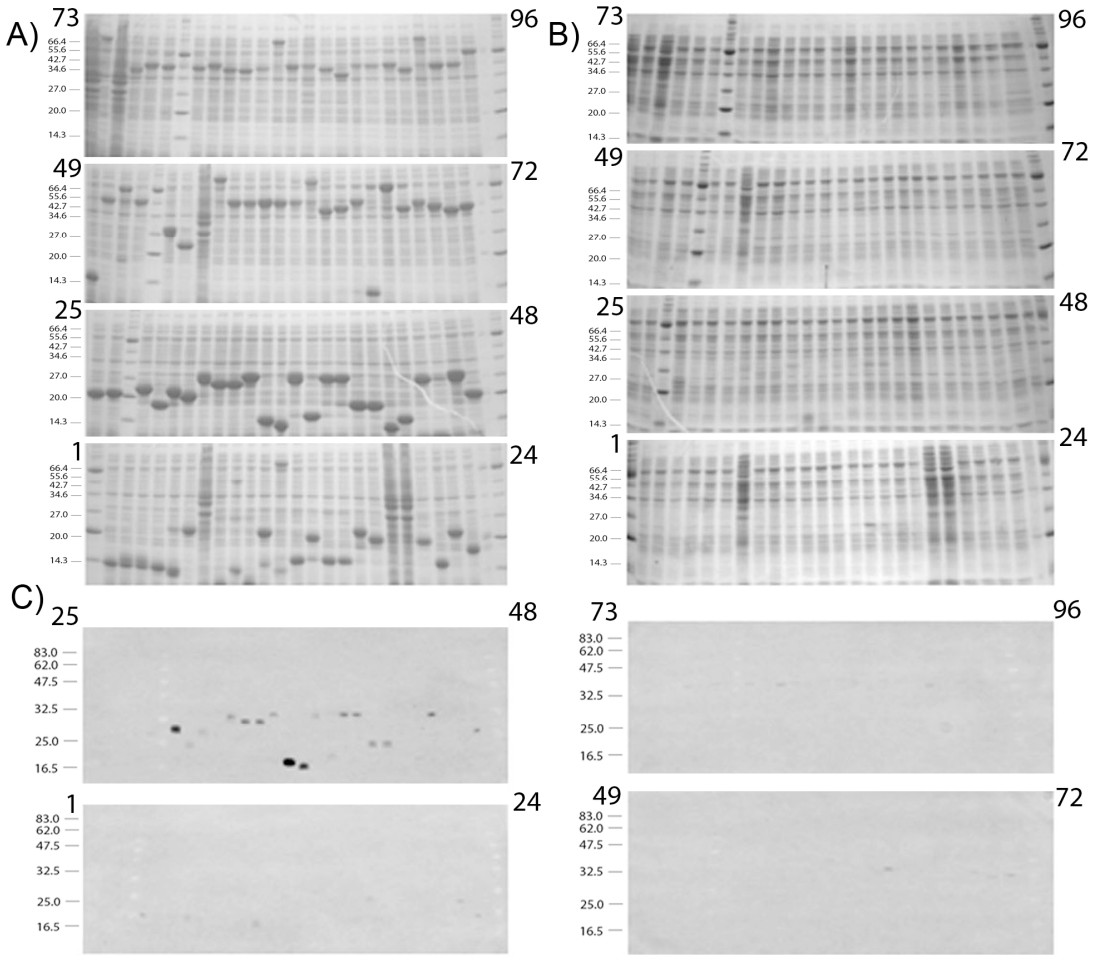


FIGURE 4.6: Ni^{2+}NTA PURIFICATION SCREEN ON IKKβ LARGE FRAGMENTS.
(A) total fractions; (B) purified fractions; (C) streptavidin western blot on the purified fractions. Numbers indicate ranking of the clones after the first part of the screen.

Rescreening libraries with coexpressed lambda phosphatase

Based upon a significant literature precedent, we reasoned that active kinases might be toxic to *E. coli* resulting in their absence in the library: transformed cells would not grow and so would not be picked and tested for expression (Wang, *et al.*, 2006). It was also possible that phosphorylation might negatively affect the solubility of the constructs. These potential problems were addressed by coexpression with the IKKβ library of the dual specificity, serine/threonine lambda phosphatase from a second pACYC184-derived coexpression plasmid. Competent cells expressing the flag-tagged phosphatase were transformed with the 3′ IKKβ library and processed through ESPRIT as previously. After purification screens no visible purified protein was observed by Coomassie-stained SDS-PAGE but streptavidin western blot again revealed the same pattern as previously with low quantities of soluble purified proteins mainly ranging between 15 kDa and 25 kDa (Figure 4.7). The larger fragments are approximately the size of the kinase catalytic domain while the smaller fragments correspond to the N-terminal lobe only of the catalytic domain.

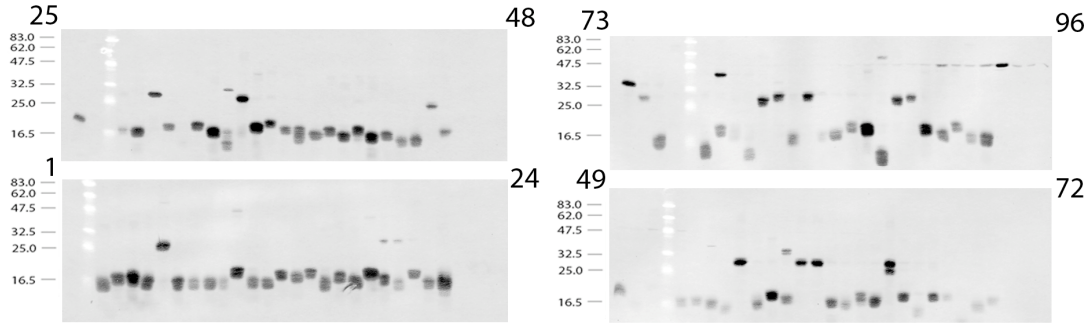


FIGURE 4.7: STREPTAVIDIN WESTERN BLOT ON PURIFIED PROTEINS FROM IKKB LIBRARY SCREENED WITH COEXPRESSED LAMBDA PHOSPHATASE.

A selection of clones detected by western blot was sequenced to identify the exact domain boundaries of the slightly soluble expression clones. Although no precise domain definition

was observed and some among the inserts appeared matching approximately the size of half of the kinase catalytic domain (Fig. XXX).

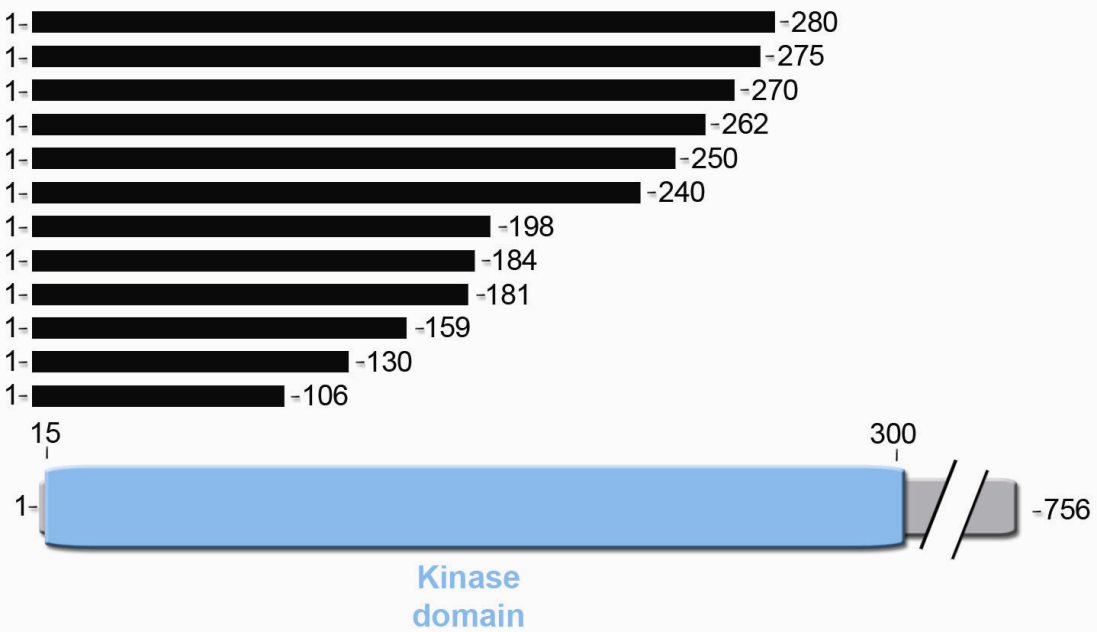


FIGURE 4.8: SELECTED CONSTRUCTS SEQUENCED FROM IKKB SCREEN MAPPED ONTO ITS DOMAIN PREDICTION.

A large distribution of sizes can be observed with long constructs corresponding to approximately a putative kinase catalytic domain and smaller constructs that would comprise the N-terminal lobe plus additional sequence. (Illustration is not in exact scale)

Solubility screen on TBK1

Thirty-two well-biotinylated clones were selected from the fluorescent colony blots and submitted to the second step of purification screening. The results were similar to that of IKKβ with a number of slightly soluble purifiable clones corresponding to the predicted size of its N-terminal lobe (Table 4.3). Although the library was performed on a mix of two plasmids encoding alternative start positions, all clones selected started at residue 12 and not

at the native amino-terminus, perhaps representing better translation efficiency from this position.

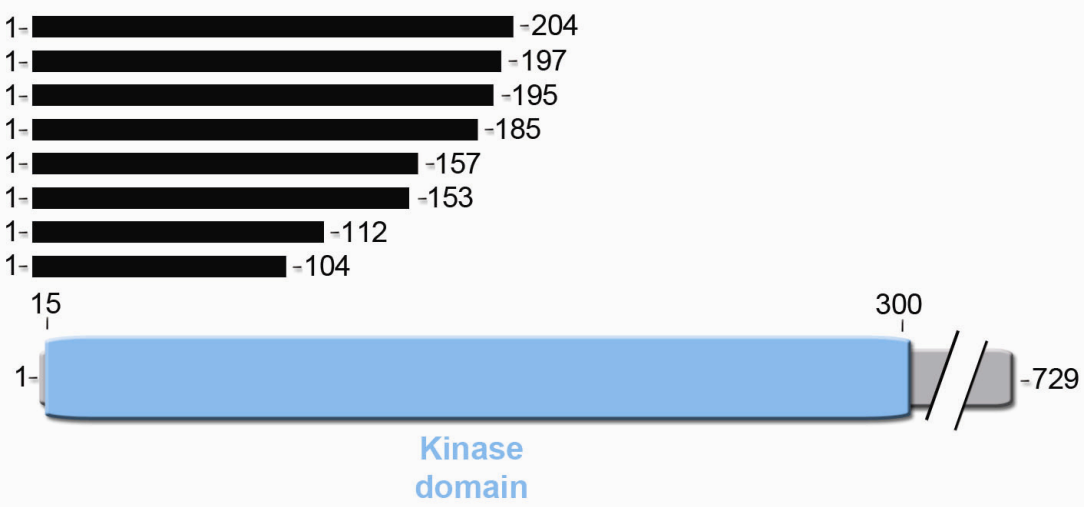


FIGURE 4.9: SELECTED CONSTRUCTS SEQUENCED FROM TBK1 SCREEN MAPPED ONTO ITS DOMAIN PREDICTION.

Similar to IKKβ, a large distribution of sizes can be observed. (Illustration is not in exact scale)

Due to faint signals and little visibility of the purified protein, the results were confirmed by scaling up the expression for five of the longest clones to 1 L TB expression cultures. Ni^{2+}NTA purification was performed with two-step elution collecting two elution fractions at 300 mM imidazole. No inclusion bodies were observed and there was a small amount of purified protein only visible by streptavidin western blot (Figure 4.8).

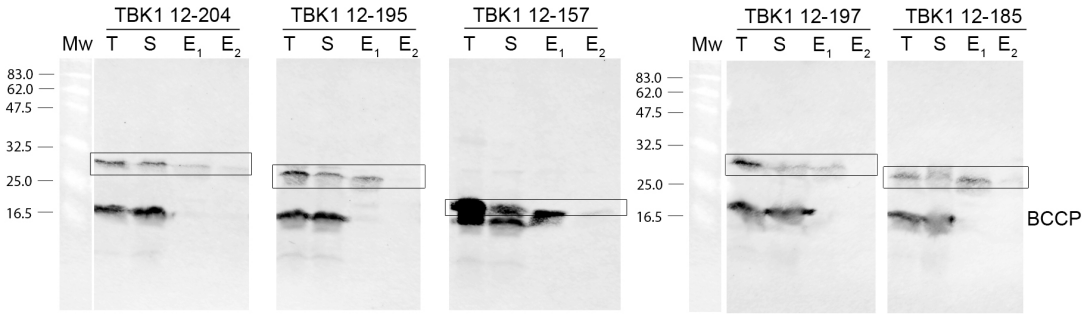


FIGURE 4.10: STREPTAVIDIN PROBED WESTERN BLOTS ON PURIFIED TBK1 FROM 1 L EXPRESSION CULTURE (IN RECTANGLES)..

Legend: T-total fraction, S-soluble fraction, E1-first elution, E2-second elution

To summarise the above sections, the library-based construct screening of IKKβ and TBK1 did not result in the identification of soluble over-expressing kinase catalytic domain constructs. IKKβ exhibited strong inclusion body formation with small amounts of catalytic domain and N-terminal lobe, whilst TBK1 only yielded low levels of soluble protein corresponding to the N-terminal lobe only. Both of the screen results suggest that the N-terminal lobe of the kinase catalytic domains may be autonomously folded in the absence of the rest of the domain.

Attempts to rescue the solubility of the partially soluble IKKβ catalytic domains

It was decided to focus upon the longest IKKβ constructs (residues 1-275 and 1-280) identified from the screen as slightly soluble in the presence of large inclusion bodies. A number of strategies were attempted briefly summarised below.

Reducing expression levels in E. coli

High levels of over-expression are commonly a cause of inclusion body formation. Tuner strains of *E. coli* with a mutant Lac permease were tested with no obvious decrease in expression level upon IPTG titration (not shown). The expression levels of IKKβ 1-275 were successfully decreased with in an induction time course at 30 °C (Figure 4.10A). The catalytic domain remained largely insoluble but purified material was faintly visible by SDS-PAGE. Quantification of the fluorescent signal from the Alexa488 streptavidin western blot demonstrated that the highest amount of soluble and purified protein was obtained after 30 min induction at 30 °C and that it decreased after that point.

In the same experiment, lysis in a series of additives was tested. Urea in combination with 1 % TritonX-100 (Figure 4.10 c, lanes A & B) resulted in slight a solubilisation effect but improvements were only marginal.

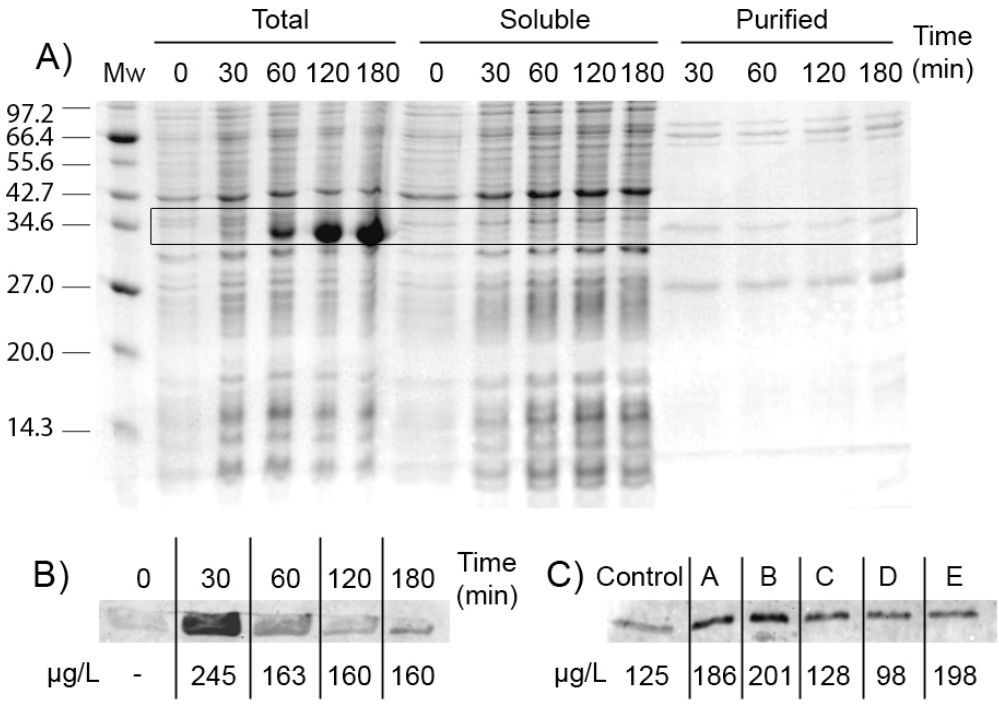


FIGURE 4.11: EFFECT ON SOLUBILITY OF IKKB 1-275 UPON TIME COURSE OF THE INDUCTION AND ADDITIVES.

(A) Coomassie-stained SDS-PAGE for expression at 30 °C (total fractions normalized at 3 µg; soluble at 5 µg and purified at 1 µg); (B) western blot for quantification of the time course purified fractions with calculated estimation of the yield per litre of expression culture; (C) western blot for quantification of purified fractions prepared with additives upon lysis (Control – no additives, A – 1 M urea and 1 % TritonX-100, B – 2 M urea and 1 % Triton X-100, C – 1 M NDSB201 and 1 % Triton X-100, D - 20µM staurosporin and E - 100µM AMP-PNP);

Chaperone coexpression

Inclusion body formation suggests problems of folding and is commonly observed during expression of human proteins in *E. coli*. A screen of bacterial strains that over-express different combinations of chaperones in *E. coli* has been shown to improve soluble expression in some cases (de Marco*, et al.*, 2007). Vectors were obtained (gift of B. Bukau, Heidelberg, Germany) and coexpression strains generated for expression trials.

Coexpression of GroEL/GroES with IKKβ catalytic domain constructs resulted in a four-fold increase in purified soluble material. Increase in soluble yields was also observed with

DnaK/DnaJ/GrpE or in combination with ClpB and GroEL/GroES. This effect has already been observed in a large study on protein kinase coexpression with chaperones (Haacke, *et al.*, 2009). Coexpression with GroEL/GroES was scaled up and subsequently size exclusion chromatography revealed that approximately 20% of this solubilised material was in free form whilst the majority was chaperone-bound. Further attempts to release the chaperone by addition of ATP, or 2M urea were unsuccessful.

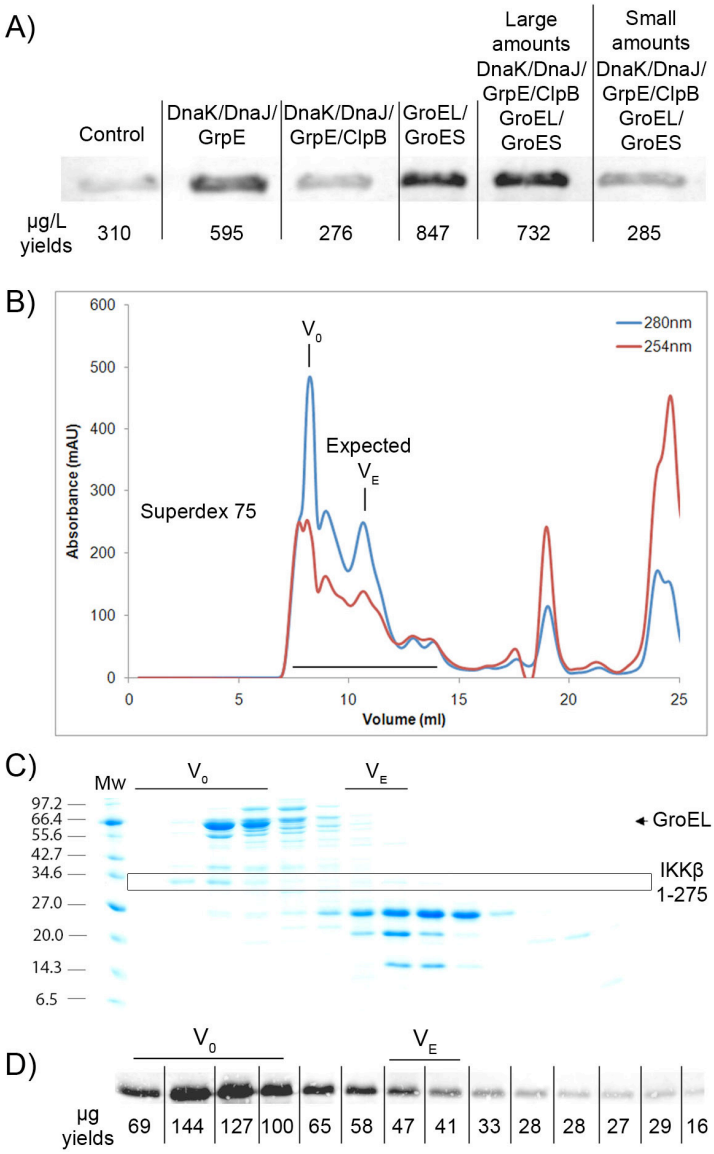


FIGURE 4.12: EFFECT ON SOLUBILITY OF IKKB 1-275 UPON COEXPRESSION WITH CHAPERONES.

(A) quantitative western blot on purified kinase catalytic domain from the screen with calculated estimation of the yield per litre of expression culture; (B) scale-up and size exclusion chromatography on coexpression with GroEL/GroES; (C) 15% SDS-PAGE gel on the size exclusion fractions; (D) western analysis of fractions confirmed majority of protein was coeluting with the chaperone although approximately 20% eluted as free monomer;

Refolding of the IKKβ catalytic domain inclusion bodies

A simple protocol for refolding of kinase domains (personal communication: P. Ramage, Novartis, Basel) was performed on purified inclusion bodies for IKKβ 1-275 resulting in milligram quantities of soluble material that remained stable for more than a week at 4 °C. The refolded material bound Ni^{2+}NTA resin and was easily affinity purified to homogeneity (Figure 4.12a). Despite this promising result, further analyses reveal large polydisperse soluble aggregates by DLS which remained unchanged upon incubation of the protein with various additives (Figure 4.12b). The protein eluted in the void volume of S200 size exclusion column (Figure 4.13c). Attempts to improve the result by applying a high throughput kinase refolding screen (Cowan*, et al.*, 2008) were performed in collaboration with Richard Cowan during a visit to AstraZeneca, Alderley Park, UK and similar results were achieved.

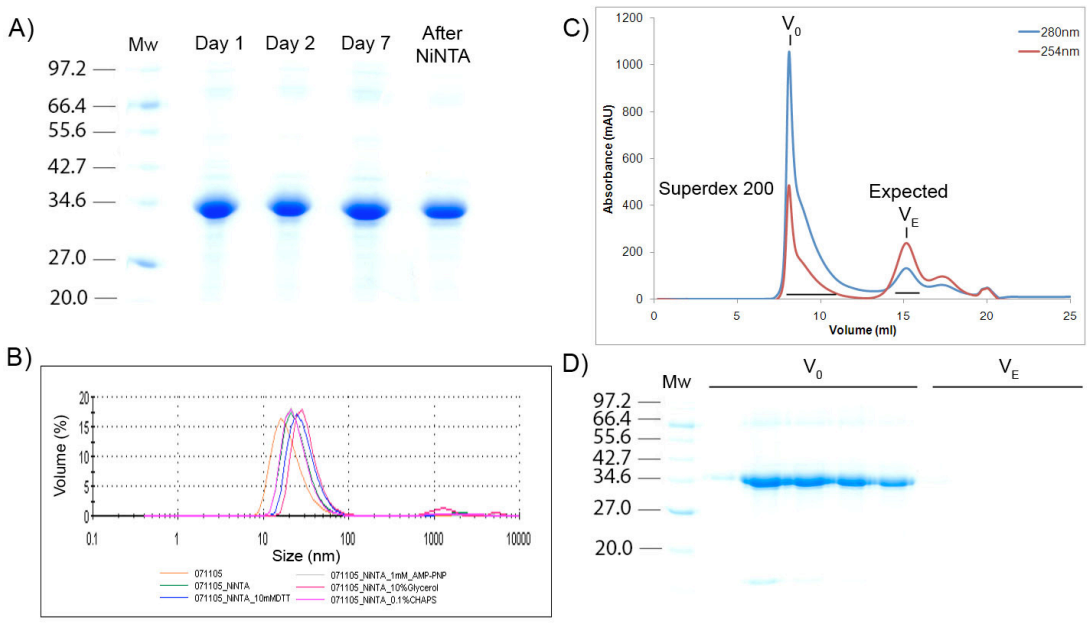


FIGURE 4.13: REFOLDING OF IKKB 1-275.

(A) SDS-PAGE gel of the refolded material shows stable and soluble product; (B) Analysis of the refolded material by dynamic light scattering with a screen of additives; (C) size exclusion chromatography on the refolded material; (D) SDS-PAGE analysis of the size exclusion chromatography confirms the presence of the refolded material in the void volume. Lanes in panel D correspond to fractions indicated in panel C;

MBP fusions & N-terminal variations

MBP is a common protein fusion used for purification and enhancement of solubility (Hammarstrom*, et al.*, 2002). Sequential N-terminal truncations of the IKKβ 1-280 kinase domain were fused downstream of a 6xHis-TEV tag or MBP fusion partner. While 6xHis-tagged proteins were well-expressed, none were soluble. On the contrary, MBP-tagged proteins were soluble and purifiable. Moreover, TEV cleavage resulted in a sufficient release of the IKKβ kinase domain from MBP (Fig. 4.13). Further analysis demonstrated that the protein was eluting in the void volume during size exclusion chromatography indicating the formation of soluble aggregates.

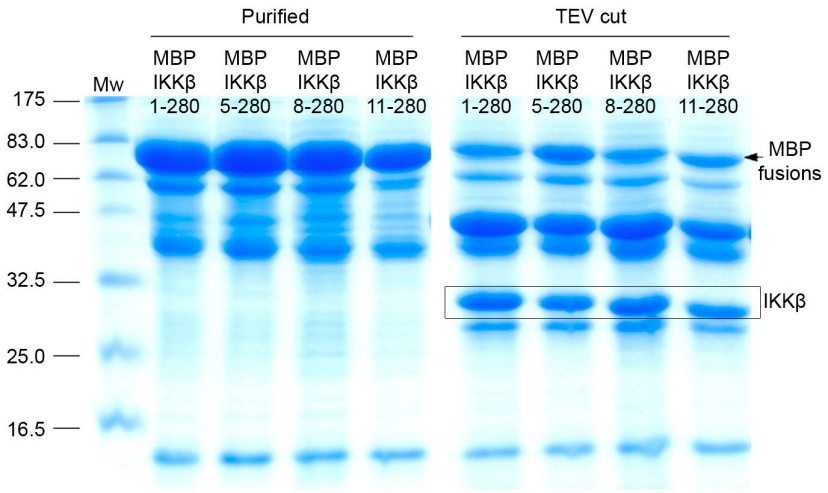


FIGURE 4.14: MBP FUSION ENHANCES THE SOLUBLE EXPRESSION OF IKKB.
Incubation with TEV results in kinase domain release but the protein appears to be a soluble aggregate.

Insect cell expression

Small amounts of full length IKKβ have been produced in insect cells (Pengelley, *et al.*, 2006) but the quantities were insufficient for structural studies. Insect cell expression of IKKβ 1-280 and 1-407 (kinase catalytic domain with ubiquitin like domain was attempted in collaboration with AstraZeneca. Separately insect cell expression was performed at EMBL Grenoble in the Eukaryotic Expression Facility (Fitzgerald, *et al.*, 2006). Both constructs expressed in both experiments, but with the majority of the protein in the insoluble fraction.

Summary of expression attempts of the IKKβ kinase catalytic domain

Although catalytic domain constructs of IKKβ identified through ESPRIT screening showed small levels of solubility and purifiability (but insufficient for structural biology studies), attempts to improve yields of purified protein were unsuccessful. The amount of purified protein after one-step Ni^{2+}NTA purification did not exceed 4-5% of the material found in the inclusion bodies. Attempts at processing larger volumes of culture were not successful, largely because the initial poor recovery coupled to the additional purification steps resulted in further sample loss. Current methods of crystallisation require at a minimum, several

milligrams of pure and homogeneous protein. Due to the low yields for the kinase catalytic domain of IKKβ, its structural characterization was not pursued further.

Screening IKKβ for additional domains

The IKKβ protein contains a number of internally located domains (Fig 4.1). These have been predicted from sequence analyses, or from biochemical experiments, but well expressing constructs suitable for structural studies have not been described. An ESPRIT library approach was applied to the identification of expression constructs encoding these soluble domains.

Library generation and quality

A bidirectional truncation library was performed on the gene encoding IKKβ (see Fig 1.xxx). Similarly to the 3′ truncation library, the full-length gene was cloned into pYUM1001 with flanking *Aat*II/*Asc*I and *Not*I/*Nsi*I restriction sites permitting exonuclease truncation from each end. The reactions were performed and DNA fragments encoding protein fragments from 200-400 amino acids were excised from the agarose gel. This size range permits the identification of regulatory domains as well as testing the definition of both the termini of the kinase catalytic domain (Figure 4.4).

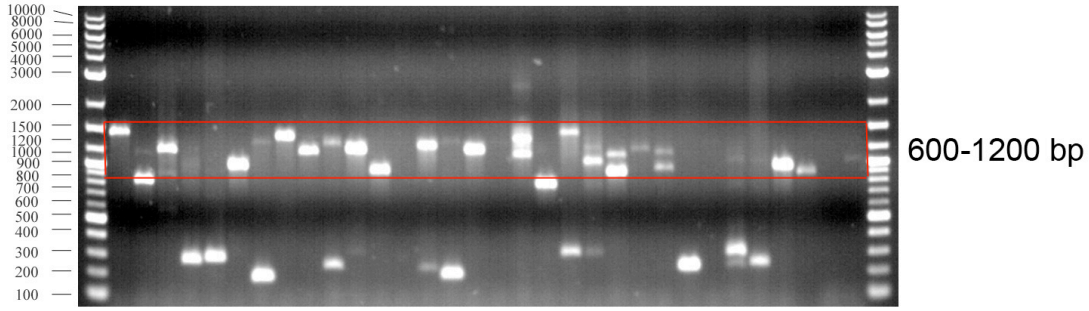


FIGURE 4.15: REPRESENTATIVE COLONY PCR SCREEN FOR THE EVALUATION OF IKKB BIDIRECTIONAL TRUNCATION LIBRARY QUALITY.

32 of 96 tested are shown. Insert size is equal to PCR product minus 320bp added with the primers for PCR screen.

The plasmid library was transformed in arabinose inducible *E. coli* BL21-RIL and 27 648 colonies were isolated with the picker robot. Due to the limited resolving power of agarose gels with this early generation (larger) ESPRIT vector (elaborated in chapter 2) 24% of the constructs in this library are smaller than the desired size range. Moreover, due to the nature of the library where both ends of the gene vary simultaneously, the level of diversity is much higher than the previous unidirectional truncation libraries resulting in 2.8% of all constructs being tested (Table 4.4).

	IKKβ
Construct (aa)	1-756
Truncation Type	bidirectional
Size Range (bp)	600-1200
Quality (%)	48.9
Contamination (%)	23.9
Diversity	455400
Completeness (%)	2.8

TABLE 4.2: SUMMARY OF THE GENERATION AND SCREENING OF IKKB BIDIRECTIONAL TRUNCATION LIBRARY.

*Completeness is calculated using server http://guinevere.otago.ac.nz/aef/STATS/index.html.

Solubility screens

The first 14 clones ranked according to their biotinylation levels on the colony array were selected for the second level of screening by purification directly from 50 ml expression cultures. Ni^{2+}NTA purification was performed and the purified proteins analyzed by 15% SDS-PAGE (Fig. 4.15). Two constructs were prominently soluble and purifiable in large amounts. Others appeared partially soluble with small amounts visible in the purified fractions.

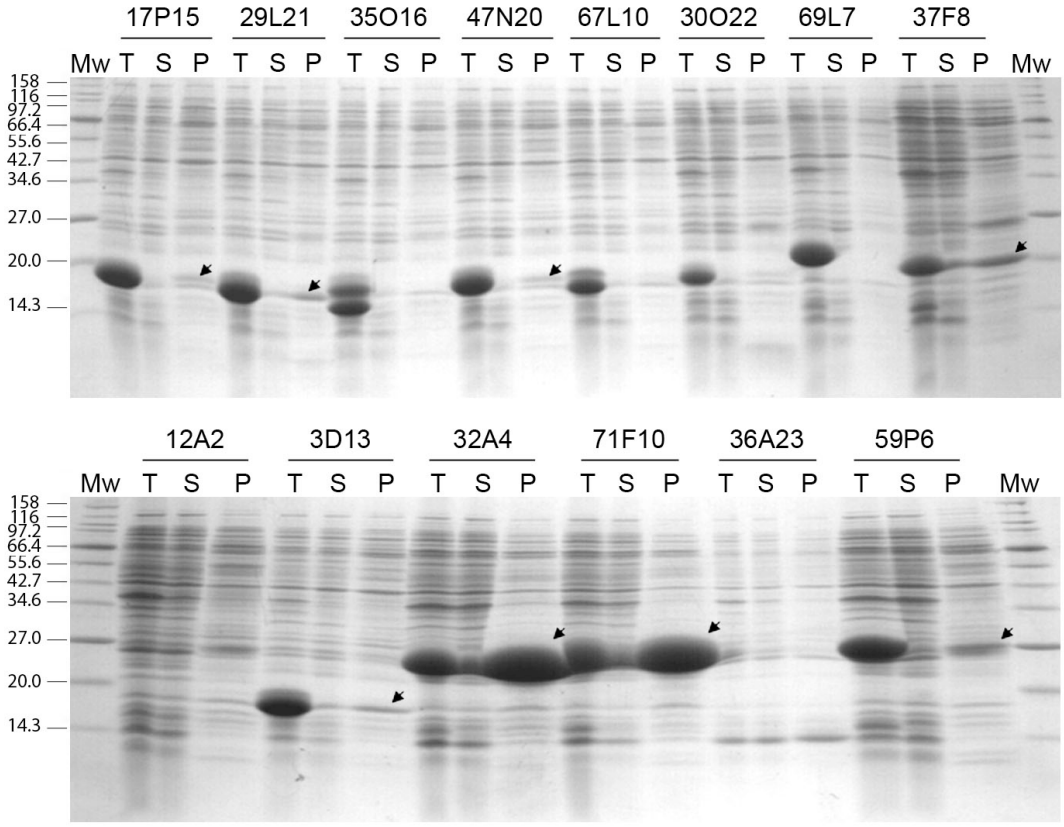


FIGURE 4.16: SDS-PAGE ANALYSIS OF THE FIRST 14 PROTEINS RANKED ON BIOTINYLATION FROM THE BIDIRECTIONAL TRUNCATION LIBRARY.

The purified proteins were in the size range 20-30 kDa and although an additional 96 clones were tested, no significantly longer constructs were identified. Sequencing of the positive clones reveals that they fall in three regions of the gene: i) N-terminal lobe of the kinase catalytic domain, ii) encompassing the ubiquitin-like domain (ULD) and iii) the NEMO binding region (Fig. 4.16). All of the aforementioned were present in the 14 clones initially tested and no significantly better-behaved constructs were found upon further testing of clones deeper in the library. Among them, ULD-containing constructs (32A4 and 71F10) were highly soluble and purifiable whilst the N-terminal lobe of the kinase catalytic domain (3D13) and the NEMO binding region (59P6) were partially soluble. The identification of constructs for the N-terminal lobe supports the hypothesis that the kinase catalytic domain of

at least some kinases can be split into separate lobes. This is the third instance where such constructs are isolated and the majority of these constructs are around 110 amino acids long. The constructs overlapping with the ULD domain identified in the central region of the protein are the most stable and well-behaved from IKKβ.

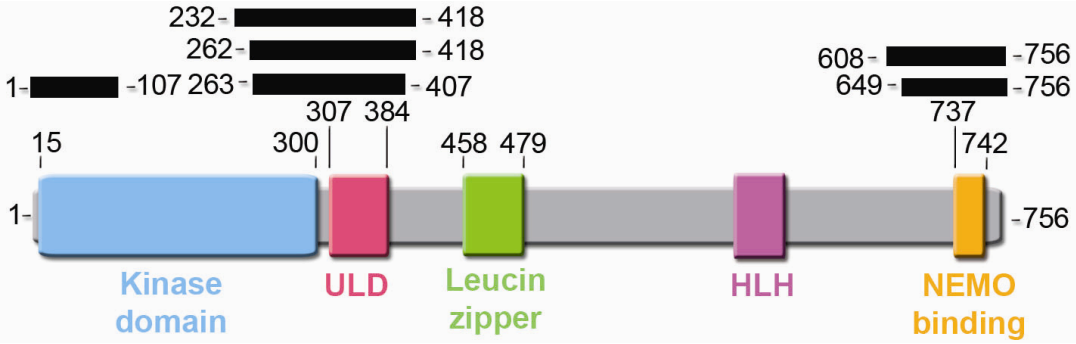


FIGURE 4.17: SOLUBLE EXPRESSION CONSTRUCTS IDENTIFIED BY BIDIRECTIONAL TRUNCATION MAPPED ONTO A LINEAR DOMAIN REPRESENTATION OF IKKB.

Purification of the constructs identified in the ULD region yielded 19 mg/L of culture for IKKβ 263-407 and 33 mg/L for IKKβ 262-418 (constructs after genetic deletion of the biotin acceptor peptide). TEV cleavage to remove the His tag from IKKβ 263-407 was only partial but complete for IKKβ 262-418. Size exclusion chromatography revealed that a significant amount of both purified proteins eluted as a single peak (Fig. 4.17). Furthermore, both proteins eluted earlier than the expected implying that they were either oligomers or elongated: IKKβ 263-407 eluted at 14.85 ml instead of 16.3 ml as expected for the 19.8 kDa predicted from the DNA sequence, corresponding to a globular protein of 35 kDa. Similarly, IKKβ 262-418 eluted at 14.8 ml (35 kDa) instead of 16.4 ml.

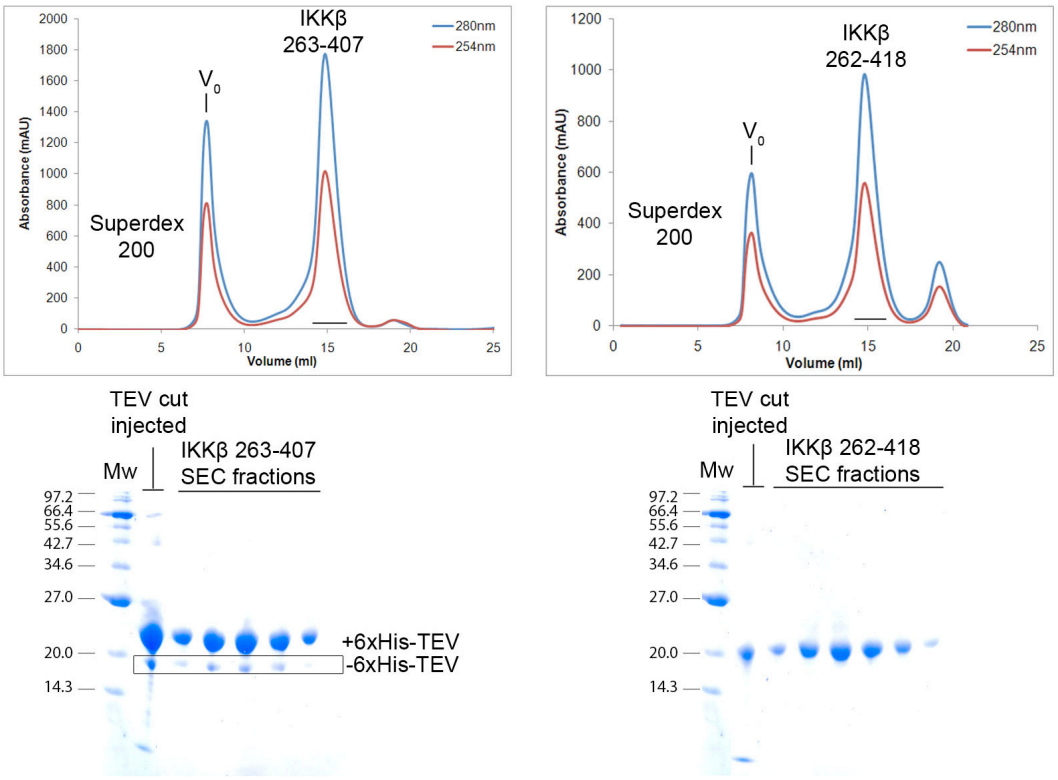


FIGURE 4.18: IKKB ULD 263-407 AND 262-418 CHARACTERISATION BY SIZE EXCLUSION CHROMATOGRAPHY.

Significant amounts of the protein eluted in volume fraction corresponding to a dimeric form.

Further characterization of ULD constructs using biophysical approaches

Circular dichroism was used to analyse the purified domains and indicated that they were folded. Deconvolution of the 195-260 nm region of the spectrum for IKKβ 263-407 suggested that there may be some disordered regions (30% random coil), perhaps in part due to the presence of the His tag and TEV site (Fig. 4.18A).

In order to define the oligomerisation state of this domain, two complementary techniques were applied to IKKβ 263-407 (19.2 kDa predicted from sequence): size exclusion chromatography multi-angle laser light scattering (SEC MALLS; with assistance from M.

Jamin, UVHCI) and small angle X-ray scattering (SAXS; with assistance from A. Round, beamline XXX, ESRF). SEC MALLS with a Superdex 200 column resulted in a peak at 15.56-16.99 ml. Subsequent analysis of the data demonstrated the protein was monodisperse, with an observed size of 22.1 kDa, similar to the monomeric size (Fig. 4.18B). SAXS measurements also indicated a monodisperse, monomeric, elongated domain rather than a dimer, as visualized with the *ab initio* model created by DAMMIN (Fig. 4.18C). This model exhibited a good fit to the experimental data (Fig. 4.18D) The maximum dimension (D_{max}) defined for the protein was 7 nm and molecular weight estimated through calibration protein BSA (Guinier) and molecular volume (Porod) is close to that expected for the monomer.

In conclusion, the ULD domains were elongated in shape, monomeric and monodisperse, the former explaining the size exclusion chromatography observation.

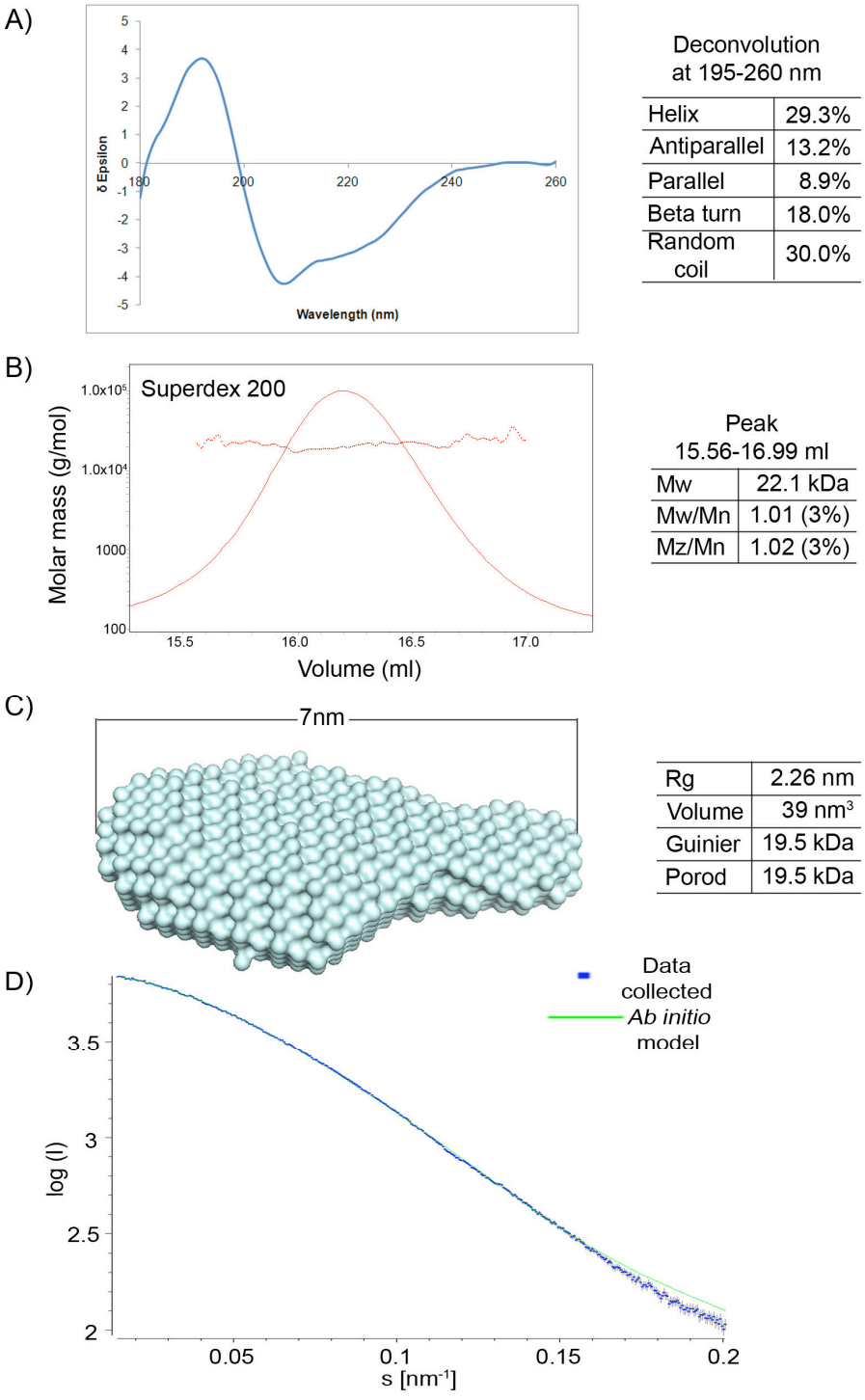


FIGURE 4.19: BIOPHYSICAL CHARACTERIZATION ON 6XHIS-TEV IKKB 263-407.

(A) Circular dichroism demonstrated the protein to be folded; (B) SEC MALLS showing monodisperse protein of 22 kDa; (C) SAXS ab initio model and measurements suggest an elongated protein of 19.5 kDa; (D) SAXS ab initio model shows good fit to the collected data;

Crystallization trials on IKKβ ULD constructs

Crystallization trials were performed on IKKβ 263-407 with suppression of its biotin acceptor peptide but with 6xHis-TEV due to the inefficient TEV cut. At high concentrations, precipitates were formed and at low concentrations, predominantly clear drops were observed. Similar phenotype was observed with crystallization trials on IKKβ 262-418 after removal of His tag and biotin acceptor peptide over a range of concentrations: 5 mg/ml, 12 mg/ml, 17 mg/ml and 30 mg/ml. Although, the proteins appeared very well monodisperse, no condition where crystals appeared was found.

Soluble domains can often support disordered termini and the fragments identified in the ULD region were considerably larger than the annotated ULD with CD suggesting the presence of some random coil. Therefore, limited proteolysis was attempted on the construct (with tags) to identify a more compact domain. Initial analysis by mass spectrometry confirmed the expected molecular weight and the cleavage of the initiator methionine (Fig. 4.19A). Elastase, trypsin and chymotrypsin were used at two different concentrations to interrogate the protein (Fig. 4.19B). While elastase and chymotrypsin both resulted in rapid degradation to small fragments of approximately 7 kDa, trypsin yielded a stable fragment at about 14 kDa. The trypsin digests were analysed by ESI TOF mass spectrometry (IBS service, Grenoble) identifying N-terminal proteolysis (Fig. 4.19C). Combining the results from limited proteolysis and library truncation, a construct IKKβ 302-407 was cloned in pYUM6002 with then-terminal 6xHis-TEV tag, expressed and purified to homogeneity. This construct did not aggregate as observed by size exclusion chromatography for the longer variants. Preliminary crystallization trials were performed with protein at 4 mg/ml and 5 mg/ml, but to date have been unsuccessful.

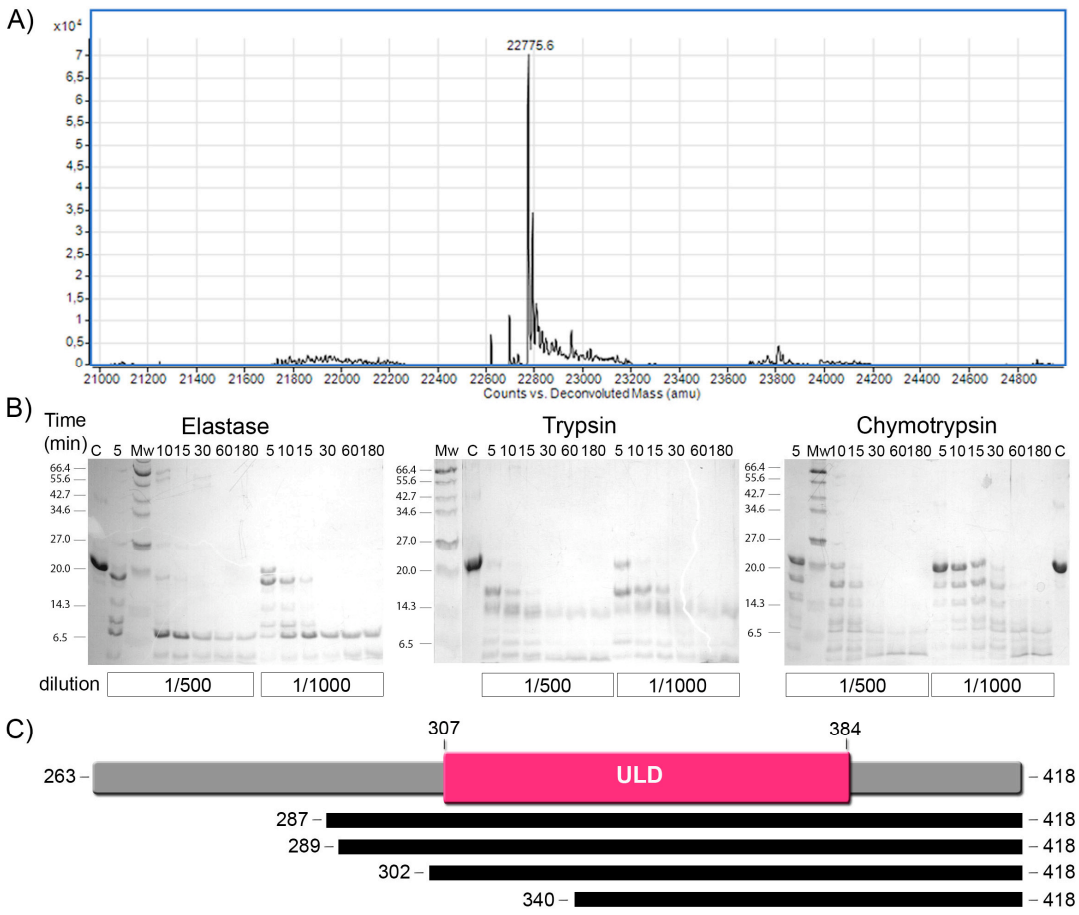


FIGURE 4.20: IKKB 262-418 LIMITED PROTEOLYSIS.

(A) mass spectrometry of the expressed protein showed it was of the size predicted from DNA sequencing (B) time course of elastase, trypsin and chymotrypsin digestion at two different dilutions (C – undigested control) (B) constructs identified by mass spectrometry after 30 minutes digestion with 1/500 dilution of trypsin at 37 °C.

In summary of this section, the bidirectional truncation library of IKKβ did not identify kinase catalytic domain constructs, but regions comprising the N-ter lobe were partially soluble suggesting that splitting of the kinase catalytic domain is possible. The NEMO binding region constructs were also soluble and were in agreement with previously identified constructs in cell-based studies (May, *et al.*, 2000,May, *et al.*, 2002), but were not characterized further. Finally, the best-behaved constructs contained the annotated ULD

region (May*, et al.*, 2004). Biophysical analyses confirmed them to be intact, folded, elongated proteins, which after several trials did not result in crystals. Limited proteolysis revealed more compact constructs but subsequent trials again did not yield crystals.

In vivo functional studies

Constructs identified from the truncation libraries corresponding to IKKβ kinase catalytic domain (1-275), ULD (262-418) and NEMO binding domain (649-756) were cloned in pIRES vectors in order to study the effects of their over-expression in eukaryotic cells on NF-κB signalling (collaboration: M. Pasparakis, Cologne, Germany). Although proteins were not visible by western blot, the constructs exhibited pronounced effects in the assay. It was found that expression of the kinase catalytic domain construct led to activation of NF-κB and expression of the luciferase reporter at low concentrations of transfected plasmid (Fig. 4.20). In contrast, over-expression of the ULD and NEMO binding domain lead to inhibition of signalling. This observation correlates with previously reported inhibition effects of the similar fragments (May*, et al.*, 2000,May*, et al.*, 2004). The *in vivo* study shows that the fragments are functional and therefore presumably natively folded in mammalian cells. Further work will be necessary in order to understand the full mechanistic significance of these data.

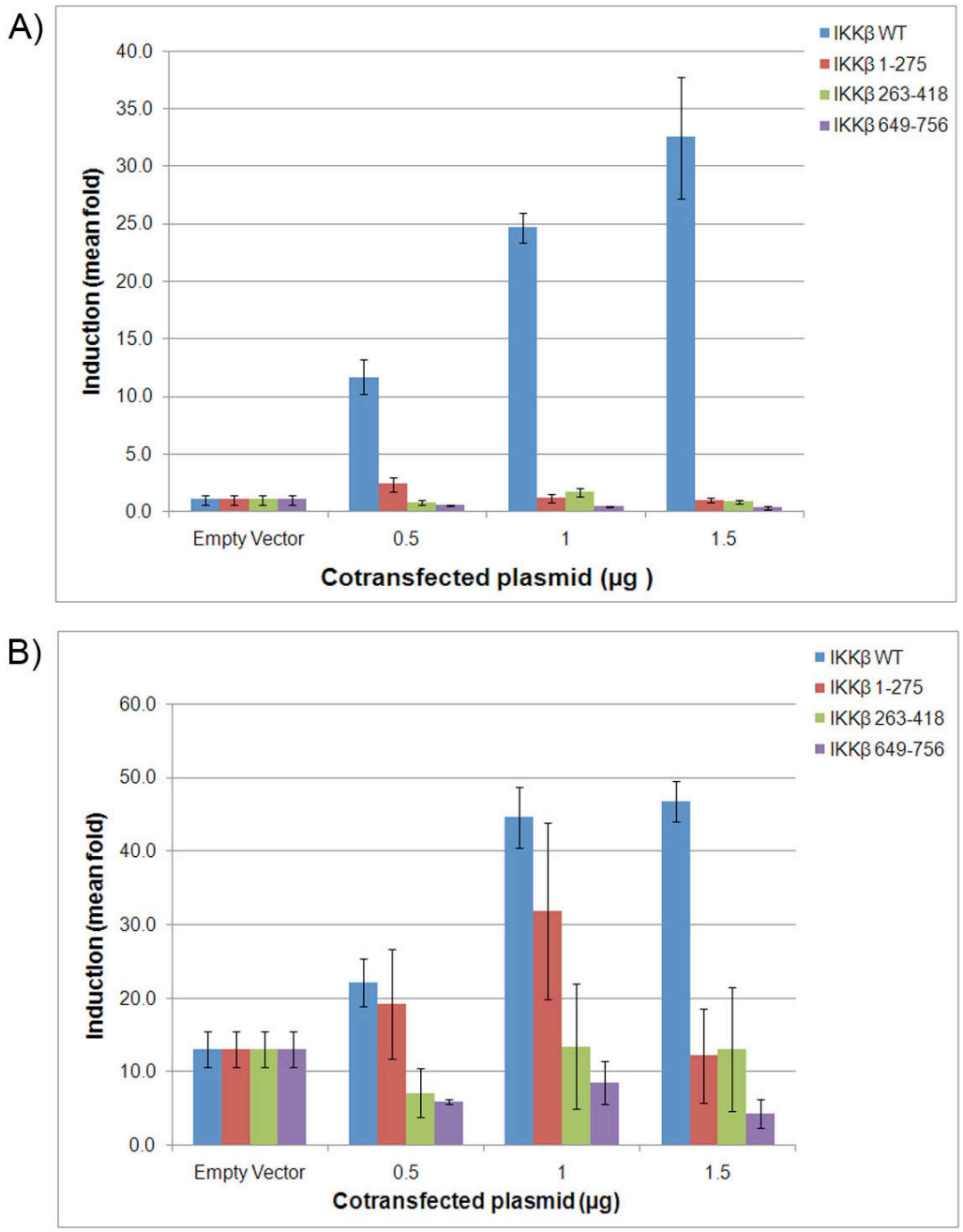


FIGURE 4.21: LUCIFERASE REPORTER ASSAY ON CONSTRUCTS OF IKKB IDENTIFIED FROM ESPRIT SCREENING.

(A) over-expression in unstimulated HEK293 cells with control IKKβ wild-type (WT); (B) over-expression in stimulated HEK293, 24h post transfection;

Discussion & Conclusions

Although much studied due to their medical importance, no structures from the IKK-like kinases have been reported to date, with the exception of a short helical 48 aa region from the IKK C-ter in complex with a similar-sized small fragment of NEMO (May, *et al.*, 2000). Sequence analysis and homology modelling clearly established the N-ter location of the catalytic domain, and strong similarities with DAPK1 suggested that no obvious barrier should exist to its expression. However, extensive construct screening by ESPRIT of IKKβ and TBK1 did not yield overexpressing soluble kinase catalytic domain constructs despite oversampling of all C-ter boundaries. Small amounts of IKKβ catalytic domain were purified from clones that expressed the majority in an insoluble form, although the yields were insufficient for setting up crystallisation trials. Attempts to increase the proportion of soluble material by optimising expression parameters, purifying with additives and using chaperone strains made only small differences. Analysis using size exclusion chromatography of the most promising constructs coexpressed with GroEL/GroES at low temperature showed that the purified protein was irreversibly bound to the chaperones. Refolding results were initially promising since multimilligram quantities of purified soluble protein were obtained, but it was later shown to be polydisperse and so unsuitable for crystallisation trials. There is a report that IKKβ associates with the Hsp90 chaperone *in vivo* and this may reflect an intrinsically low stability of this protein (Broemer, *et al.*, 2004). Moreover, IKKβ is not found a monomer *in vivo* but always in a complex with partner proteins that may provide stabilising interactions. This work is probably the most extensive construct testing ever performed on this target and the results strongly indicate that the catalytic domain of IKKβ (and TBK1) cannot be expressed soluble at high yield in *E. coli*.

The screen of a bidirectional truncation library (of mainly internal gene fragments) demonstrated that other soluble domains distal to the catalytic domain could be identified. N lobe constructs from were identified for both IKKβ and TBK1 suggesting that these fold autonomously from the rest of the domain. By far the best constructs identified from these screens were those spanning the ULD. It expressed well and purified at very good yields, and circular dichroism demonstrated it to be folded. While it appeared dimeric by size exclusion chromatography, complementary methods SEC MALLS and SAXS (ESRF ID14-3 beamline) both confirmed it as monomeric with the size exclusion chromatography explained by an elongated shape. Neither the longer construct identified in the screen or a shorter one defined by limited proteolysis have yet crystallised.

In an attempt to derive further information on the constructs identified from IKKβ, we collaborated with a cell biology laboratory (M. Pasparakis, Cologne) and transfected the fragments as plasmids into HEK293 cells using a standard luciferase assay. Down-regulation of NF-κB-dependent gene expression of the reporter was observed for constructs 262-418 (ULD) and 649-756 (NEMO Binding Domain) suggesting that these were competitively inhibiting the natural protein interactions. The kinase catalytic domain also appeared to exert inhibitory effects. Further studies are required to establish the significance of these results, although they hint that the constructs identified by ESPRIT screening may have native-like structure.

Materials & Methods

Open Reading Frames

Open reading frames corresponding to the following proteins were used: full length wild-type IKKB_human (Swiss-Prot Accession Number: O14920) and synthetic TBK1_human (Swiss-

Prot Accession Number: Q9UHD2) 1-500 aa. Genes were provided by AZ with synthesis of the latter at Geneart.

TBK1 Library Entry Clones

Two constructs (2-500 aa and 12-500 aa) of the synthetic TBK1_human gene were cloned in the ESPRIT vector pYUM6002. They were amplified by overlap extension PCR in order to add H6TEV tag and provide *Nde*I restriction site for cloning. The first PCR amplifying the gene was done with 400 nM gene specific primers TBK1for1 (5′ ACTAC CGAGA ATTTG TATTT TCAGG GGCAG AGCAC CAGCA ACCAT CTGTG 3′), TBK1for2 (5′ ACTAC CGAGA ATTTG TATTT TCAGG GGTCT GATAT TCTGG GCCAG GGCGC 3′) and reverse TBK1 rev1 (5′ GCCTC ATTAC GCCGC GGCCG CGAGC TCATC AGAAA TTTCG CCCAG 3′) with *Nsi*I restriction site amplified by *Pwo* polymerase in 25 µl reaction for 10 cycles with T_m=57 °C and 3 min of elongation at 72 °C on 10 ng of synthetic gene DNA. Amplification of 1 µl product of the first PCR reaction as template was prepared in the same manner but with 400 nM H6TEVfor (5′ CCATG ATTAT GATAT TCCAA CTACC GAGAA TTTGT ATTTT CAGGG G 3′) instead of DAPK1for1 and T_m=56 °C. The final amplification with 1 µl of the last reaction was done with 200 nM PB2synthfor3 (5′ GAGAT ATACA TATGG GCCAC CATCA TCACC ACCAT GATTA TGATA TTCCA ACTAC CGAGA 3′) (*Nde*I restriction site) as forward primer with T_m=57 °C, 5 min of elongation at 72 °C and 20 cycles of amplification in 50 µl total reaction volume. The PCR product was gel purified from 1% gel with QIAEXII kit. Both PCR product and 2 µg of pYUM6002 was cut in buffer 2 (NEB) with 20 U *Nde*I for 1h at 37 °C and buffer exchanged with Nucleospin kit (MachereyNagel). A second cut in buffer 3 (NEB) with 20 U *Nsi*I was performed for 1 h at 37 °C for both. The vector was dephosphorylated with 1 µl of SAP phosphatase for 1 h at 37 °C added in the last reaction. Both were buffer exchanged, ligated and transformed. The plasmids were sequence verified.

IKKβ Library Entry Clones Cloning

The wild-type IKKB_human gene corresponding to construct 2-756 was cloned in pTAR007 (bidirectional truncation entry vector created by Dr. Franck Tarendeau) by restriction digest cloning with *Asc*I and *Not*I. The wild-type gene (15 ng) was amplified by *KOD* polymerase in a 25 µl reaction supplemented with 1 mM $MgSO_4$ for 30 cycles with Tm=57 °C and 2 min of elongation at 72 °C with 500 nM of each IKK2for1 (5′ TAATG AGGCG CGCCG AGCTG GTCAC CTTCC CTGAC AAC 3′) (adding *Asc*I restriction site) and IKK2rev1 (5′ TTACGC CGCGG CCGCT GAGGC CTGCT CCAGG CAG 3′) (adding *Not*I restriction site). 2 µg of purified PCR product were digested consecutively with 10 U of *Asc*I in 30 µl reaction supplemented with BSA followed by buffer exchange (Qiagen) and second digestion with 10 U of *Not*I, again in 30 µl. Both reactions were incubated for 1 h at 37 °C and terminated by heat inactivation at 65 °C for 25 min. 2 µg of pTAR007 was digested with 10 U of *Asc*I for 1 h at 37 °C. The restricted plasmid was dephosphorylated by addition of 2 µl of SAP for 30 min. After buffer exchange it was digested with 10 U of *Not*I. The vector was dephosphorylated with 2 µl of SAP phosphatase for 1.5 h at 37 °C and was buffer exchanged with Nucleospin (MachereyNagel).

It was ligated and OMNIMax competent cells transformed. The sequence verified bidirectional truncation plasmid was used to generate a derivative in-frame construct with an N-terminal 6xHis-TEV entry clone for 3′ truncation library by cutting with *Asc*I and *Aat*II restriction enzymes, blunt-ending and religation. The resulting plasmid was sequence verified.

Library Generation and Testing

The library generation was performed largely as described in chapter 2 (*see* ESPRIT Protocol, Materials & Methods section) with the following modifications. IKKβ bidirectional truncation was performed with two consecutive exonuclease-mediated truncations. For the first, 6 µg of the entry plasmid were digested with 30 U of each *Asc*I and *Aat*II for 2h at 37 °C to linearise the plasmid. Then, it was buffer exchanged with the Montage kit (Whatman) and 2.7 µg of DNA were used for th exonuclease III reaction supplemented with 16 mM NaCl to limit the deletion rate to 38 bp/min. 1.2 µl of reaction mix were quenched in 2 M NaCl each min to inhibit the reaction. After evaluation of the quality of the first truncation library, plasmid was pooled from 14,400 colonies. 6 µg were digested with 12 U of *Not*I and *Nsi*I for 2 h at 37 °C. The exonuclease III reaction was prepared with 3µg of the linearised plasmid and performed as the first truncation. Finally, an arabinose indicible BL21 (DE3) codon plus (RIL) *E. coli* strain was transformed with the bidirectional truncation library.

The IKKβ 3′ truncation entry plasmid (6 µg) was linearised with *Not*I and *Nsi*I at 37 °C for 6 h. After buffer exchange with Nucleospin, 3.7 µg were digested with exonuclease III supplemented with 16 mM NaCl at 22 °C. 1.8 µl were quenched per min in 2 M NaCl. The same strain as above was transformed with the library.

Each of the TBK1 entry clones (10 µg) were digested with 80 U of *Nsi*I and *Not*I enzymes at 37 °C. 2 µg of the linearised vectors were submitted to time course exonuclease III reaction limited to 25 bp/min truncation by 44 mM NaCl at 22 °C quenching 2 µl of reaction every min in 3 M NaCl at 70 °C. Subsequent steps are as described in the ESPRIT protocol in chapter 2. The library was tested in the IPTG inducible BL21 (DE3) *E. coli* strain.

Time Course Expression Test

Arabinose inducible BL21 (DE3) codon plus (RIL) cells transformed with the IKKβ 1-275 expression vector were cultured in 500 ml TB expression broth until OD_{600} 0.5. A 100 ml aliquot of induced cells (time point 0 min) was taken. Protein expression was induced with 0.1% w/v arabinose at 30 °C. 100 ml aliquots were collected after 30, 60 120 and 180 min. The purification steps were then performed as described in the appendix (*see* Typical *E. coli* expression test).

Test of Additives to the Purification

Arabinose inducible BL21 (DE3) codon plus (RIL) cells transformed with the IKKβ 1-275 expression plasmid was cultured in 500 ml TB expression broth until OD_{600} 0.4. Protein expression was induced with 0.1% w/v arabinose at 30 °C for 4 h. The culture was split in even volumes and pelleted. Pellets were resuspended with buffer (50 mM Na_2HPO_4/NaH_2PO_4, 300 mM NaCl, 10 mM β-ME and protease inhibitors). One of the pellets was kept aside as a reference and for the remainder, five conditions were tested by addition of: 1 M urea and 1% Triton X-100; 2 M urea and 1% Triton X-100; 1 M NDSB 201 and 1% Triton X-100; 20 μM Staurosporin; and 100 μM AMP-PNP. The purification steps were then performed as described in the appendix (*see* Typical *E. coli* expression test).

Chaperones Coexpression Test

E. coli chaperone strains (de Marco*, et al.*, 2007) were transformed with the IKKβ 1-275 expression plasmid. Expression was performed in 50 ml TB cultures for the chaperones and a control (IPTG inducible BL21 (DE3) codon plus RIL transformed with the same plasmid). Cells were grown at 37 °C until OD_{600} 0.4, then continued at 20 °C until OD_{600} 0.8. Protein expression was then induced with 0.1 mM IPTG and cultures grown overnight at 20 °C. Purification was performed as in appendix (*see* Typical *E. coli* expression test).

Refolding

Cell pellets from IKKβ 1-275 expression cultures were resuspended with 50 mM Tris pH 7, 5 mM DTT, 5 mM EDTA and 5 mM Benz-HCl. They were sonicated for 10 min with 12 μs pulses, then centrifuged for 30 min at 15,550 g. The inclusion bodies were washed four times by resuspension in the same buffer until the supernatant was visibly clear. Then they were washed twice with water supplemented with 5 mM DTT. Estimated 685 mg of protein were dissolved to final 10 mg/ml in 8 M urea, 50 mM Tris pH 8.5, 100 mM DTT and 5 mM EDTA. Approximately 20 mg of protein were diluted rapidly in 200 ml ice cold buffer 500 mM Tris pH 8 supplemented with 5 mM DTT. The solution was stirred slowly at 4 °C overnight. Then, it was concentrated to 10 ml with in a spin concentrator (Millipore) with 20 kDa cut-off. After centrifugation at 10,000 rpm for 10 min it was dialysed overnight at 4 °C against 2 l of 50 mM Tris pH 8 supplemented with 5 mM DTT. After a second dialysis for 1 h against 5 l of the same buffer the amount of protein remaining in the solution was evaluated. The extension of this method to a high throughput screen is comprehensively described in Cowan *et al* (Cowan*, et al.*, 2008).

IKKβ Cloning in pIRES Mammalian Expression Vectors

All constructs were prepared by restriction digest with *Not*I at the 5′ and *EcoR*I at the 3′. Constructs were amplified with and without N-ter 6xHis tag where after the first round of amplification with gene specific primer, the tagged constructs were then amplified with second primer IRES.HisNotIfor (5′ CGATA TCTGC GGCCG CGTCA TGGGC CACCA TCATC ACCAC CAT 3′) in a similar way to previously described. Below is a table providing the gene specific primers used for the PCR amplifications. All constructs were sequence verified.

Construct	**With 6xHis tag**	**Without 6xHis tag**
IKKβ 1-275	IRES.IKK2CDhisfor 5′GGGCCACCATCATCACCACCATA GCTGGTCACCTTCCCTGACAACG 3′ IRES.IKK2CDrev 5′TAGTGGATCCACTGAATTCCGTCT CACTTCTCCAGTCGCTCAGCCAGG 3′	IRES.IKK2CDfor 5′CGATATCTGCGGCCGCGTCATGAGC TGGTCACCTTCCCTGACAACG 3′ IRES.IKK2CDrev 5′TAGTGGATCCACTGAATTCCGTCTC ACTTCTCCAGTCGCTCAGCCAGG 3′
IKKβ 262-418	IRES.ULD12hisfor 5′GGGCCACCATCATCACCACCATA ATAATCTTAACAGTGTCCTGGCTGA GC 3′ IRES.ULD12rev 5′GATCCACTGAATTCCGTCTCACTT GGGCTCTTGAAGGATACAGCTGA 3′	IRES.ULD12for 5′CGATATCTGCGGCCGCGTCATGAAT AATCTTAACAGTGTCCTGGCTGAGC 3′ IRES.ULD12rev 5′GATCCACTGAATTCCGTCTCACTTG GGCTCTTGAAGGATACAGCTGA 3′
IKKβ 649-756	IRES.NBDhisfor 5′GGGCCACCATCATCACCACCATA AGCGGCAGAAGGAGCTCTGG 3′ IRES.NBDrev 5′GTGGATCCACTGAATTCCGTCTCA TGAGGCCTGCTCCAGGCAGC 3′	IRES.NBDfor 5′CGATATCTGCGGCCGCGTCATGAAG CGGCAGAAGGAGCTCTGG 3′ IRES.NBDrev 5′GTGGATCCACTGAATTCCGTCTCAT GAGGCCTGCTCCAGGCAGC 3′

pFastBac Cloning

DNA encoding IKKβ 1-280 and 1-407 were subcloned from IKKβ 3′ truncation entry construct amplified with restriction flanking sites *Bam*HI or *Not*I at the 5′ and *Hind*III at the 3′. All clones were sequence verified.

Construct	Forward Primer	Reverse primers
IKKβ 1-280	pFastBacBamHIfor 5′CCATCGGGCGCGGATCCCGATG GGCCACCATCATCACCACCATGAT TATG 3′	pFastBacIKK2_QLMrev 5′TTCTCGACAAGCTTGTTACATCAGTT GCAGCCACTTCTCCAGTCGCTCAG 3′
IKKβ 1-407	pFastBacNotIfor 5′CTCACTAGTCGCGGCCGCTATG GGCCACCATCATCACCACCATGAT TATG 3′	pFastBacIKK2_RPQPrev 5′TTCTCGACAAGCTTGTTAAGGTTGG GGCCGTGGGGAGATCTGAGTCTC 3′

Cloning of MBP Fusions

MBP fusions were cloned into pMAL-c2g plasmid modified to introduce a *Sfi*I restriction site into the multiple cloning site used for the cloning, 3′ to the MBP ORF. All inserts were prepared by overlap extension PCR with gene specific primers in the table below and a second amplification with forward primer newPT (5′ CCAAC TACGG CCAAC TCGGC CGAGA ATTTG TATTT TCAGG GT 3′) in order to add *Sfi*I restriction site at the 5′. The inserts and vector were then digested with *Sfi*I at 50 °C for 2-3h. Following ligation and transformation reactions, all constructs were sequence verified.

Construct	Forward Primers	Reverse primer
IKKβ 1-280	IKK2_SWSPfor 5′GCCGAGAATTTGTATTTTCAGGGTAGCTG GTCACCTTCCCTGACAACGCA 3′	IKK2_QLMrev 5′ACTTAGTGGCCGAGGCGGCCT TACATCAGTTGCAGCCACTTCTC CAGT 3′
IKKβ 5-280	IKK2_TCGAfor 5′GCCGAGAATTTGTATTTTCAGGGTACATG TGGGGCCTGGGAAATGAAAGA 3′	
IKKβ 8-280	IKK2_PSLfor 5′GCCGAGAATTTGTATTTTCAGGGTCCTTCC CTGACAACGCAGACATGTGG 3′	
IKKβ 11-280	IKK2_TTQfor 5′GCCGAGAATTTGTATTTTCAGGGTACAAC GCAGACATGTGGGGCCTGG 3′	

Limited Proteolysis and Mass Spectrometry

Crystallisation purity IKKβ 262-418 (ULD construct) was diluted to 1 mg/ml and digestion reactions performed at 30 °C for the various dilutions and proteases. For mass spectrometry the protein was digested with trypsin in a mass ratio 1/500 (protease/protein) at 37 °C for 30 min. The data was collected and analysed by Luca Signor (IBS, Grenoble).

SAXS

The protein was prepared in a similar manner to the crystallisation trials protocol as described in the materials and methods section of chapter 3. Three concentrations were measured 1.5, 3 and 6 mg/ml. Data were collected and analysed with Dr. Adam Round (EMBL, Grenoble).

Crystallisation Trials

Protein preparation and crystallisation trials were performed in a similar manner as described in the materials and methods section of chapter 3.

Circular Dichroism

A Chirascan CD spectrometer (AppliedPhotophysics) was used to measure circular dichroism between 180-260 nm with 1.2 s per point. Data were analysed with DataPro software.

"*Being defeated is often only a temporary condition. Giving up is what makes it permanent.*" --- Marilyn vos Savant

Chapter 5: PI3Kβ & mTOR

Introduction

The secondary messenger inositol triphosphate (IP_3) is derived from the phosphatyd-inositol 4,5 biphosphate ($PI(4,5)P_2$) after cleavage by phospholipase C (PLC). A small water-soluble molecule, it mediates the increase of Ca^{2+} concentration in the cytoplasm by binding to IP3-gated Ca^{2+} release channels in the endoplasmic reticulum. Phosphoinositide3-kinases (PI3K) phosphorylate PI at the 3′-OH group of the inositol ring, hence, when they target PI (4, 5) P2, they create docking sites in the plasma membrane for proteins like PLC that relay various intracellular signals (Alberts, 2002).

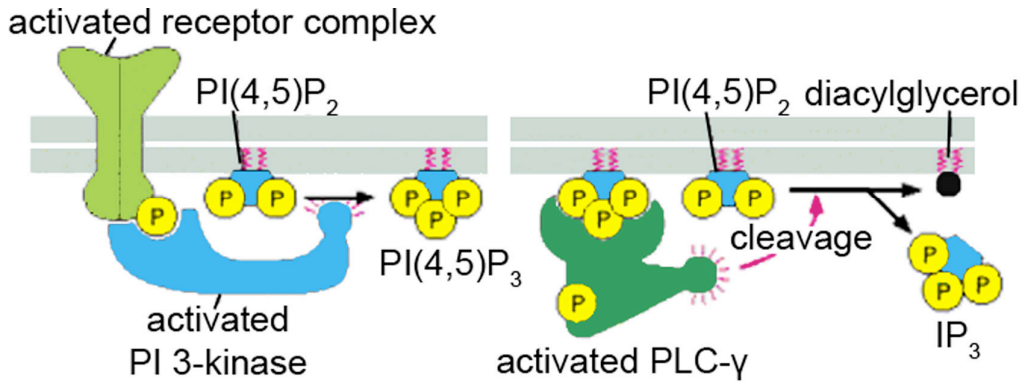


FIGURE 5.1: PHOSPHOINOSITIDE 3-KINASE PATHWAY IS INITIATED UPON ACTIVATION OF A RECEPTOR COMPLEX.

Then, activated PI 3-kinase phosphorylates PI(4,5)P2 and create a docking site (PI(4,5)P3) for proteins like PLC-γ. In the case of PLC-γ binding to the latter, it mediates the cleavage of PI(4,5)P2 and release of the secondary messenger IP3. *adapted from* (Alberts, 2002)

PI3Ks are a family of related lipid kinases divided into three classes: Class I, Class II and Class III. They differ in structure, regulation and substrate specificity. Class IA consists of p110α, p110β and p110δ catalytic subunit and several regulatory subunits p85α, p55α, p50α, p85β and p55γ (Leevers*, et al.*, 1999,Vanhaesebroeck*, et al.*, 2001). PI3Ks are obligate heterodimers: the catalytic subunit coexists with the regulatory one in an inactive conformation and is activated upon binding via a SH2 domain in the regulatory subunit to tyrosine kinase receptors (Geering*, et al.*, 2007,Koyasu, 2003). Alternatively, activated Ras can independently activate the catalytic p110 subunits. Among the catalytic subunits, p110α and p110β are ubiquitous with the latter being prevalent whilst p110δ is specifically expressed in the leukocytes. The most abundant regulatory subunit is p85α. Class IB is only present in mammals and has a single p101 regulatory unit and a p110γ catalytic domain, preferentially expressed in leukocytes (Koyasu, 2003). Atomic resolution structures have been determined for p110α-p85α / PI3Kα (Figure 5.1) (Huang*, et al.*, 2007) and p110γ / PI3Kγ (Walker*, et al.*, 1999).

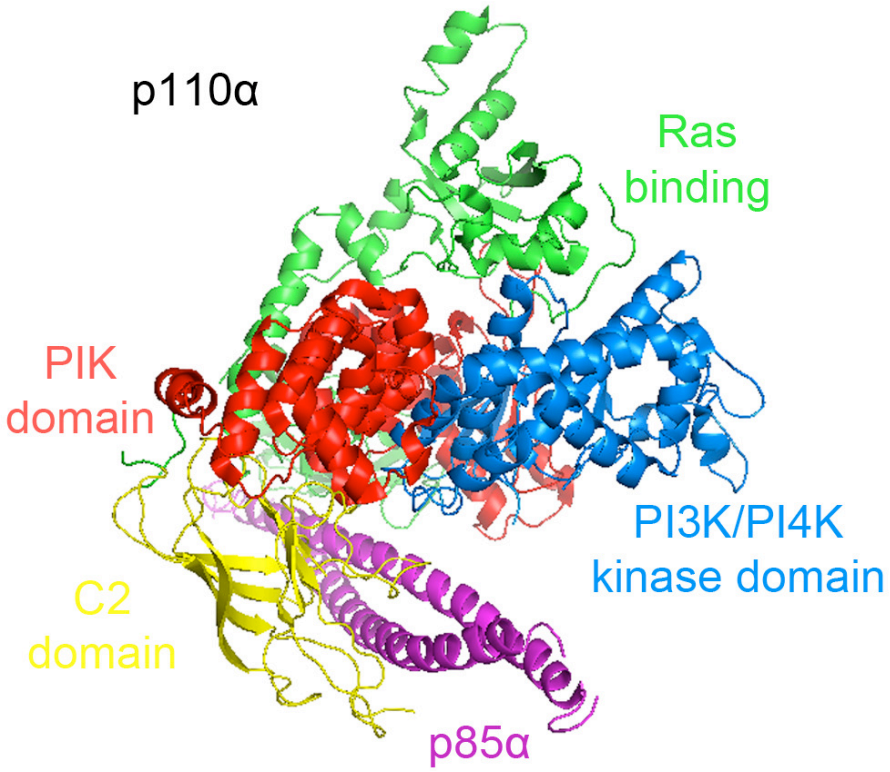


FIGURE 5.2: STRUCTURAL REPRESENTATION OF PI3KA (PDB CODE: 2RD0) SHOWING CATALYTIC DOMAIN (BLUE) AND ADDITIONAL REGULATORY DOMAINS (OTHER COLOURS).

PI3Kβ is composed of the p110β catalytic subunit with additional regulatory subunits. The catalytic subunit is more than 120 kDa and is homologous to the crystallized p110α but to date there is no structure reported. While p110β$^{-/-}$ mice die early in embryonic development (Bi, *et al.*, 2002), conditional knockouts have allowed a more precise investigation of its role *in vivo*. This suggests roles in i) kinase independent signalling leading to insulin insensitivity and impaired glucose homeostasis in the liver, and ii) kinase dependent signalling that hampers tumorigenesis by diminishing Akt phosphorylation in prostate tumors (Jia, *et al.*, 2008). Many cancers show mutation of this gene though data on how it contributes to disease are just emerging (Zhao and Vogt, 2008).

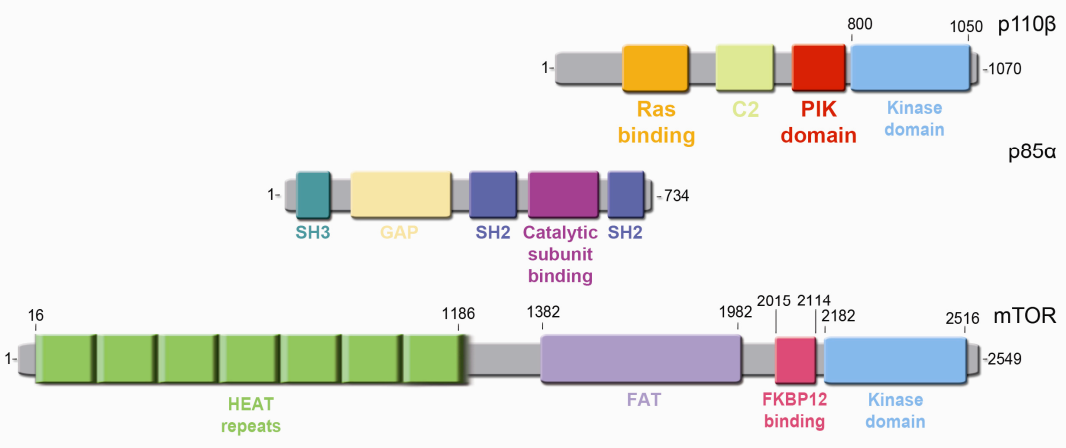


FIGURE 5.2: DOMAIN COMPOSITION OF PI3KB (110B CATALYTIC AND P85A REGULATORY SUBUNIT), AND mTOR

There exists another class of PI3K related proteins, termed Class IV. In contrast to the other classes of PI3K kinases, the members of this class exhibit ser/thr catalytic activity but have PI3K-like catalytic domains. A prominent member of this class is the mammalian target of rapamycin (mTOR).

The story of the discovery of mTOR is fascinating. It starts with the finding of a bacterial species endemic to the Easter Islands, which secreted an antifungal product (Sehgal, *et al.*, 1975,Vezina, *et al.*, 1975). This was the macrolide rapamycin named after the islands of its origin. Soon, it was realised that this molecule possessed immunosuppressive and antiproliferative properties, which initiated research to identify its target. It was soon established that rapamycin bound to FKBP causing inhibition of a large complex exhibiting kinase catalytic activity (Jin, *et al.*, 1992). The catalytic activity in this complex was attributed to the mammalian target of rapamycin (mTOR) also known as rapamycin target protein. This is a striking example of a case where an inhibitor is found before its target protein.

The protein mTOR is a large, 289 kDa ser/thr kinase which exists in two distinct complexes *in vivo*: mTORC1 and mTORC2 (Hara, *et al.*, 2002,Sarbassov, *et al.*, 2004). They regulate a range of processes controlling cell survival, translation, ribosome biogenesis, glucose metabolism and others (Guertin and Sabatini, 2005,Wullschleger, *et al.*, 2006). Mice null for mTOR died as embryos after 5.5-6.5 days (Gangloff, *et al.*, 2004) and a more recent mouse study suggested that specific inhibition of this kinase contributed to their longevity (Harrison, *et al.*, 2009). Structural characterization of the small FKBP12 binding domain of mTOR was successful permitting an understanding of the inhibition mechanism invoked by rapamycin (Choi, *et al.*, 1996). Beyond this, little is known about structure and activity regulation of mTOR.

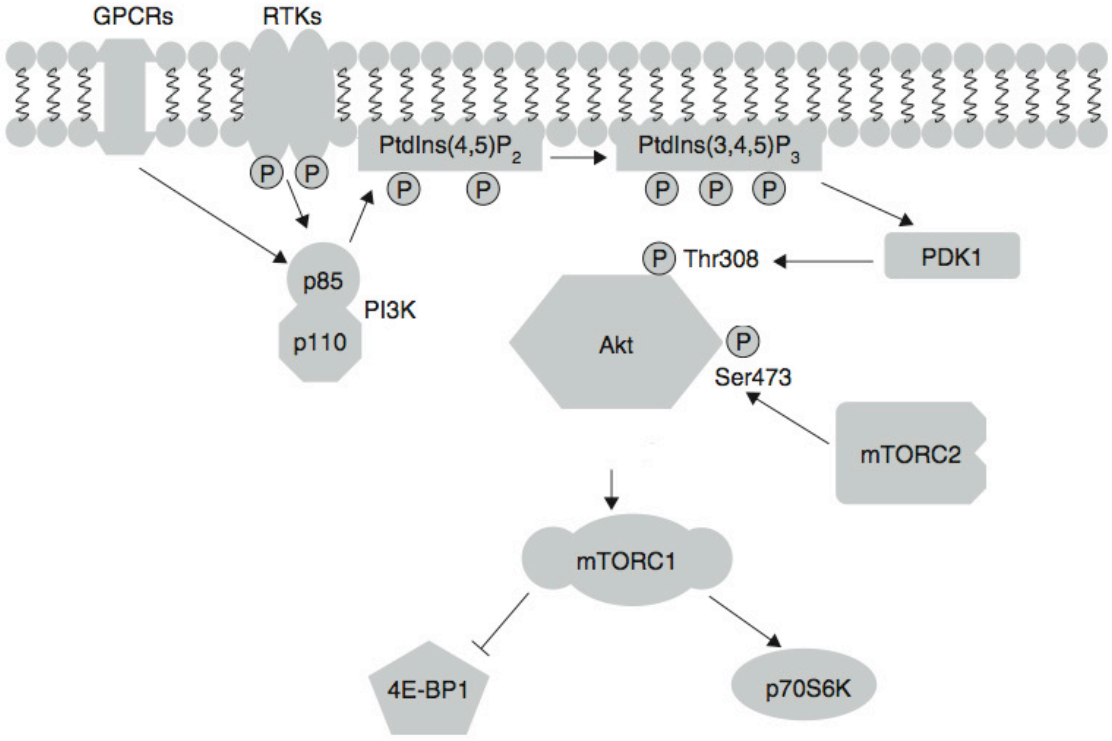


FIGURE 5.3: PI3K/AKT/MTOR PATHWAY (MODIFIED FROM (MARTELLI, ET AL., 2009)

PI3K and mTOR are involved together in cell proliferation and growth pathways (Figure V.III). Upon signal from a tyrosine receptor kinase or other kinases, activation of PI3K leads to PI (4, 5) P2 phosphorylation to PI (3, 4, 5) P3. The latter serves as a docking site recognized by protein kinase PDK1, which then activated, phosphorylates Akt (also known as PKB). Akt was shown to be phosphorylated by mTORC2 and indirectly activated the mTORC1 complex leading to suppression of initiation factor 4E-BP1 and activation of p70S6K (known also as S6K1). This pathway results in phosphorylation of ribosomal subunit 70S and production of more ribosomal proteins. The result of this complex process is increased translation levels and thus enhanced cell growth (Martelli*, et al.*, 2009).

Due to their key roles in this pathway, PI3Kβ and mTOR have been studied intensively for their possible roles in tumorigenesis and attempts made to identify efficient inhibitors. Inhibiting mTOR with rapamycin proved insufficient and was limited only to mTORC1. Notably, neither rapamycin nor analogous molecules were effective at inactivating the catalytic activity of mTOR. It would be expected that blocking the activity of mTOR by inhibition of its catalytic domain or by using a combined therapy of rapamycin and/or analogues along with PI3Kβ inhibition would result in potent way to fight cell proliferation diseases like cancer (Ballou and Lin, 2008). Common inhibitors for PI3K are also known to inactivate mTOR probably due to structural similarity implied by the sequence homology. Wortmannin is a well-characterized inhibitor and is obtained as a natural product from *Penicillium wortmanni*; it inhibits all PI3Ks and has proved useful in the study of different tumor types. The counter side is that it binds covalently to the proteins, making it too toxic for clinical applications. Structural characterization of the catalytic domains of p110β and mTOR has been elusive so far and would be a major breakthrough allowing structure-guided drug design. Moreover, this group of kinases is still poorly characterized with only one structure revealing the coordination of ATP in p110γ (PDB code: 1E8X).

In this work, definition of soluble constructs for the kinase catalytic domains of mTOR and p110β was desired and we therefore applied the ESPRIT random construct screening approach.

Results

Initial considerations

The kinase catalytic domains of p110β and mTOR share only about 20% identity with no significant homology to the targets studied in earlier chapters. In contrast, when the catalytic domain of p110β is compared with p110α and p110γ, a high resemblance is found with 56% and 51% identity respectively. Therefore, a sequence homology model on p110β is expected to be quite accurate.

I-TASSER analysis on the catalytic domains of p110β and mTOR showed that they most closely resembled p110γ (PDB code: 1E8XA). As expected, the homology model for p110β is more accurate with a RMSD from its template of 0.69 Å. A reliable model was not obtained for mTOR. Both proteins have C-terminal kinase catalytic domains hence a 5′ truncation library strategy was selected for ESPRIT. Since they are both very large proteins and only the kinase catalytic domains were being targeted, a comprehensive screen was performed on the regions containing the catalytic domains plus additional upstream sequence, and not the full-length forms. To define how much extra sequence to add to the core domain prior to beginning truncations, a study was made of resolved structures, homology modeling results and domain predictions. The two most homologous structures, p110γ (PDB code: 2E8XA) and p110α (2RD0, Fig. 5.1), revealed large protein molecules containing interacting domains. Therefore, the starting construct for library analysis of p110β was designed to span multiple domains prior to the kinase domain and was similar to the crystallized form of

p110γ. The domain prediction boundaries were also taken into consideration suggesting two potential C-termini: native at amino acid position 1070 and a second at - 1050 corresponding to predicted kinase catalytic domain boundaries by Swiss-Prot. In the case of mTOR, the FKBP12 binding domain was used to guide the N-terminal upstream sequence prior to commencing truncation. Like p110β the native (aa 2549) and slightly shorter (aa 2516) C-termini were investigated. The strategies are summarized in Fig. XXX and each library was made using an equimolar mix of the two starting plasmids encoding the slightly different C-termini.

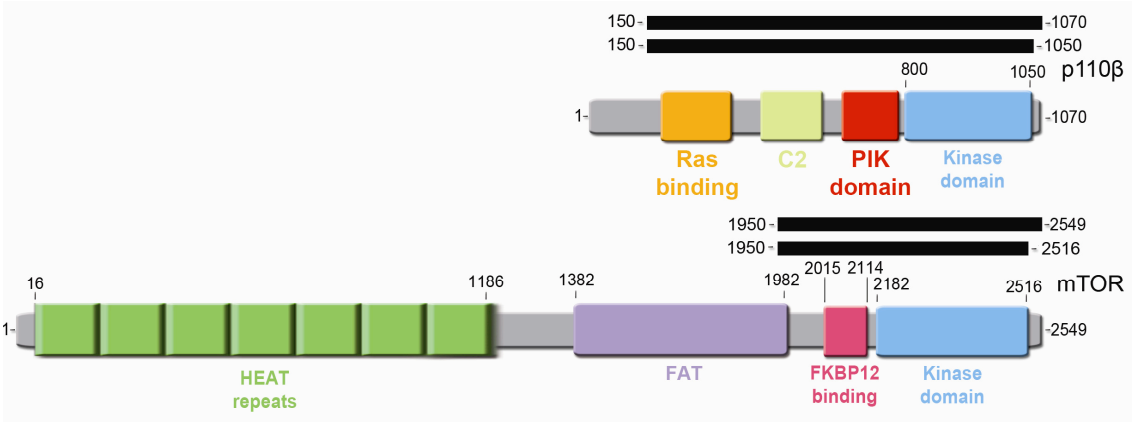


FIGURE 5.4: TRUNCATION LIBRARY ENTRY CONSTRUCTS MAPPED ONTO 110B AND mTOR

Generation of Libraries

Starting constructs of p110β were synthesized as codon-optimized genes (Geneart) and cloned in the minimal library vector pYUM6002. Equal concentrations of the plasmid pair were mixed, the exonuclease truncation reaction was performed and the library size-fractionated into two sublibraries corresponding to inserts encoding 50-400 amino acids and 400 to full-length (Figure V.IVa). The mTOR library was made in a similar manner, again from a codon-optimised gene, and fractionated into two sublibraries containing inserts

encoding 50-250 amino acids and 250 to full length (Figure V.1Vb). *E. coli* BL21 was transformed with the mTOR and p110β plasmid sublibraries and 4608 individual clones isolated by automated colony picking into 384 well plates. The oversampling for the library was calculated at least 3-fold. The library statistics are presented in Table V.I.

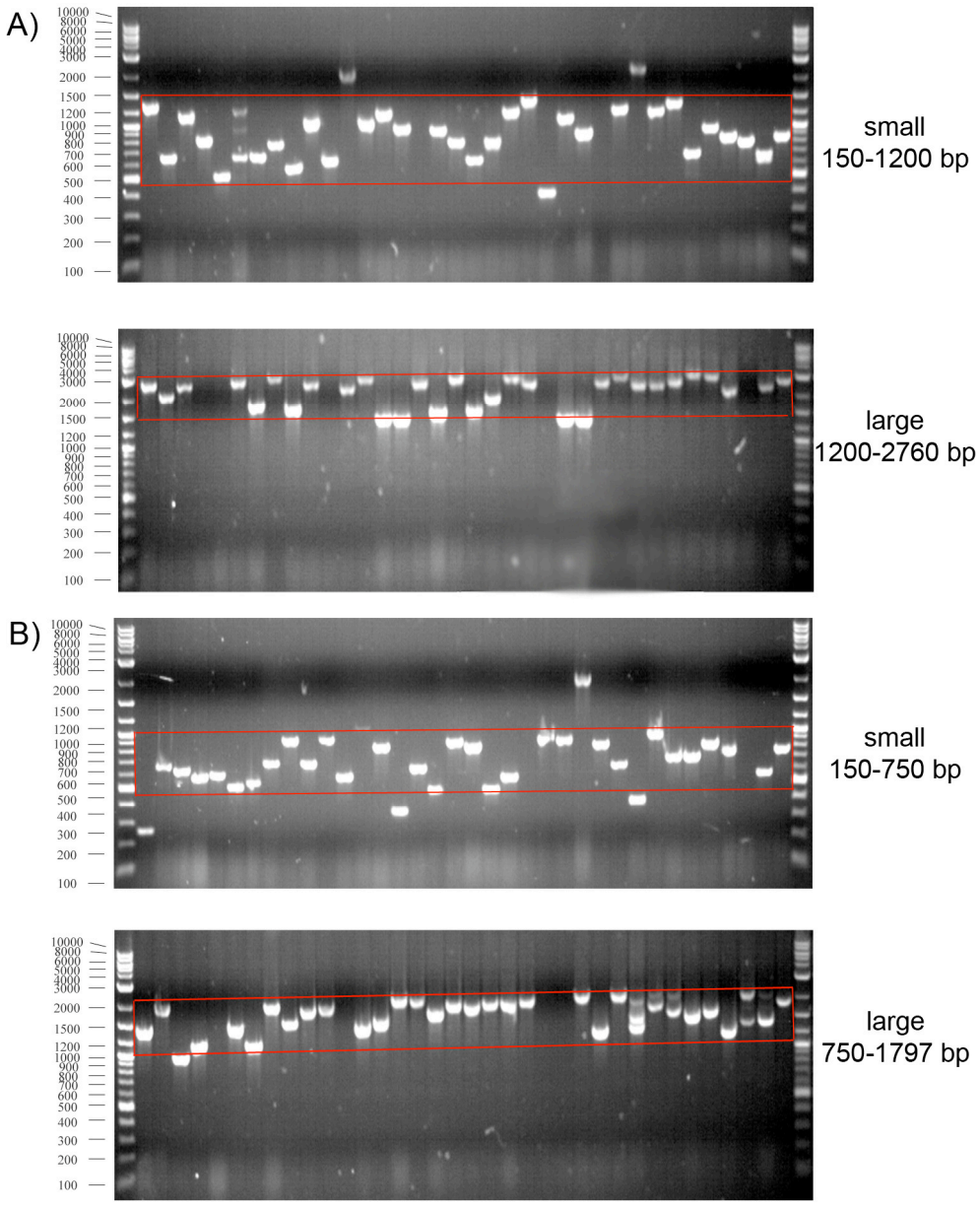


FIGURE 5.5: COLONY PCR SCREEN FOR THE EVALUATION OF TRUNCATION LIBRARY QUALITY.

36 of 96 tested per sublibrary are shown. (A) p110β 5′ truncation library quality and (B) mTOR 5′ truncation library quality. Insert size is equal to PCR product minus 320bp added with the primers for PCR screen. The PCR screen revealed a high insert quality with little contamination by DNA fragments outside of the desired size ranges.

E. coli BL21 was transformed with the plasmid library and 4608 individual clones isolated by automated colony picking into 384 well plates. The oversampling for the library was calculated at about 3-fold and higher. The library statistics are presented in Table V.I. The PCR screen demonstrated high quality of the libraries generated. Little contamination is seen and the size separation of the truncation libraries has allowed comprehensive coverage.

	p110β		**mTOR**	
Construct (aa)	**150-1070**		**1-599**	
Truncation Type	5′	5′	5′	5′
Size Range (bp)	150-1200	1200-2760	150-750	750-1797
Diversity	1050	1560	600	1047
Oversampling	4.4	2.95	7.7	4.4
Quality (%)	86.1	80.6	83.3	86.1
Contamination (%)	8.3	5.6	8.3	2.8
Completeness (%)	97.7	90.8	99.8	97.7

TABLE 5.1: SUMMARY OF THE GENERATION AND SCREENING OF P110B AND MTOR 5′ TRUNCATION LIBRARIES.

*Completeness is calculated using server: http://guinevere.otago.ac.nz/aef/STATS/index.html.

The isolated clones were arrayed onto nitrocellulose membranes and colony blots probed for N-ter 6xHis and C-ter biotinylation levels by hybridisation of fluor-conjugated anti-6xHis mAb and streptavidin. According to the usual protocol, constructs with no 6xHis tag signal were eliminated from further analysis removing a very large portion of the library, and the

remaining clones ranked according to their biotinylation signals. The first observation was that the signals from the colony array blot were significantly less intense than previous libraries, especially for the large-sized sublibraries. After image analysis and signal quantification with Visual Grid software (GPC-Biotech), it was clear that most of the signals were close to the background levels from the endogenous *E. coli* BCCP, but that a small number of clones exhibited significant signals. Therefore, a smaller selection of clones was isolated from the library plates for the next step of purification screening from 4 ml expression cultures (26 for each of p110β and mTOR small fragments and 6 from the large fragments).

Ni^{2+}NTA purification was carried out using a liquid handling robot and purified proteins analyzed by 15% SDS-PAGE. No visible purified protein was observed on Coomassie blue stained gels, but three constructs from the highest ranked small fragments of the p110β library were visible by western blot with fluorescent conjugated streptavidin. They appeared approximately the same size and DNA sequencing revealed a cluster similar to the domain prediction for the C-terminal kinase catalytic domain (Fig. V.VA). Moreover, the native C-terminus was selected for all of them despite the presence of the shorter second construct in the library pool. The three constructs were scaled up to 1 L expression cultures and purified with Ni^{2+}NTA agarose resin. Again, protein was not visible by Coomassie blue stained SDS-PAGE gel but some was detected by western blot, especially for the longer construct aa 806-1070 (Fig. V.VB).

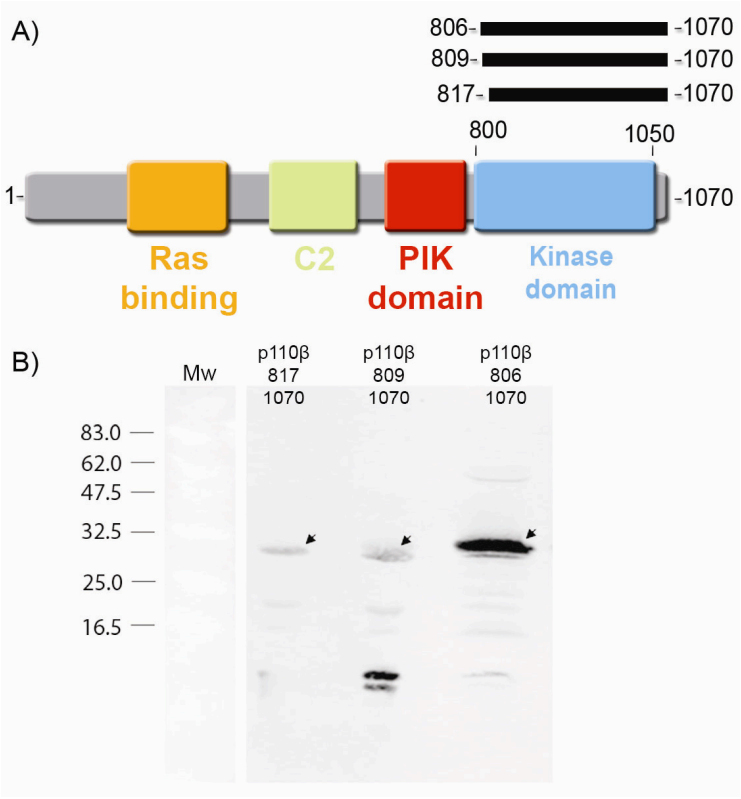


FIGURE 5.6: EXPRESSION SCREEN RESULTS ON P110B.

(A) constructs identified are mapped onto the protein domain predictions; (B) western blot of purified protein fractions from 1L TB expression cultures reveals small amounts of purified protein, especially for the longest construct;

Insect cell expression of the library-derived p110β catalytic domain

For structural characterisation, both PI3Kα and PI3Kγ were over-expressed in insect cells (Huang, *et al.*, 2007,Walker, *et al.*, 1999). Therefore, this was tried for the two longest p110β constructs identified from the screen using the latest baculovirus technology (EMBL Eukaryotic Expression Facility, Grenoble) (Fitzgerald, *et al.*, 2006). Both constructs were subcloned minus biotin tag into bacmids containing a separate YFP-expression cassette (for monitoring late gene expression), and recombinant baculovirus DNA isolated subsequently from DH10EMBacY cells. Sf21 cells were transfected with the viral DNA and V_0 virus

harvested after approximately 60 h. It was further amplified in 50 ml of medium to generate V_1 virus. Subsequently, after the viral stock was collected, cells were resuspended in 50 ml of medium and the expression test performed (Fig. V.VI). High levels of protein expression were observed of which the majority was found in the pellet fraction.

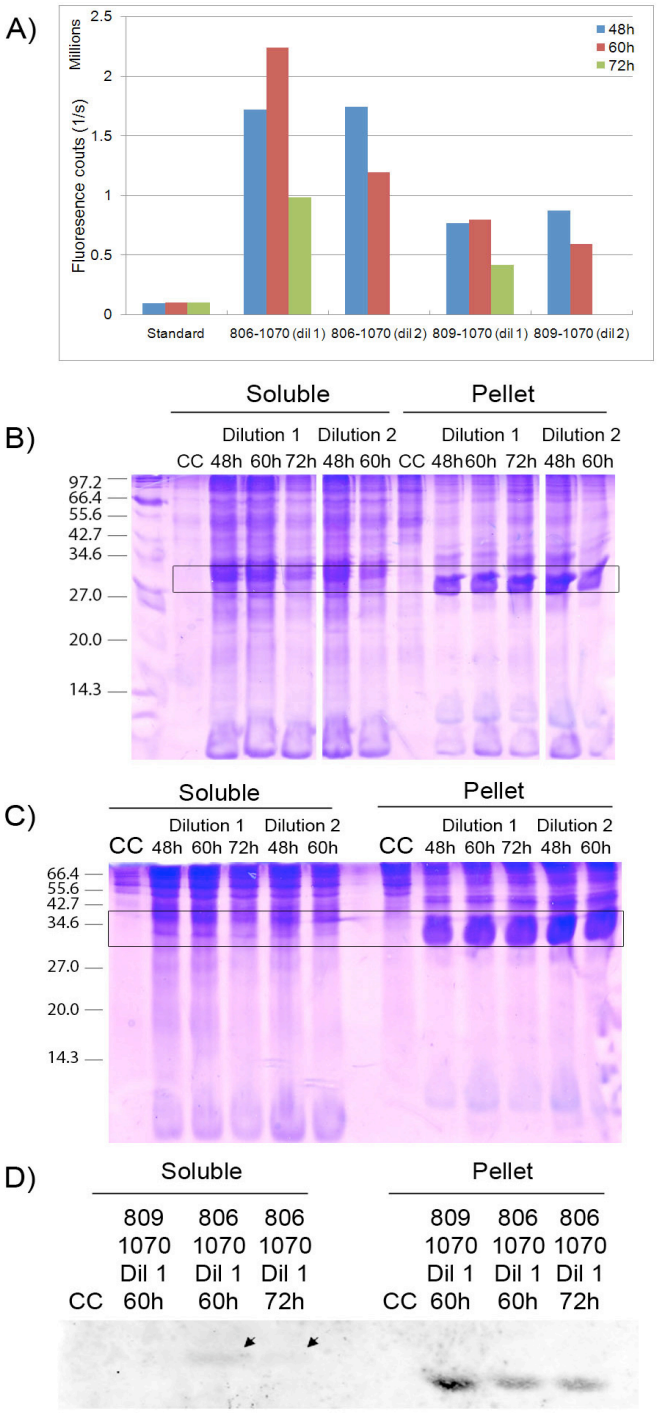


FIGURE 5.7: BACULOVIRUS V1 GENERATION AND EXPRESSION RESULTS ON P110B.
(A) fluorescence measurements following the generation of V1 virus post cell arrest; (B) SDS-PAGE of the expression of p110β 806-1070 and (C) p110β 809-1070 show the majority of recombinant protein is in the pellet (D) western blot confirmed the presence of the protein in the pellet but also reveals small amounts in the soluble fractions;

Due to the observation of large fraction of insoluble protein, an alternative purification protocol was used, with buffers containing a low concentration of the non-ionic detergent NP-40 and KCl. The expression was tested in 40 ml of Hi5 cell culture with two different concentrations of V_1. The result was clear expression of the protein as before, but the slightly longer construct (806-1070) purifying well with no obvious difference between the different virus titer used (Fig. V.VIIA). A consecutive scale up on p110β 806-1070 to 1.5 L with 40 µl/ml V_1 infection of Hi5 cells yielded approximately 1 mg purified protein (Fig. V.VIIB). There was still protein remaining in the pellet fraction that migrated at a different, lower size on SDS-PAGE to the purified protein suggesting that the purified soluble material was a different conformational population to the insoluble material, perhaps being post-translationally modified. Indeed, O-glycosylation is predicted (OGPET, University of Texas at El Paso) to occur at Thr1009.

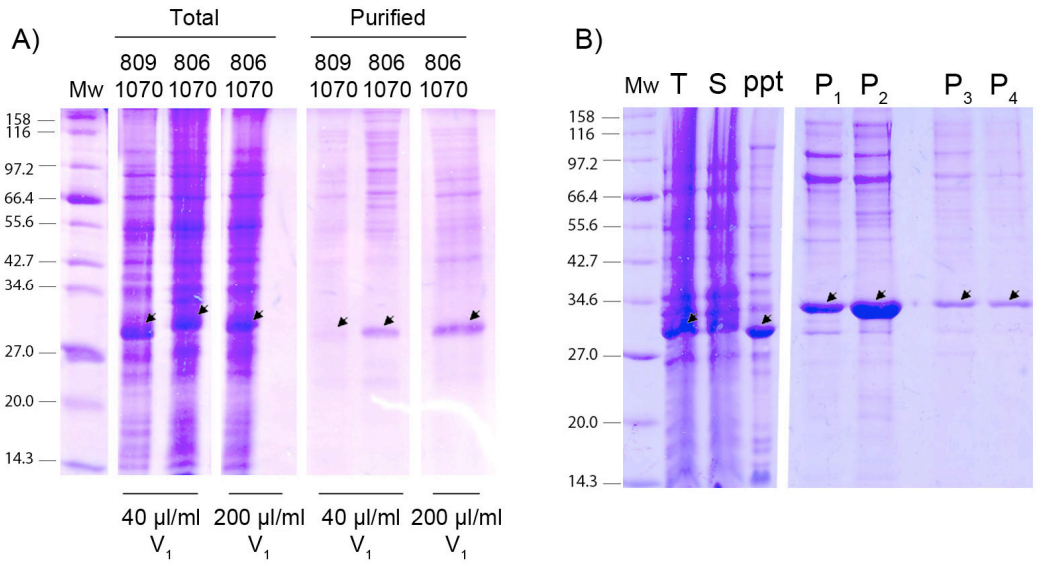


FIGURE 5.8: ALTERNATIVE PURIFICATION PROTOCOL USED ON P110β CATALYTIC DOMAIN CONSTRUCTS.

(A) SDS-PAGE analysis of expression trials with two different viral titers and incorporating a low NP40 concentration and KCl in the purification protocol. Protein was purified from the longer construct; (B) SDS-PAGE of the 1.5 L scale-up of expression of p110β 806-1070 demonstrates that whilst some protein remains in the pellet a significant fraction of soluble material can be purified (T – total, S – Soluble, ppt – pellet, P1, P2, P3, P4 – different purification fractions eluted with 300 mM Imidazole);

Although promising, when the purified p110β catalytic domain (806-1070) was analysed by size exclusion chromatography, the majority eluted in its void volume. Attempts to reduce the polydispersity by varying salt concentrations or addition of CHAPS resulted in no further improvement. Protein that eluted in the correct elution volume was concentrated to 1.2 mg/ml and submitted for crystallization trials with no crystals observed to date.

Discussion & Conclusions

The structural characterization of phosphoinositide 3-kinases is much less advanced than classical protein kinases. It is of great importance to obtain structural information on their catalytic domains since inhibitors appear promising as antitumour drugs. Previous reports have demonstrated that some members of this family can be over-expressed in insect cells and successfully crystallized (Huang*, et al.*, 2007,Walker*, et al.*, 1999), although this has not been achieved for mTOR or p110β (catalytic subunit of PI3Kβ) despite intensive studies. The approach undertaken in this work was soluble construct definition of the kinase catalytic domains for mTOR and p110β using the ESPRIT technology. The screening strategy for both, defined after initial modeling and other considerations, was i) unidirectional truncations since both catalytic domains were C-terminally located, and ii) libraries constructed on sub-full-length starting fragments in order to focus on the catalytic domains only.

The screen of mTOR yielded no positives and was initially disappointing. However, it is often advantageous to "fail fast, fail cheaply" and one positive aspect to this result was that in a single, short experiment, all N-ter junctions were tested against two C-terminal variations. Therefore, a definitive conclusion can probably be drawn about the potential of achieving mTOR catalytic domain expression in *E. coli* without embarking of long iterative cycles of construct synthesis and testing.

In contrast, more promising results were achieved with p110β. Three constructs were identified that exhibited low, but measurable levels of soluble expression in *E. coli*. When two of these constructs were transferred to the baculovirus expression system, much higher levels of expression was found. With further optimization, one of the constructs yielded low but useable quantities of purified protein that enabled the setting up of crystallisation trials. No crystals have been identified to date, but the observation of possible post-translational modification(s) suggests firstly that it may well be folded and secondly that identification and subsequent suppression of the modification could favorably influence crystallisability.

More broadly, this work provides a demonstration that even when very low levels of soluble expression are observed in *E. coli* during library screening, the construct boundaries identified may be structurally relevant and that expression of these fragments in a higher eukaryotic system (e.g. insect cells) may result in more efficient folding and higher yields. The difference in behavior of kinase constructs in *E. coli* and insect cells is well appreciated and suggests potential advantages for construct screening approaches functioning directly in insect cells.

Materials & Methods

Open Reading Frames

Open reading frames corresponding to the following proteins were used: synthetic PK3CB_human/p110beta (Swiss-Prot Accession Number: P42338) 150-1070 aa (GeneArt) and synthetic FRAP1_human (Swiss-Prot Accession Number: P42345) 1950-2549 aa. Genes were provided by AZ and were codon optimised by Geneart.

FRAP1 Library Entry Clones Cloning

The synthetic FRAP1_human gene constructs encoding for 1950-2549 aa and 1950-2516 aa were cloned into pYUM6002 (*see* Chapter 2). Both were amplified with flanking *Asc*I and *Nsi*I restriction sites introduced with the following primers: RAPT1for (5′ GGGGT AATGA GGCGC GCCGC GTCCG CTGGT TGGCC G 3′), RAPT1rev (5′ CCATT GTTCG ATGCA TTCCA AAACG GGCAC CAGCC 3′) and RAPT1rev2 (5′ CACGA CGCAT CGATG CATTA TCATG GCTAA AATCA CGGCC G 3′). The cloning procedure was as described for DAPK1 (*see* Materials and Methods section, Chapter 3).

PI3Kβ p110β Library Entry Clone Cloning

The synthetic p110beta_human gene constructs encoding for 150-1070 aa and 150-1050 aa were cloned in pYUM6002 (*see* Chapter 2). Both were amplified with flanking *Asc*I and *Nsi*I restriction sites introduced with the following primers: PI3Kbfor (5′ GGGGG TAATG AGGCG CGCCG CGCAA AATGC GTAAA TTCAG CG 3′), PI3Kbrev (5′ CCCAT TGTTC GATGC ATTGC TACGA TAATC TTTAC GCACG GTATG 3′) and PI3Kbrev2 (5′ ATCCC GTGCA CTCGA TGCAT TACGC AGCGC TTCAT CGAAT TTC 3′). The cloning procedure was as described for DAPK1 (*see* Materials and Methods section, Chapter 3).

Library Generation and Testing

The library generation was performed largely as described in chapter 2 (*see* ESPRIT Protocol, Materials, Methods and Protocols section) with the following modifications: Both of the plasmid sets were truncated in a similar manner to DAPK1 and TBK1 (mix of the two constructs per protein in a single exonuclease mediated truncation reaction). Both were linearised with *Asc*I and *Aat*II and truncation were limited to 30 bp/min (mTOR) and 46 bp/min (p110β) by 35.63 mM NaCl and 4.24 mM NaCl respectively. The library was tested in IPTG inducible BL21 (DE3) cells.

Insect Cell Expression

Library clones were subcloned in pFastBac1 (Invitrogen) with restriction cloning using *Bam*HI and *Hin*dIII. Flanking restriction sites were added with the following generic pFastBacfor (5′ CCATC GGGCG CGGAT CCCGA TGGGC CACCA TCATC ACCAC CATGA TTATG 3′) and pFastBacPI3Kb_DYRSrev (5′ CTCGA CAAGC TTGTT AGCTA CGATA ATCTT TACGC ACGGT ATGCG CCATC 3′) and bacmids were obtained by transformation of DH10EMBacY (provided by Imre Berger, EMBL, Grenoble) with the the sequence-verified pFastBac1 clones, followed by DNA purification and virus generation (Fitzgerald*, et al.*, 2006). The best yielding viruses were scaled up and tested for protein expression in a Hi5 cell line (Invitrogen). Largely insoluble protein expression was improved using a modified

purification protocol: sonicating the pellets in lysis buffer (50 mM Tris, 250 mM KCl, 0.1% NP40 pH 7) followed by centrifugation at 15,500 g for 30 min. Soluble fractions were immobilised on Ni^{2+}NTA resin (Qiagen) overnight at 4 °C shaking at 70 rpm on an orbital shaker, then transferred to a gravity column and washed with lysis buffer, high salt buffer (50 mM Tris, 1 M KCl, 5 mM imidazole and 10 mM β-ME pH 7) and finally washing buffer (50 mM Tris, 100 mM KCl, 10-20 mM imidazole and 10 mM β-ME pH 7). Protein was eluted in buffer with three steps of increasing imidazole concentration (50 mM Tris, 1M NaCl, 100 mM /200 mM /400 mM imidazole and 10 mM β-ME).

"*The aim of argument, or of discussion, should not be victory,*

but progress." --- Joseph Joubert

Chapter 6: Discussion

Discussion

Library-based construct screening is a recent methodological development in the field of structural biology that uses the principles of random mutation and screening/selection, similar to the protein engineering approach of directed evolution. As a strategy, it aims to replace iterative cycles of time consuming cloning and expression testing, the success of which is largely dependent on the quality of starting information for construct design, with a single linear process where all possible constructs of a given protein are assayed. In the laboratory, it has been applied by coworkers both to novel targets where no starting information is available, and to improving expressing, but misbehaving constructs. Here we have applied it to targets with good levels of annotation, but which have resisted overexpression previously. Due to the large number (potentially saturation-level) of constructs tested, the expected output is an answer to the question, "Will my protein express solubly in *E. coli*?" If the answer is yes, purified material is generated that can be used for further studies such as crystallisation or functional assays. A negative result, although disappointing, has a high value in that it may prevent many months of fruitless work and guide the researcher towards other expression systems, or stopping work on the target.

In order to have confidence in the answer, the requirements for such a screen are high levels of domain boundary coverage and a reliable detection of the soluble phenotype. The exonuclease protocols used in this work (Ostermeier and Lutz, 2003) seem unbiased, evidenced from analysis of both construct size diversity and sequencing data. The orientation of the insert is maintained throughout, a particular advantage for the bidirectional truncation procedure. Related to this last point, since the gene is truncated within the expression vector, it is simple to obtain tens of thousands of colonies from the one-part ligation which is much more efficient than a classical two part ligation as used for cloning fragmented DNA.

High throughput solubility screens with the capacity to assay such large numbers of constructs as to get a statistically reliable answer inevitably suffer from false positive or false negative results, ideally at a low frequency. Once characterised, steps should be taken to eliminate these. With kinases, a false negative could occur due to a negative selection against active clones from activity-mediated toxicity; catalytically inactive mutants should reduce this. A false positive signal may occur from some inclusion bodies that show some biotinylation, either due to binding of misfolded proteins to chaperones prior to inclusion body formation, or the slow aggregation of folded, transiently soluble molecules. In the ESPRIT workflow, this has been addressed by the combination of very high throughput, but indirect solubility screen employing *in vivo* modification of a peptide tag, with subsequent direct assessment of the enriched, putatatively soluble constructs in a 96 well Ni^{2+}NTA purification screen. In this way, we have seen that many constructs are neither fully soluble or insoluble, but exhibit a mixed phenotype with a minor purifiable fraction that, in several cases in the group, has enabled a structure determination.

The study on DAPK1, the best characterised of the kinase set, demonstrated that the construct screening approach can precisely identify a folded catalytic domain from a multidomain

human kinase. Moreover, the identification of several soluble constructs led to crystals with different crystal packing revealing small differences in conformation to the existing (overtruncated) structures (Tereshko*, et al.*, 2001). When studying targets such as these where small molecule inhibitors are desired, alternative crystal forms can be very valuable because they may allow soaking of compounds into pregrown crystals, thereby speeding up drug discovery. In the case of DAPK1, we have identified longer constructs than those reported previously; these reveal structured regions involved in Ca^{2+}/calmodulin binding and regulation of this interaction by post-translational modification. Furthermore, the minimal binding site region was defined by rescreening the truncation library in the presence of coexpressed calmodulin, and the results differ slightly from previous studies on a Ca2+/calmodulin/synthetic peptide complex (PDB code: 1YR5). These soluble constructs of DAPK1 would be useful to study further the contribution of phosphorylation at Ser289 and Ser308 to its activity. Mass spectrometry analysis might reveal autophosphorylation of Ser308 in *E. coli*-produced material. Ser289 has been shown to be phosphorylated by RSK (Anjum*, et al.*, 2005); thus coexpression of this kinase would be required, or perhaps a S289E mutant could be analysed to see if it affects the conformation of the C-ter helix.

When ESPRIT was applied to the IKK-like kinases TBK1 and IKKβ it was found that the N-ter catalytic domains resisted over-expression in *E. coli* despite the 3-fold oversample of all C-ter boundaries. Low expression of the IKKβ catalytic domain was observed, but attempts to increase the yield failed, despite initially promising results with refolding protocols. The results further suggest, unexpectedly, that the N-terminal lobe of the kinase catalytic domain could be expressed individually resulting in partially soluble protein. A bidirectional truncation screen on IKKβ identified overexpressed, purifiable regulatory domains covering the ubiquitin-like domain (ULD) and NEMO-binding regions described previously in cell biology studies (May*, et al.*, 2000). The ULD was pursued towards structural characterization

and simple functional experiments suggesting *in vivo* interactions. In current models of NF-κB activation, the protein kinases are activated upon binding to anchored, or unanchored, polyubiquitin chains (Xu*, et al.*, 2009). It seems plausible that the ULDs of the IKK kinases may interact with this ubiquitin. To test this hypothesis, the interaction of the purified ULDs with ubiquitin could be investigated. Attempts to crystallize the longer ULD constructs identified from the screen were unsuccessful but biophysical characterisation including SAXS measurements at the ESRF beamline demonstrated that it is monodisperse and monomeric. Limited proteolysis resulted in a better-behaved protein that still did not crystallise.

The work on the p110β subunit of PI3Kβ has demonstrated how construct screening is capable of identifying domain boundaries in *E. coli* even when the amounts of expressed soluble proteins are very low. Subsequent expression of the identified construct in insect cells resulted in improved yields, therefore suggesting that future construct screening approaches directly in insect cells might be advantageous for some targets.

Finally, while construct screening approaches in *E. coli* have proven useful for the identification of soluble domains in proteins with no homologues, they are clearly more limited for finding such solutions in proteins as well characterised as the kinase set studied here. It would seem that expression is dependent on more than just optimised construct boundaries. For example, IKKβ, formed irreversible complexes with coexpressed chaperones, and otherwise formed inclusion bodies. Refolding yielded highly soluble, but aggregated material. All suggest problems at the level of folding, either because of the limitations of *E. coli*, or because the folding pathway requires full-length material. Indeed, a current model of IKKβ activation *in vivo* suggests the catalytic domain is in contact with regions that are distant in primary sequence (Karin, 1999) (Hayden and Ghosh, 2008).

Future perspectives

Construct libraries explore structural boundaries and, where this is the solution to the problem, purified material is obtained rapidly. However, as discussed above, some of the catalytic domain constructs identified appear to have good boundaries identified by library screening, but misfold, as evidenced by expressing the PI3Kβ catalytic domain construct in insect cells, and by the small amount of material produced for the IKKβ catalytic domain. Therefore, improving the folding pathway and reducing post-folding aggregation due to unfavorable amino acids on the surface of the protein might be achieved by random mutagenesis on the identified constructs in exactly the way reported previously for GFP (Pedelacq*, et al.*, 2006). With such an approach, it would be important to proceed with caution, since a risk exists that beneficial mutations may change the protein in an unwanted way.

The observation of partially soluble catalytic domain lobes could also be explored further. It is perhaps interesting from an evolutionary perspective that the lobes appear autonomously folded. To our knowledge, this has not been reported previously. From a practical perspective, if the separate lobes were to yield enough material, they could be used to reconstitute the kinase catalytic domain by *in vitro* intein-mediated ligation. This might be a new way to obtain catalytic domain expression for some kinases and would also permit differential isotopic labelling of lobes for NMR studies on protein dynamics.

From a technical perspective, increasing the screening capacity of ESPRIT would be advantageous for assaying larger libraries, or obtaining better coverage of the bidirectional truncation constructs that are undersampled; currently 5-10% coverage of total theoretical diversity is tested. A life-or-death selection potentially offers this, however it is often difficult to rank positives relative to each other, or set a cut-off, in contrast to using a screen. The

automation employed in the current form of ESPRIT provides a major advantage in clone-handling and downstream manipulation of hits because all the constructs are ordered in labelled plates. However it comes with the limitation of about 30,000 constructs per screen. Already started in the laboratory is work towards incorporation of an open reading frame (ORF) selection system (Gerth*, et al.*, 2004) meaning the clones picked and tested are all in frame, resulting in a potential 9-fold increase in efficiency for bidirectional libaries.

Focussing on a subset of the library using additional parameters beyond solubility could provide similar advantages. For example, hybridising the truncation library colony macroarrays with fluorescent oligonucleotides that bind to plasmid DNA encoding a region of interest could be combined with the solubility readout to identify clones expressing a precise domain. Furthermore, one could probe the membranes for example with an antibody against kinase (auto) phosphorylation sites to identify active enzymes.

The coexpression method demonstrated here for the DAPK1-Ca^{2+}/calmodulin interaction could be of interest where co-folding occurs between the library protein of interest and a partner supplied as a fixed bait. Already experiments in the lab have shown that this can yield expression in otherwise difficult cases. Moreover, an alternative promoter regulating the expression of the partner protein could be useful since on a single array verification of the influence on solubility upon switching on and off the expression of its partner could be observed, providing a high throughput assay on folding-upon-binding effects.

"*We can't solve problems by using the same kind of thinking*

as we used when we created them." --- Albert Einstein

References

1. Adhikari, A., Xu, M. & Chen, Z.J. Ubiquitin-mediated activation of TAK1 and IKK. *Oncogene* **26**, 3214-26 (2007).
2. Aharoni, A. *et al.* Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. *Proc Natl Acad Sci U S A* **101**, 482-7 (2004).
3. Ahmad, R., Raina, D., Meyer, C., Kharbanda, S. & Kufe, D. Triterpenoid CDDO-Me blocks the NF-kappaB pathway by direct inhibition of IKKbeta on Cys-179. *J Biol Chem* **281**, 35764-9 (2006).
4. Alberts, B., Johnson A., Lewis, J., Raff, M., Roberts, K. and Walter, P. *Molecular Biology of the Cell*, (Garland Science, New York, 2002).
5. Anderson, S. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucl Acids Res* **9**, 3015-3027 (1981).
6. Angelini, A. *et al.* Expression of Helicobacter pylori CagA domains by library-based construct screening. *FEBS J* **276**, 816-24 (2009).
7. Anjum, R., Roux, P.P., Ballif, B.A., Gygi, S.P. & Blenis, J. The tumor suppressor DAP kinase is a target of RSK-mediated survival signaling. *Curr Biol* **15**, 1762-7 (2005).
8. Ballou, L.M. & Lin, R.Z. Rapamycin and mTOR kinase inhibitors. *J Chem Biol* **1**, 27-36 (2008).
9. Baneyx, F. *Protein Expression Technologies*, (Horizon Bioscience, 2004).
10. Banner, D.W., Kokkinidis, M. & Tsernoglou, D. Structure of the ColE1 rop protein at 1.7 A resolution. *J Mol Biol* **196**, 657-75 (1987).
11. Beckett, D., Kovaleva, E. & Schatz, P.J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**, 921-9 (1999).
12. Berridge, M.J. *Cell Signalling Biology*, (Portland Press Limited, 2008).
13. Bhagwat, S.S. Kinase inhibitors for the treatment of inflammatory and autoimmune disorders. *Purinergic Signal* **5**, 107-15 (2009).
14. Bi, L., Okabe, I., Bernard, D.J. & Nussbaum, R.L. Early embryonic lethality in mice deficient in the p110beta catalytic subunit of PI 3-kinase. *Mamm Genome* **13**, 169-72 (2002).
15. Bialik, S. & Kimchi, A. The death-associated protein kinases: structure, function, and beyond. *Annu Rev Biochem* **75**, 189-210 (2006).
16. Bieniossek, C. *et al.* Automated unrestricted multigene recombineering for multiprotein complex production. *Nat Methods* **6**, 447-50 (2009).
17. Blundell, T.L., Jhoti, H. & Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* **1**, 45-54 (2002).

18. Bolivar, F., Rodriguez, R.L., Betlach, M.C. & Boyer, H.W. Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. *Gene* **2**, 75-93 (1977).
19. Broemer, M., Krappmann, D. & Scheidereit, C. Requirement of Hsp90 activity for IkappaB kinase (IKK) biosynthesis and for constitutive and inducible IKK and NF-kappaB activation. *Oncogene* **23**, 5378-86 (2004).
20. Brown, E.J. *et al.* A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature* **369**, 756-8 (1994).
21. Brown, N.R. *et al.* Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J Biol Chem* **274**, 8746-56 (1999).
22. Brown, T.A. *Gene Cloning & DNA Analysis: An Introduction*, (Blackwell Publishing, 2006).
23. Burley, S.K. An overview of structural genomics. *Nat Struct Biol* **Suppl:**, 932-934 (2000).
24. Byrd, D.R. & Matson, S.W. Nicking by transesterification: the reaction catalysed by a relaxase. *Mol Microbiol* **25**, 1011-22 (1997).
25. Cabantous, S. *et al.* Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis. *J Struct Funct Genomics* **6**, 113-9 (2005).
26. Chang, A.C. & Cohen, S.N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**, 1141-56 (1978).
27. Chapman-Smith, A. & Cronan, J.E., Jr. In vivo enzymatic protein biotinylation. *Biomol Eng* **16**, 119-25 (1999).
28. Chen, F., Castranova, V., Shi, X. & Demers, L.M. New insights into the role of nuclear factor-kappaB, a ubiquitous transcription factor in the initiation of diseases. *Clin Chem* **45**, 7-17 (1999).
29. Chen, M.J. *et al.* The functional human dihydrofolate reductase gene. *J Biol Chem* **259**, 3933-43 (1984).
30. Chen, Z.J., Parent, L. & Maniatis, T. Site-specific phosphorylation of IkappaBalpha by a novel ubiquitination-dependent protein kinase activity. *Cell* **84**, 853-62 (1996).
31. Cheng, Y., Zhang, Y. & McCammon, J.A. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: an ab initio QM/MM study. *J Am Chem Soc* **127**, 1553-62 (2005).
32. Cho, H.S. *et al.* Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**, 756-60 (2003).
33. Choi, J., Chen, J., Schreiber, S.L. & Clardy, J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273**, 239-42 (1996).
34. Chou, M.M. & Hanafusa, H. A novel ligand for SH3 domains. The Nck adaptor protein binds to a serine/threonine kinase via an SH3 domain. *J Biol Chem* **270**, 7359-64 (1995).
35. Clark, K., Plater, L., Peggie, M. & Cohen, P. Use of the pharmacological inhibitor BX795 to study the regulation and physiological roles of TBK1 and IkappaB kinase epsilon: a distinct upstream kinase mediates Ser-172 phosphorylation and activation. *J Biol Chem* **284**, 14136-46 (2009).

36. Cohen, O., Feinstein, E. & Kimchi, A. DAP-kinase is a Ca2+/calmodulin-dependent, cytoskeletal-associated protein kinase, with cell death-inducing functions that depend on its catalytic activity. *Embo J* **16**, 998-1008 (1997).
37. Cook, A. *et al.* Structural studies on phospho-CDK2/cyclin A bound to nitrate, a transition state analogue: implications for the protein kinase mechanism. *Biochemistry* **41**, 7301-11 (2002).
38. Cornvik, T. *et al.* An efficient and generic strategy for producing soluble human proteins and domains in E. coli by screening construct libraries. *Proteins* **65**, 266-73 (2006).
39. Cowan, R.H., Davies, R.A. & Pinheiro, T.T. A screening system for the identification of refolding conditions for a model protein kinase, p38alpha. *Anal Biochem* **376**, 25-38 (2008).
40. Cozzetto, D. *et al.* Evaluation of template-based models in CASP8 with standard measures. *Proteins* **77**, 18-28 (2009).
41. Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-291 (1998).
42. Crameri, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **14**, 315-9 (1996).
43. Crivici, A. & Ikura, M. Molecular and structural basis of target recognition by calmodulin. *Annu Rev Biophys Biomol Struct* **24**, 85-116 (1995).
44. Dahlroth, S.L., Nordlund, P. & Cornvik, T. Colony filtration blotting for screening soluble expression in Escherichia coli. *Nat Protoc* **1**, 253-8 (2006).
45. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC Biotechnol* **7**, 32 (2007).
46. De Strooper, B. & Woodgett, J. Alzheimer's disease: Mental plaque removal. *Nature* **423**, 392-3 (2003).
47. Deiss, L.P., Feinstein, E., Berissi, H., Cohen, O. & Kimchi, A. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. *Genes Dev* **9**, 15-30 (1995).
48. Delhase, M., Hayakawa, M., Chen, Y. & Karin, M. Positive and negative regulation of IkappaB kinase activity through IKKbeta subunit phosphorylation. *Science* **284**, 309-13 (1999).
49. Dobson, C.M. Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**, 3-16 (2004).
50. Duffy, S., Tsao, K.L. & Waugh, D.S. Site-specific, enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides. *Anal Biochem* **262**, 122-8 (1998).
51. Dyson, M.R. *et al.* Identification of soluble protein fragments by gene fragmentation and genetic selection. *Nucleic Acids Res* **36**, e51 (2008).
52. Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol* **4**, 32 (2004).
53. Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289-98 (2006).

54. Fabian, M.A. *et al.* A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* **23**, 329-36 (2005).
55. Fisher, A.C., Kim, W. & DeLisa, M.P. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. *Protein Sci* **15**, 449-58 (2006).
56. Fitzgerald, D.J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nat Methods* **3**, 1021-32 (2006).
57. Funakoshi-Tago, M. *et al.* Functional role of c-Src in IL-1-induced NF-kappa B activation: c-Src is a component of the IKK complex. *J Biochem* **137**, 189-97 (2005).
58. Gangloff, Y.G. *et al.* Disruption of the mouse mTOR gene leads to early postimplantation lethality and prohibits embryonic stem cell development. *Mol Cell Biol* **24**, 9508-16 (2004).
59. Geering, B., Cutillas, P.R., Nock, G., Gharbi, S.I. & Vanhaesebroeck, B. Class IA phosphoinositide 3-kinases are obligate p85-p110 heterodimers. *Proc Natl Acad Sci U S A* **104**, 7809-14 (2007).
60. Gerth, M.L., Patrick, W.M. & Lutz, S. A second-generation system for unbiased reading frame selection. *Protein Eng Des Sel* **17**, 595-602 (2004).
61. Gozuacik, D. & Kimchi, A. Autophagy as a cell death and tumor suppressor mechanism. *Oncogene* **23**, 2891-906 (2004).
62. Grothues, D., Cantor, C.R. & Smith, C.L. PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nucl Acids Res* **21**, 1321-1322 (1993).
63. Guertin, D.A. & Sabatini, D.M. An expanding role for mTOR in cancer. *Trends Mol Med* **11**, 353-61 (2005).
64. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15**, 500-6 (2008).
65. Haacke, A., Fendrich, G., Ramage, P. & Geiser, M. Chaperone over-expression in Escherichia coli: apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates. *Protein Expr Purif* **64**, 185-93 (2009).
66. Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H. & Hard, T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. *Protein Sci* **11**, 313-21 (2002).
67. Hanks, S.K. & Hunter, T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *Faseb J* **9**, 576-96 (1995).
68. Hantschel, O., Rix, U. & Superti-Furga, G. Target spectrum of the BCR-ABL inhibitors imatinib, nilotinib and dasatinib. *Leuk Lymphoma* **49**, 615-9 (2008).
69. Hara, K. *et al.* Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell* **110**, 177-89 (2002).
70. Hardie, G.a.H., S. *The Protein Kinase FactsBook*, (Academic Press, San Diego, 1995).
71. Harikumar, K.B. *et al.* Modification of the cysteine residues in IkappaBalpha kinase and NF-kappaB (p65) by xanthohumol leads to suppression of NF-kappaB-regulated gene products and potentiation of apoptosis in leukemia cells. *Blood* **113**, 2003-13 (2009).
72. Harris, J. *et al.* Nuclear accumulation of cRel following C-terminal phosphorylation by TBK1/IKK epsilon. *J Immunol* **177**, 2527-35 (2006).
73. Harrison, D.E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392-5 (2009).

74. Hart, D.J. & Tarendeau, F. Combinatorial library approaches for improving soluble protein expression in Escherichia coli. *Acta Crystallogr D Biol Crystallogr* **62**, 19-26 (2006).
75. Hartmann-Petersen, R. & Gordon, C. Integral UBL domain proteins: a family of proteasome interacting proteins. *Semin Cell Dev Biol* **15**, 247-59 (2004).
76. Hayden, M.S. & Ghosh, S. Shared principles in NF-kappaB signaling. *Cell* **132**, 344-62 (2008).
77. Heddle, C. & Mazaleyrat, S.L. Development of a screening platform for directed evolution using the reef coral fluorescent protein ZsGreen as a solubility reporter. *Protein Eng Des Sel* **20**, 327-37 (2007).
78. Henikoff, S. Ordered deletions for DNA sequencing and in vitro mutagenesis by polymerase extension and exonuclease III gapping of circular templates. *Nucleic Acids Res* **18**, 2961-6 (1990).
79. Henikoff, S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359 (1984).
80. Hofinger, E.S. *et al.* Recombinant human hyaluronidase Hyal-1: insect cells versus Escherichia coli as expression system and identification of low molecular weight inhibitors. *Glycobiology* **17**, 444-53 (2007).
81. Hu, P., Mondino, A., Skolnik, E.Y. & Schlessinger, J. Cloning of a novel, ubiquitously expressed human phosphatidylinositol 3-kinase and identification of its binding site on p85. *Mol Cell Biol* **13**, 7677-88 (1993).
82. Huang, C.H. *et al.* The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations. *Science* **318**, 1744-8 (2007).
83. Hudes, G.R. mTOR as a target for therapy of renal cancer. *Clin Adv Hematol Oncol* **5**, 772-4 (2007).
84. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275-82 (2002).
85. Ikeda, F. *et al.* Involvement of the ubiquitin-like domain of TBK1/IKK-i kinases in regulation of IFN-inducible genes. *Embo J* **26**, 3451-62 (2007).
86. Inbal, B., Bialik, S., Sabanay, I., Shani, G. & Kimchi, A. DAP kinase and DRP-1 mediate membrane blebbing and the formation of autophagic vesicles during programmed cell death. *J Cell Biol* **157**, 455-68 (2002).
87. Inbal, B. *et al.* DAP kinase links the control of apoptosis to metastasis. *Nature* **390**, 180-4 (1997).
88. Inbal, B., Shani, G., Cohen, O., Kissil, J.L. & Kimchi, A. Death-associated protein kinase-related protein 1, a novel serine/threonine kinase involved in apoptosis. *Mol Cell Biol* **20**, 1044-54 (2000).
89. Jacobs, S.A., Podell, E.R., Wuttke, D.S. & Cech, T.R. Soluble domains of telomerase reverse transciptase identified by high-throughput screening. *Protein Sci* **14**, 2051-2058 (2005).
90. Jang, C.W. *et al.* TGF-beta induces apoptosis through Smad-mediated expression of DAP-kinase. *Nat Cell Biol* **4**, 51-8 (2002).
91. Jia, S. *et al.* Essential roles of PI(3)K-p110beta in cell growth, metabolism and tumorigenesis. *Nature* **454**, 776-9 (2008).
92. Jin, Y.J., Burakoff, S.J. & Bierer, B.E. Molecular cloning of a 25-kDa high affinity rapamycin binding protein, FKBP25. *J Biol Chem* **267**, 10942-5 (1992).
93. Johnson, L.N. Protein kinase inhibitors: contributions from structure to clinical compounds. *Q Rev Biophys* **42**, 1-40 (2009).

94. Kannan, N., Taylor, S.S., Zhai, Y., Venter, J.C. & Manning, G. Structural and functional diversity of the microbial kinome. *PLoS Biol* **5**, e17 (2007).
95. Karin, M. How NF-kappaB is activated: the role of the IkappaB kinase (IKK) complex. *Oncogene* **18**, 6867-74 (1999).
96. Karin, M., Yamamoto, Y. & Wang, Q.M. The IKK NF-kappa B system: a treasure trove for drug development. *Nat Rev Drug Discov* **3**, 17-26 (2004).
97. Karkkainen, H.R. *et al.* A 96-well format for a high-throughput baculovirus generation, fast titering and recombinant protein production in insect and mammalian cells. *BMC Res Notes* **2**, 63 (2009).
98. Kawai, T., Matsumoto, M., Takeda, K., Sanjo, H. & Akira, S. ZIP kinase, a novel serine/threonine kinase which mediates apoptosis. *Mol Cell Biol* **18**, 1642-51 (1998).
99. Kawasaki, M. & Inagaki, F. Random PCR-based screening for soluble domains using green fluorescent protein. *Biochem Biophys Res Commun* **280**, 842-844 (2001).
100. Keenan, R.J., Siehl, D.L., Gorton, R. & Castle, L.A. DNA shuffling as a tool for protein crystallization. *Proc Natl Acad Sci U S A* **102**, 8887-92 (2005).
101. Kendrew, J.C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-6 (1958).
102. King, D.A. *et al.* Domain structure and protein interactions of the silent information regulator Sir3 revealed by screening a nested deletion library of protein fragments. *J Biol Chem* **281**, 20107-19 (2006).
103. Knaust, R.K.C. & Nordlund, P. Screening for soluble expression of recombinant proteins in a 96-well format. *Anal Biochem* **297**, 79-85 (2001).
104. Kornev, A.P., Haste, N.M., Taylor, S.S. & Eyck, L.F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A* **103**, 17783-8 (2006).
105. Kornev, A.P., Taylor, S.S. & Ten Eyck, L.F. A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci U S A* **105**, 14377-82 (2008).
106. Koyasu, S. The role of PI3K in immune cells. *Nat Immunol* **4**, 313-9 (2003).
107. Krappmann, D. *et al.* The I kappa B kinase (IKK) complex is tripartite and contains IKK gamma but not IKAP as a regular component. *J Biol Chem* **275**, 29779-87 (2000).
108. Kristensen, P. & Winter, G. Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des* **3**, 321-8 (1998).
109. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-8 (2003).
110. Kute, T. *et al.* Development of Herceptin resistance in breast cancer cells. *Cytometry A* **57**, 86-93 (2004).
111. Kwak, Y.T., Guo, J., Shen, J. & Gaynor, R.B. Analysis of domains in the IKKalpha and IKKbeta proteins that regulate their kinase activity. *J Biol Chem* **275**, 14752-9 (2000).
112. Leevers, S.J., Vanhaesebroeck, B. & Waterfield, M.D. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr Opin Cell Biol* **11**, 219-25 (1999).
113. Lesley, S.A., Graziano, J., Cho, C.Y., Knuth, M.W. & Klock, H.E. Gene expression response to misfolded protein as a screen for soluble recombinant protein. *Protein Eng* **15**, 153-60 (2002).
114. Li, Q., Van Antwerp, D., Mercurio, F., Lee, K.F. & Verma, I.M. Severe liver degeneration in mice lacking the IkappaB kinase 2 gene. *Science* **284**, 321-5 (1999).

115. Li, S., Wang, L. & Dorf, M.E. PKC phosphorylation of TRAF2 mediates IKKalpha/beta recruitment and K63-linked polyubiquitination. *Mol Cell* **33**, 30-42 (2009).
116. Liu, Y.C., Penninger, J. & Karin, M. Immunity by ubiquitylation: a reversible process of modification. *Nat Rev Immunol* **5**, 941-52 (2005).
117. Lutz, S., Fast, W. & Benkovic, S.J. A universal, vector-based system for nucleic acid reading-frame selection. *Protein Eng* **15**, 1025-30 (2002).
118. Lutz, S., Ostermeier, M. & Benkovic, S.J. Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. *Nucleic Acids Res* **29**, E16 (2001).
119. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-34 (2002).
120. Margutti, S. & Laufer, S.A. Are MAP kinases drug targets? Yes, but difficult ones. *ChemMedChem* **2**, 1116-40 (2007).
121. Marsden, B.D. & Knapp, S. Doing more than just the structure-structural genomics in kinase drug discovery. *Curr Opin Chem Biol* **12**, 40-5 (2008).
122. Martelli, A.M. *et al.* Targeting the PI3K/AKT/mTOR signaling network in acute myelogenous leukemia. *Expert Opin Investig Drugs* **18**, 1333-49 (2009).
123. Maxwell, K.L., Mittermaier, A.K., Forman-Kay, J.D. & Davidson, A.R. A simple in vivo assay for increased protein solubility. *Protein Sci* **8**, 1908-1911 (1999).
124. May, M.J. *et al.* Selective inhibition of NF-kappaB activation by a peptide that blocks the interaction of NEMO with the IkappaB kinase complex. *Science* **289**, 1550-4 (2000).
125. May, M.J., Larsen, S.E., Shim, J.H., Madge, L.A. & Ghosh, S. A novel ubiquitin-like domain in IkappaB kinase beta is required for functional activity of the kinase. *J Biol Chem* **279**, 45528-39 (2004).
126. May, M.J., Marienfeld, R.B. & Ghosh, S. Characterization of the Ikappa B-kinase NEMO binding domain. *J Biol Chem* **277**, 45992-6000 (2002).
127. Minard, P., Scalley-Kim, M., Watters, A. & Baker, D. A "loop entropy reduction" phage-display selection for folded amino acid sequences. *Protein Sci* **10**, 129-34 (2001).
128. Mozzarelli, A. & Rossi, G.L. Protein function in the crystal. *Annu Rev Biophys Biomol Struct* **25**, 343-65 (1996).
129. Nelson, D.a.C., M. *Lehninger Principles of Biochemistry*, (Worth Publishers, New York, 2003).
130. Noble, M.E., Endicott, J.A. & Johnson, L.N. Protein kinase inhibitors: insights into drug design from structure. *Science* **303**, 1800-5 (2004).
131. Nowell, P.C. & Hungerford, D.A. Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* **25**, 85-109 (1960).
132. Oppermann, H., Levinson, A.D., Varmus, H.E., Levintow, L. & Bishop, J.M. Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src). *Proc Natl Acad Sci U S A* **76**, 1804-8 (1979).
133. Ostermeier, M. & Lutz, S. The creation of ITCHY hybrid protein libraries. *Methods Mol Biol* **231**, 129-141 (2003).
134. Patrick, W.M., Firth, A.E. & Blackburn, J.M. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng* **16**, 451-7 (2003).

135. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88 (2006).
136. Pedelacq, J.D. *et al.* Engineering soluble proteins for structural genomics. *Nat Biotechnol* **20**, 927-932 (2002).
137. Pengelley, S.C. *et al.* A suite of parallel vectors for baculovirus expression. *Protein Expr Purif* **48**, 173-81 (2006).
138. Petsko, G.a.R., D. *Protein Structure and Function*, (New Science Press, London, 2004).
139. Prodromou, C., Savva, R. & Driscoll, P.C. DNA fragmentation-based combinatorial approaches to soluble protein expression Part I. Generating DNA fragment libraries. *Drug Discov Today* **12**, 931-8 (2007).
140. Qian, B. *et al.* High-resolution structure prediction and the crystallographic phase problem. *Nature* **450**, 259-64 (2007).
141. Raveh, T. & Kimchi, A. DAP kinase-a proapoptotic gene that functions as a tumor suppressor. *Exp Cell Res* **264**, 185-92 (2001).
142. Reich, S. *et al.* Combinatorial Domain Hunting: An effective approach for the identification of soluble protein domains adaptable to high-throughput applications. *Protein Sci* **15**, 2356-65 (2006).
143. Rensing, L. Periodic geophysical and biological signals as Zeitgeber and exogenous inducers in animal organisms. *Int J Biometeorol* **16 Suppl**, 113-25 (1972).
144. Rosenberg, O.S., Deindl, S., Sung, R.J., Nairn, A.C. & Kuriyan, J. Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* **123**, 849-60 (2005).
145. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-5 (2008).
146. Rowley, J.D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-3 (1973).
147. Rushe, M. *et al.* Structure of a NEMO/IKK-associating domain reveals architecture of the interaction site. *Structure* **16**, 798-808 (2008).
148. Sambrook, J.a.R., D. *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, 2001).
149. Sanjo, H., Kawai, T. & Akira, S. DRAKs, novel serine/threonine kinases related to death-associated protein kinase that trigger apoptosis. *J Biol Chem* **273**, 29066-71 (1998).
150. Sarbassov, D.D. *et al.* Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Curr Biol* **14**, 1296-302 (2004).
151. Savva, R., Prodromou, C. & Driscoll, P.C. DNA fragmentation based combinatorial approaches to soluble protein expression Part II: library expression, screening and scale-up. *Drug Discov Today* **12**, 939-47 (2007).
152. Schatz, P.J. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (NY)* **11**, 1138-1143 (1993).
153. Scholle, M.D., Collart, F.R. & Kay, B.K. In vivo biotinylated proteins as targets for phage-display selection experiments. *Protein Expr Purif* **37**, 243-52 (2004).

154. Sehgal, S.N., Baker, H. & Vezina, C. Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *J Antibiot (Tokyo)* **28**, 727-32 (1975).
155. Service, R.F. Structural biology. Protein structure initiative: phase 3 or phase out. *Science* **319**, 1610-3 (2008).
156. Shih, Y.P. *et al.* High-throughput screening of soluble recombinant proteins. *Protein Sci* **11**, 1714-9 (2002).
157. Sieber, V., Pluckthun, A. & Schmid, F.X. Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16**, 955-60 (1998).
158. Stehelin, D., Fujita, D.J., Padgett, T., Varmus, H.E. & Bishop, J.M. Detection and enumeration of transformation-defective strains of avian sarcoma virus with molecular hybridization. *Virology* **76**, 675-84 (1977).
159. Szarek, P., Dyguda-Kazimierowicz, E., Tachibana, A. & Sokalski, W.A. Physical nature of intermolecular interactions within cAMP-dependent protein kinase active site: differential transition state stabilization in phosphoryl transfer reaction. *J Phys Chem B* **112**, 11819-26 (2008).
160. Tagwerker, C. *et al.* A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol Cell Proteomics* **5**, 737-48 (2006).
161. Tarendeau, F. *et al.* Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14**, 229-33 (2007).
162. Tarendeau, F. *et al.* Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* **4**, e1000136 (2008).
163. Tereshko, V., Teplova, M., Brunzelle, J., Watterson, D.M. & Egli, M. Crystal structures of the catalytic domain of human protein kinase associated with apoptosis and tumor suppression. *Nat Struct Biol* **8**, 899-907 (2001).
164. Thorstenson, Y.R., Hunicke-Smith, S.P., Oefner, P.J. & Davis, R.W. An automated hydrodynamic process for controlled, unbiased DNA shearing. *Genome Methods* **8**, 848-855 (1998).
165. Tojima, Y. *et al.* NAK is an IkappaB kinase-activating kinase. *Nature* **404**, 778-82 (2000).
166. Vanhaesebroeck, B. *et al.* Synthesis and function of 3-phosphorylated inositol lipids. *Annu Rev Biochem* **70**, 535-602 (2001).
167. Velentza, A.V., Schumacher, A.M., Weiss, C., Egli, M. & Watterson, D.M. A protein kinase associated with apoptosis and tumor suppression: structure, activity, and discovery of peptide substrates. *J Biol Chem* **276**, 38956-65 (2001).
168. Vezina, C., Kudelski, A. & Sehgal, S.N. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *J Antibiot (Tokyo)* **28**, 721-6 (1975).
169. Waldo, G.S. Improving protein folding efficiency by directed evolution using the GFP folding reporter. *Methods Mol Biol* **230**, 343-59 (2003).
170. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17**, 691-695 (1999).
171. Walker, E.H., Perisic, O., Ried, C., Stephens, L. & Williams, R.L. Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* **402**, 313-20 (1999).

172. Wang, W. *et al.* Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. *Proc Natl Acad Sci U S A* **103**, 3563-8 (2006).
173. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-45 (2004).
174. Watson, J.D. *et al.* Towards fully automated structure-based function prediction in structural genomics: a case study. *J Mol Biol* **367**, 1511-22 (2007).
175. Waugh, D.S. Making the most of affinity tags. *Trends Biotechnol* **23**, 316-20 (2005).
176. Welch, M. *et al.* Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One* **4**, e7002 (2009).
177. Westbrook, C.A. The ABL oncogene in human leukemias. *Blood Rev* **2**, 1-8 (1988).
178. Wigley, W.C., Stidham, R.D., Smith, N.M., Hunt, J.F. & Thomas, P.J. Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein. *Nat Biotechnol* **19**, 131-135 (2001).
179. Woronicz, J.D., Gao, X., Cao, Z., Rothe, M. & Goeddel, D.V. IkappaB kinase-beta: NF-kappaB activation and complex formation with IkappaB kinase-alpha and NIK. *Science* **278**, 866-9 (1997).
180. Wullschleger, S., Loewith, R. & Hall, M.N. TOR signaling in growth and metabolism. *Cell* **124**, 471-84 (2006).
181. Wunderlich, M., Martin, A. & Schmid, F.X. Stabilization of the cold shock protein CspB from Bacillus subtilis by evolutionary optimization of Coulombic interactions. *J Mol Biol* **347**, 1063-76 (2005).
182. Xia, Z.P. *et al.* Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature* **461**, 114-9 (2009).
183. Xu, M., Skaug, B., Zeng, W. & Chen, Z.J. A Ubiquitin replacement strategy in human cells reveals distinct mechanisms of IKK activation by TNFalpha and IL-1beta. *Mol Cell* **36**, 302-14 (2009).
184. Xu, W., Doshi, A., Lei, M., Eck, M.J. & Harrison, S.C. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol Cell* **3**, 629-38 (1999).
185. Yamaoka, S. *et al.* Complementation cloning of NEMO, a component of the IkappaB kinase complex essential for NF-kappaB activation. *Cell* **93**, 1231-40 (1998).
186. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-19 (1985).
187. Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B. & Kuriyan, J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115-26 (2001).
188. Zaki, M.a.B., C. *Protein Structure Prediction*, (Humana Press, 2008).
189. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
190. Zhao, L. & Vogt, P.K. Class I PI3K in oncogenic cellular transformation. *Oncogene* **27**, 5486-96 (2008).

"*Measure thrice, cut once.*"

--- Bulgarian Proverb

Appendix

General Methods and Protocols

Ligation

Regularly, for cloning, 150-300 ng total DNA in miniprep elution buffer (Tris HCl pH 8) was ligated 3:1 insert:backbone molar ratio with using a rapid ligation kit (Roche) for 5-15 min at room temperature.

Colony PCR screen

Single colony were isolated from petri dishes and resuspended in 20 µl of ddH_2O. The PCR reaction master mix comprised 400 nM of T7for (5′ GCGAA ATTAA TACGA CTCAC TATAG G 3′) and T7rev (5′ GCTAG TTATT GCTCA GCGGT GGC 3′), 200 µM dNTP, 1 U *Taq* polymerase and 2 µl of resuspended cells. The PCR program used was 94 °C 2 min, 94 °C 45 sec, 57 °C 30 sec, 72 °C 1 min/kb for 25 cycles finishing with additional double-length elongation step at 72 °C. This protocol was modified in a gene-specific manner by varying primers and T_m. After completion of the amplification reaction, it was mixed with 3 µl loading blue and run on 1% agarose gel electrophoresis at 5-10 V/cm. Finally, the agarose gel was visualised at 360 nm UV with agarose gel illuminator.

DNA Sequencing

Three different DNA sequencing services have been used throughout my work. Initially, EMBL-Heidelberg, Germany, then Macrogen, Korea and Cogenics, Grenoble were used. Usually, about 5 µl of 100 ng/µl plasmid was sent for sequencing along with 20 µl of custom primers at 5 pM. Usually, readouts were within 600-800 bp after the sequencing primer.

Electrocompetent Cell Preparation

A single colony was grown in 5 ml overnight TB culture with the appropriate antibiotics. Next, 50 ml TB supplemented with antibiotics was inoculated with a 1/200 dilution of the overnight and grown at 37 °C until the cell density reached was OD_{600} 0.8. The culture was put immediately on ice and centrifuged in a refrigerated centrifuge (5804R, Eppendorf) at 4,000 g for 10 min. The supernatant was removed and pellet gently resuspended with 1 pellet volume of cold water, previously cooled on ice. The centrifugation step was repeated and supernatant removed. Then, the pellet was washed with 0.5 volumes cold water. The final

wash was with 1/50 volume cold water and cells were either used directly for transformation or frozen with 10% glycerol at -80 °C.

Bacterial Transformations

Heat shock was performed usually with te cloning strain like Mach1 (Invitrogen). 12-50 µl of cells were thawed on ice and incubated with 1 µl ligation product for 1 min on ice then put at 42 °C for 42 s and immediately back on ice for 1 min. Recovery was performed with 200-1000 µl SOC for 40-60 min at 37 °C in a shaking incubator, normally 180 rpm. For cloning, 50 µl of transformant mix was streaked on LB agar petri dishes with the appropriate antibiotics and grown overnight at 37 °C.

Electroporation was performed with lab-prepared electrocompetent arabinose or IPTG inducibile BL21 (DE3) codon plus (RIL) cells. They were thawed on ice and 50 µl were mixed with 10-100 ng plasmid in 10 mM Tris HCl pH 8. The mixture was transferred to a electroporation cuvette, also on ice. The electroporator (BioRad) was set to 1.8 kV and the cuvette subjected to an electric pulse. When an efficiency above 4.5 ms was achieved, cells were recovered in SOC medium and plated as described above.

Cell Lysis

Typically, 40 ml of resuspended culture was lysed in a French Press (FA-078, SLM Aminco) at 4 °C with pressure of 1000 psi and 3-4 cycles of the lysate through the cell (FA-072, SLM Aminco).

Sonication was performed at level 8 of the sonicator (XL2020, Misonix) for 4 min with total impulse alternating 10 s on and 30 s off. The lyses were performed on ice.

More recently for cell lysis, a Microfluidiser (M-110L, Microfluidics) was used at 18 kpsi, 4 °C. The lysate was cycled 10 times to assure good lysis.

Typical E. coli Expression Test

A fresh starter culture of 5-10 ml was started from a single colony, 180 rpm, 18-20 h, 37 °C. Typically, 50 ml - 1 L TB media supplemented with the appropriate antibiotics was inoculated with 1/100 dilution of the starter culture and grown at 37 °C, 180 rpm until OD_{600} 0.6. Cultures were induced with 0.1 mM IPTG or 0.2% w/v arabinose depending on the promoter system of the strain. Usually, protein expression was carried out at 25 °C and 160 rpm overnight. Pellets were harvested at 8,000 rpm for 10 min with Eppendorf 5804R centrifuge or 6,000 g for 20 min in a large rotor centrifuge (Beckman). Pellets were resuspended in a ratio 1:1 to 1:2 with sonication buffer (50 mM Tris HCl pH 7, 300 mM NaCl, 5 mM imidazole pH 7, 5 mM βME, 1 tablet/50 ml Roche Complete Inhibitor and 1/1000 dilution of Benzonase). Sonication for 4 min alternating 10 s pulse and 30 s rest on ice was used for lysis. Total fractions were collected and soluble protein fractionated by centrifugation at 15,550g for 30 min. The resin for affinity purification was prepared for immobilisation of the target protein by washing with water and lysis buffer. Soluble fractions were passed through the resins in gravity columns and washed with 20 resin volumes (50 mM

Tris HCl pH7, 300 mM NaCl, 5 mM imidazole and 5 mM β-ME). A second wash of 10 resin volumes with 10 mM imidazole was performed and then elution either with buffer supplemented with 300 mM imidazole, or with increasing imidazole concentration steps of two resin volumes each: 100 mM, 200 mM, 300 mM and 400 mM.

Protein Concentration Measurements

Regularly, Bradford reagent (Sigma) was used for total and soluble fractions as well as purified fractions following protein expression tests. In such cases, either 990 µl of Bradford reagent and 10 µl of protein sample or volumes scaled twofold down were used against a calibration curve in the linear concentration range from 200 µg/ml to 6mg/ml. These measurements were taken with a Biophotometer (Eppendorf) with calibration based on BSA. When precise measurements were required, a ND-1000 spectrometer (NanoDrop) was used for measuring absorption at 280 nm of the sample diluted in an equal volume of 6 M GnHCl and concentration calculated using the computed extinction coefficient of the protein.

SDS-PAGE

Usually, 0.75 mm thick gels 15% SDS-PAGE were used. The resolving gel consists of 5 ml 30% w/v acrylamide/bisacrylamide (Sigma), 2.5 ml Tris pH 8.8, 2.2 ml ddH_2O, 50 µl of 20% w/v SDS solution, 7 µl of TEMED (MP) and 50 µl of 50% v/v ammonium persulphate solution. It was cast and overlaid with 100% EtOH. Then, the stacking gel mixture comprising 1.33 ml 30% w/v acrylamide/bisacrylamide, 2.5 ml Tris pH 6.8, 6.2 ml ddH_2O, 50 µl of 20% SDS solution, 7 µl of TEMED and 50 µl of APS. In case of high-throughput gels run on the the Ruby System (Amersham), gels were prepared by multiplying these volumes 3-fold. Occasionally, the large gels were supplemented with Rhinohide (Invitrogen) to increase their strength.

Western Blot

Transfer was done on either PVDF (Millipore) or nitrocellulose membranes (Amersham) for 1 h with constant 180 V in the BioRad dry blotter or Ruby western blotting module. When PVDF membranes were used activation for 15 s with 100% methanol preceded transfer with 1× Tris-Glycine and 10% Methanol. When nitrocellulose was used, the transfer was done with 1× Tris-Glycine and 20% ethanol. After completion of the transfer the membrane was washed with water and then 0.1% Tween-PBS for 5-10 min. Then, the membrane was usually blocked overnight with SuperBlock (37516, Thermo Scientific Pierce).

Western Blot Quantification

Initially, the proteins of interest along with a well-characterised protein as a control were expressed supplementing 50 µM biotin in the culture medium. The control protein was purified to homogeneity and its concentration estimated by its absorbance at 280 nm. The control, well-characterised protein was used to prepare a calibration curve of known concentrations ranging usually from 0 to 2000 ng and then, both the proteins of interest and the calibration curve were transferred together to minimise transfer-related variation. Then, the protein blots were hybridised with streptavidin- Alexa488 (Amersham). Finally, they

were visualised with a Typhoon fluorescence scanner (Amersham) and the digital images evaluated with ImageQuant (Amersham). The pixel intensities for each of the bands were measured and the calibration curve was prepared (Fig. 1). The latter was used for the estimation of the mass of protein visible on the western blot. Finally, the concentration of the protein was calculated taking into account the mass loaded.

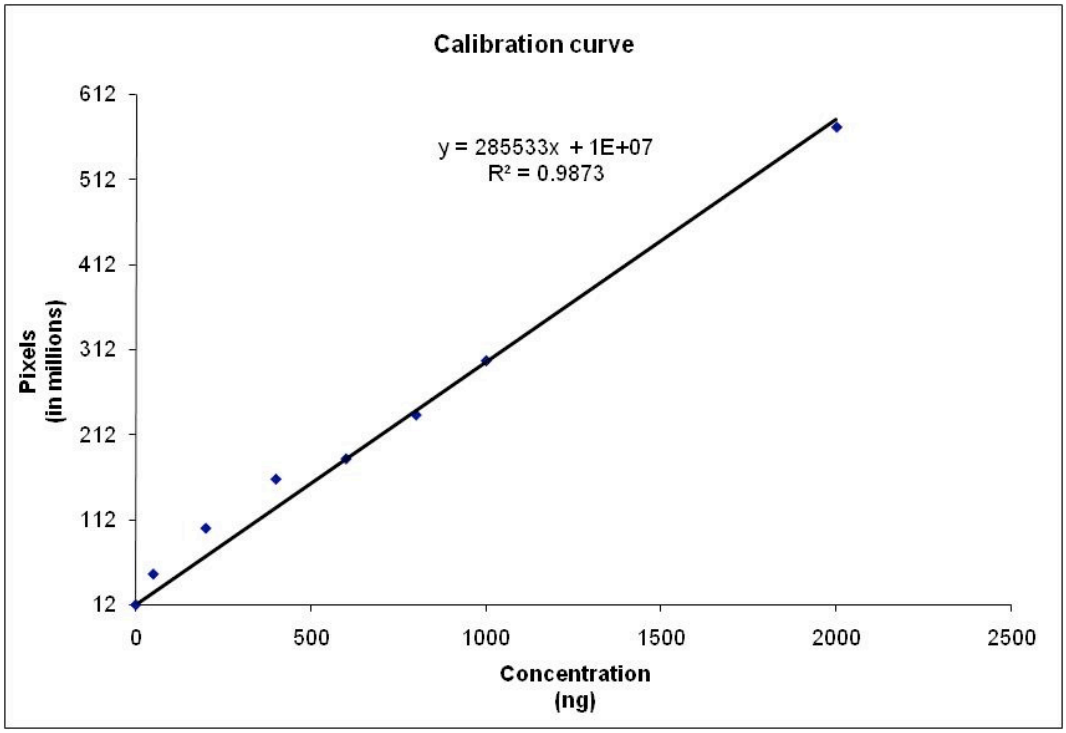


FIGURE A.1: TYPICAL CALIBRATION CURVE USED FOR FLUORESCENT WESTERN BLOT QUANTIFICATION

Materials and Kits

Molecular Weight Markers
2-log DNA Ladder (N3200S, NEB)
Broad Range Protein Marker (P7702S, NEB)
Prestained Broad Range Protein Marker (P7708S, NEB)
Rainbow Protein Marker (RPN850E, Amersham)

E. Coli Cell Strains

OMNImax (C8540-03, Invitrogen)
Genotype
F′ proAB+ lacIq lacZΔM15 Tn10(TetR) Δ(ccdAB) mcrA Δ(mrr-hsdRMS-mcrBC) φ80(lacZ)ΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD

Mach1 (C8620-03, Invitrogen)
Genotype
F- φ80(lacZ)ΔM15 ΔlacX74 hsdR(rK-mK+) ΔrecA1398 endA1 tonA

Molecular Biology Kits
Nucleospin (Macherey, Nagel)
Miniprep (Quiagen)
QIAEX (Quiagen)
QIAEX II Gel Extraction Kit (Quiagen)