Identification of protein-protein-interactions *in vitro* based on high-density protein arrays

vorgelegt von

Diplom-Ingenieur

Thomas Faupel

aus Berlin

von der Fakultät III – Prozesswissenschaften

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Helmut Görisch

Gutachter: Prof. Dr. Ulf Stahl

Gutachter: Dr. Konrad Büssow

Tag der wissenschaftlichen Aussprache: 23. September 2004

Berlin 2004

D83

Contents

1 Introduction

The recent completion of a number of genome sequencing projects has shifted the focus from individual studies of genes to large-scale studies of genes. For organisms with fully sequenced genomes, DNA microarrays are an extremly powerful tool to measure the mRNA expression level of all genes (Lockhart and Winzeler, 2000) and have proven invaluable for distinguishing where or when clustered groups of genes are turned on in different cell types or in response to different physiological conditions. The sequence data and mRNA abundance of whole genomes are in most cases, however, insufficient for understanding the biological function of the gene products. It is the proteins and their interactions which determine the functions within a living cell. Therefore, considerable attention is now being drawn to the development of large-scale screening techniques allowing for a comprehensive analysis of protein-protein-interactions on a genome-wide level.

Traditionally, protein-protein-interactions have been studied on a case-by-case basis using biochemical, biophysical or genetic techniques (reviewed in Phizicky and Fields, 1995). These traditional methods cannot realistically be applied to elucidate the interaction of every protein in a cell, tissue, or organism. It was concluded from the Human Genome Project (HGP) that there are only about 30,000 to 40,000 genes in human (Ewing and Green, 2000; Claverie, 2001), but due to splice variants and posttranslational modifications it was estimated that the human proteome could contain from as few as 100,000 different proteins, to as many as a few millions (Auerbach *et al*., 2002). Only a minority of all these proteins has yet been assigned a function. It is evident that all these proteins rarely act alone; rather, they build highly dynamic networks, in which they regulate and support each other depending on the particular cellular tasks they perform or the certain external stimuli and developmental stages they undergo within a cell (Tucker *et al*., 2001; Vidal, 2001).

The speed with which novel proteins are beeing discovered or predicted from accumulated sequence data has generated the demand for more empirical high-throughput methods that are amenable to assay thousands of different proteins simultaneously and to assign a function to proteins on a broader level. High-throughput screens on a large scale can have the advantage over individual hypothesis-driven approaches both in the number of novel protein-protein-interactions detected and in the discovery of yet uncharacterized

proteins. For such uncharacterized proteins, physical association with well-characterized proteins readily provides a valuable clue to their function or biological role within a protein network. A number of recently described technologies have gained a broad interest and have provided ways to address these issues. Depending on the location where the physical association of proteins is assayed, protein-protein-interaction screenings on a large-scale can be classified into *in vivo* and *in vitro* approaches.

1.1 Protein-protein-interaction screenings *in vivo*

1.1.1 Yeast two-hybrid screens

The most popular means to carry out a large-scale protein-protein-interaction screen *in vivo* is the yeast two-hybrid system (Fields and Song, 1989). The yeast two-hybrid system makes use of the circumstance that transcription factors are generally composed of two separable domains, a sequence-specific DNA-binding domain (DB) and an activation domain (AD) typically of the transcription factor Gal4p. When a protein X is fused to DB (DB-X) and a protein Y to AD (AD-Y), an interaction between X and Y can be detected by the reconstitution of a functional transcription factor DB-X/AD-Y. When both hybrids are co-expressed in the yeast *Saccharomyces cerevisiae* harbouring chromosomally integrated reporter genes, physical association between DB-X and AD-Y reconstitutes the function of the transcription factor Gal4p, thereby allowing transcription of reporter genes regulated by promoters to which DB can bind. The colorimetric reporter *lacZ,* encoding ß-galactosidase, which cleaves substrates such as X-Gal to produce a blue pigment, and auxotrophic reporters such as *HIS3* which allow growth on a defined medium lacking histidine, are used in yeast two-hybrid screens for detection and selection of protein-protein-interactions. In general, a two-hybrid interaction is judged positive when the expression of at least two reporter genes is activated. In the most common application, X is a protein of interest fused to DB, usually referred to as the bait and Y is an unknown protein fused to AD, usually referred to as the prey.

Pioneering small-scale two-hybrid screens for certain subsets of proteins have been reported on *Drosophila* (Finley and Brent, 1994), the T7 phage of *E. coli* (Bartel *et al*., 1996) and yeast (Fromont-Racine *et al*., 1997, Flores *et al*., 1999). Yeast was the first eukaryotic organism with a completely sequenced genome (Goffeau *et al*., 1996). All of

the proteins could be predicted and used for the comprehensive analysis of protein-protein-interactions in the yeast.

Recently, the first large-scale two-hybrid interaction analyses were undertaken for the entire *S. cerevisiae* genome, in which a bait and a prey for each of the 6,000 predicted yeast proteins was constructed. In the first two-hybrid approach 62 pools of each type of yeast transformant were generated, containing up to 96 independent hybrids each, followed by a systematic mating of bait and prey pools to yield 3,844 sets of diploids (Ito *et al*., 2000). Subsequent recovery and sequencing of DNA from diploids representing a positive phenotype identified the genes encoding the pairs of interacting proteins. A second two-hybrid approach made use of a yeast colony array format, in which each diploid colony expressed a pair of proteins (Uetz *et al*., 2000). Because the particular protein pair expressed in each colony is defined by its position in the array, positive signals identified interacting proteins without further characterization, thus obviating the need for DNA purification and sequencing.

The two independent two-hybrid approaches collectively identified over 4,000 protein-protein-interactions in *S. cerevisiae* but they suffered from a considerable number of both false positives and false negatives. Approximately 50 % of the revealed interactions are expected to be true interactions (Tucker *et al*., 2001). Moreover, among the 691 and 1,533 core binary interactions found by Uetz *et al*. and Ito *et al*., respectively, only 141 interactions are shared. The fraction shared by the two approaches showed a significantly higher rate of true positives as evaluated from already known interactions in the literature (Ito *et al*., 2001).

There are a number of reasons for both false negatives and false positives generated in the large-scale two-hybrid experiments of Ito *et al*. and Uetz *et al*., respectively. The two-hybrid technique is designed to detect binary interactions between proteins in the nucleus of a yeast cell, so many proteins are not in their native compartment and the positive or negative result for a pairwise interaction is unrelated to the physiological setting.

Interactions that escape the screening may be explained by the use of full length proteins versus variously truncated protein fragments. False negative results arise when proteins are displayed inappropriately and fail to fold correctly in yeast, so that possible interaction surfaces of the proteins are masked. A two-hybrid interaction is often hardly detectable between two full length proteins but becomes significantly stronger when

truncated protein fragments with appropriate domain boundaries are used (Fromont-Racine *et al*., 2000).

False positives are frequently generated from DB-X fusions that can activate transcription independently of an interaction with an AD-Y protein. Such self-activators include proteins that act as transcriptional activators in their respective organisms and maintain this ability in yeast (Du *et al*., 1996). The problem of transcriptionally active baits is not restricted to proteins that normally function as transcription factors. They also include proteins that act in other processes but exhibit transcriptional activity when tethered to a promoter in yeast cells (Hu *et al*., 1997). In addition, high-throughput strategies are complicated by the occurrence of spurious DB-X self activators originated from cloning artifacts. These can include cDNA fragments cloned out-of-frame in the random strategy (Ma and Ptashne, 1987). Finally, positive phenotypes do not in every case indicate direct binding of bait and prey protein. Two hybrid interactions may also be mediated by endogenous yeast proteins that link the bait to the prey.

1.1.2 *in vivo* pulldown assays

Protein-protein-interactions have recently also been assayed *in vivo* on a large scale by pulldown experiments. As for the two-hybrid approach, yeast has been the major proving ground for this technique. First, individual bait proteins of *S. cerevisiae* were epitope-tagged to generate fusion proteins, which were transiently expressed in yeast cells and used to pull down associated proteins. Protein complexes were then purified from total cell lysates by affinity chromatography on agarose beads. Obtained protein assemblies were separated by denaturing SDS-PAGE gel electrophoresis and subsequently visualized by colloidal Coomassie stain. Finally, individual protein bands were excised from the SDS-PAGE gel, digested by trypsin, analysed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and identified by their peptide mass fingerprint. This technique has been readily adapted to the parallel analysis in microtiter plate formats and allows high-throughput investigation of protein complexes in the yeast.

Two different large-scale approaches for such *in vivo* pulldowns have been reported. Tandem affinity purification (TAP) technique (Gavin *et al*., 2002) uses a cloning cassette consisting of protein A, followed by a restriction site for the tobacco etch virus (TEV) protease and calmodulin binding protein. This TAP cassette is inserted at the

C-terminus or alternatively at the N-terminus of a given yeast open reading frame (ORF) by homologous recombination to generate the TAP-tagged fusion protein. The TAP technique combines a first high affinity immunoprecipitation step on protein A agarose beads, mild elution of co-puryfying proteins from protein A beads together with the TAP-tagged fusion protein under investigation and a second affinity purification with calmodulin agarose beads to separate the protein complexes from accompanying proteins binding unspecifically to protein A agarose (Rigaut *et al*., 1999; Puig *et al*., 2001). In contrast to the TAP technique the second large-scale approach, high-throughput mass-spectrometric protein complex identification (HMS-PCI), uses a one-step immuno-affinity purification based on the Flag epitope tag N-terminally fused to a yeast ORF (Ho *et al*., 2002).

Such *in vivo* pulldown experiments do not provide information about binary protein-protein-interactions, but rather describe the assembly of multiple proteins that are stably associated with the tagged bait protein. Gavin *et al*. tested 589 bait proteins with the TAP technique and found 460 of the baits (78 %) complexed with other proteins. The size of protein complexes varied from 2 to 83 constituents, with an average of 12 proteins per complex. Ho *et al*. investigated 725 bait proteins as entry points for the HMS-PCI technique and found 493 of them (68 %) associated in protein complexes. Overall, roughly 1,500 individual proteins were detected in assemblies, which represented 25 % of all yeast proteins. A total of 531 out of 1,500 proteins correspond to hypothetical, uncharacterized proteins predicted from the yeast genome sequence.

Although *in vivo* pulldown experiments are particularly efficient to detect a large number of co-purifying proteins in yeast cells, they have several drawbacks. For *in vivo* pulldowns crude cell extracts were prepared from yeast cells growing in the mid-log phase. A previous study demonstrated that only 25 % of all yeast proteins are expressed in exponentially growing yeast cells (Washburn *et al*., 2001). Therefore, protein complexes which occur in physiological states other than exponential growth are likely to be simply not present or at least under-represented in this approach. Irrespective of the physiological settings, weak protein-protein-interactions may not survive the purification process and may be washed off during purification. Finally, it is questionable whether the success of *in vivo* pulldown techniques in yeast will extend to higher eukaryotes. Haploid yeast cells were used to test the TAP-tagged genes in the absence of their wild-type allele. *In vivo* pulldowns in yeast are particularly successful because the TAP-tagged gene

is inserted into a proper location in the genome under the normal transcriptional control. Higher eukaryotes lack an efficient system for homologous recombination and therefore making *in vivo* pulldowns for higher eukaryotes on a genome-wide scale not feasible.

1.1.3 Data integration and critical evaluation

The various yeast two-hybrid experiments and *in vivo* pulldown assays have confirmed a number of already known and unraveled many new protein-protein-interactions. All approaches have high rates of false positives. Furthermore, of all interactions detected, only 3 % are found by more than one approach (von Mehring *et al*., 2002). To interprete the results of any single large-scale *in vivo* screen for protein-protein-interaction correctly, a critical evaluation of its accuracy, complementarity and overlap with different approaches is crucial. By integrating and comparing the results obtained with different experimental methods, both an increased confidence and a better coverage of protein-protein-interactions studied on a genome-wide scale can be achieved (Gerstein *et al*., 2002). When protein-protein-interactions are found using completely different approaches, their reliability is increased significantly. This has created the demand for the development of new and independent screens to elucidate protein-protein-interactions on a broad level.

1.2 Protein-protein-interaction screenings *in vitro*

1.2.1 Protein arrays

Arrays of positionally adressable biomolecules were first introduced on the DNA level. DNA chips are now widely being used for genetic analyses. The success of miniaturized formats for immobilization of DNA molecules has motivated the development of protein arrays which aim to provide an assay platform to rapidly assess the function of thousands of proteins in parallel. All protein arrays rely on recombinantly expressed proteins studied *in vitro* and therefore do not directly assess protein function in physiological settings. In contrast to *in vivo* methods like the yeast two-hybrid system, such array formats provide a well-controlled *in vitro* environment which allows for screening of protein function under a wide variety of different conditions on a genome-wide scale. To accomplish this goal, it is necessary to immobilize proteins on a solid

support in a way that preserves their folded conformation. To date, several methods for the generation and immobilization of recombinantly expressed proteins on a solid support do exist.

1.2.1.1 Protein microarrays

Protein microarrays can be classified into systems for protein detection and systems for analysis of protein function (reviewed in Kodadek, 2001; MacBeath, 2002). Microarrays for protein detection frequently consist of antibodies directed against a defined set of proteins from a cell, tissue or organism (Haab and Brown, 2001; De Wildt *et al*., 2002). Each antibody captures its antigen from a complex mixture such as serum or cell lysate and captured antigens are typically detected with techniques earlier developed for enzyme-linked immunosorbent assays (ELISAs) (Crowther, 2001). In contrast to antibody arrays, for protein function analysis large sets of proteins are arrayed onto solid supports with high spatial density to investigate enzyme-substrate reactions or binding events of labeled biomolecules such as DNA, proteins, carbohydrates or lipids to particular proteins on the array.

In a proof-of-concept experiment for the analysis of protein-protein-interactions using protein microarrays three commercially available proteins were arrayed on microscopic glass slides and their functions were tested based on their ability to bind their respective naturally known interaction partners (MacBeath *et al*., 2000). Glass slides were pretreated with an aldehyde-containing silane reagent and the proteins were covalently attached onto the glass surface through a reaction of their primary amines with the aldehyde groups. Typically, besides the primary amine at their NH_2-termini, proteins display many lysines with primary amines on their surface, so that they can attach to the slide in a variety of orientations. To generate the microarray, Mac Beath *et al*. used a high-precision contact printing robot which produced 1600 spots per square centimeter with a spot diameter of approximately 200 μm each. Unreacted aldehydes on the slides were quenched with bovine serum albumine (BSA) which by forming a molecular layer additionally reduced nonspecific binding of other proteins in subsequent incubation steps. The microarray was probed with different fluorescently labeled proteins in solution which specifically detect their interaction partners immobilized on the microarray.

The first large-scale protein-protein-interaction screen was performed using the proteome of *S. cerevisiae* (Zhu *et al*., 2001). A library of approximately 5,800 individual

yeast ORF´s were fused to a His-GST tag and cloned in yeast by homologous recombination (Hudson *et al*., 1997). The fusion proteins were overexpressed in yeast, purified, and spotted in duplicates at high density on nickel-(II)-coated slides. Through its affinity to nickel, the His-tag ensured a site-specific immobilization in which the fusion proteins oriented uniformly away from the slide surface. A proof-of-principle experiment demonstrated that this protein microarray is suitable for protein-protein-interaction screening. By incubating the protein array first with biotinylated calmodulin in the presence of calcium and then with fluorescently labeled streptavidin, Zhu *et al*. detected six out of twelve known and an additional 33 novel calmodulin-binding proteins. This approach was the first reported to comprehensively analyse protein binding using ordered protein microarrays with a nearly complete set of proteins from the yeast.

Other approaches generated protein microarrays with predefined subsets of proteins (Ge, 2000) or protein domains (Espejo *et al*., 2002) for the examination of protein-protein-interactions, but they were all limited to several hundred individual proteins per array. This is because the major bottleneck in generating a protein array representing an entire protein collection of a particular organism has been the high-throughput expression and purification of all the proteins. As the set of proteins nominally encoded by the human genome is at least five-fold greater than the total number of yeast proteins, it is an open question how to obtain all the different proteins in sufficient quantity and purity for spotting them on solid supports while keeping their native biological activity (Mitchell, 2002).

Moreover, protein microarray approaches have four broad limitations. First, the unspecific adsorption of a soluble protein probe can compete with selective protein-protein-interactions, leading to increased background signals and thereby particularly complicating the investigation of low affinity binding events. Second, proteins are either immobilized in nonproductive orientations or are denatured to some extent, both of which prevents selective interactions with soluble protein probes. The third limitation can be attributed to the labeling technique used to detect binding events on protein microarrays. The frequently applied fluorescence labeling technique is problematic as chemical modifications caused by labeling may result in changing the protein surface characteristics, thus preventing the probe proteins from interaction with the immobilized proteins on the microarray.

1.2.1.2 Arrayed cDNA expression libraries

An alternative way to immobilize proteins in a high spatial density on solid supports is a technique which uses arrayed cDNA expression libraries and was described by Büssow *et al*. (Büssow *et al*., 1998, Büssow *et al*., 2000). In this approach, a cDNA library was constructed by reverse transcription of the poly(A)$^+$ mRNA population from a human fetal brain. The cDNA fragments were directionally cloned into a vector for expression of fusion proteins with an N-terminal His-tag and transformed into the bacterial host *E. coli*. After transformation, bacterial expression clones were picked into microtiter plates with a picking robot. The cDNA expression library that was obtained has been used to generate high-density protein arrays on polyvinyl difluoride (PVDF) membranes.

Individual clones of the cDNA library were spotted on the PVDF membrane, grown on agar plates and induced for recombinant protein expression. Bacterial cells were lysed and cellular proteins were released together with the recombinant protein leading to an immobilization of the proteins on the membrane. PVDF membranes have a high protein binding capacity and the released proteins are bound with a high local protein concentration per spot in a non-covalent but irreversible manner. Once the expression clones have been picked into microtiter plates, they can be used multiple times as starting material for the production of protein arrays without the need for time-consuming and laborious purification of individual proteins.

Usually, cDNA expression libraries contain many clones that do not produce a recombinant protein. Productive clones can be detected with an antibody directed against the His-tag of the recombinant protein, whereas most unproductive clones are not detected. Such unproductive clones were removed from the library by rearraying the subset of productive clones which represented approximately 20 % of the initial cDNA expression library.

The arrayed cDNA expression library was assayed for antigen-antibody binding in a first proof-of-concept experiment. A total of 80,640 individual clones were probed with an antibody directed against the His-tag, a cDNA probe directed against human HSP90α and an antibody directed against human HSP90α. 56 clones were detected with the cDNA probe. 25 % of the clones tested positive for the cDNA probe were also positive for the anti-His-tag antibody, whereas 72 % of the clones tested positive for the cDNA probe were also detected with the protein-specific anti-HSP90α antibody. DNA

sequencing turned out that the fraction of clones negative for the anti-His-tag antibody contained reading frames not correctly fused to the His-tag but nevertheless expressed a protein detected by the anti-HSP90α antibody. This result suggests that such clones contain a translational start site within the cDNA insert enabeling the expression of HSP90α fragments which lack the His-tag.

Even though cDNA expression libraries contain many clones expressing proteins which begin and end at random positions, they provide an economical way to produce a large set of different proteins without the need of laborious PCR amplification using individual primer pairs for each gene. Certainly, a number of proteins will prove refractory to expression in *E. coli* and subsequent biochemical manipulation, but cDNA expression libraries have the advantage to study a comprehensive set of proteins, especially from higher eukaryotes both simultaneously and repeatedly. Furthermore, a prescreening step for detection of unproductive clones followed by rearraying of the productive ones undoubtedly improves the quality of cDNA expression libraries.

1.2.2 Peptide arrays

Peptide arrays can be particularly useful to investigate protein-protein-interactions in those cases in which one of the interacting partners participates in complex formation by docking a relatively short amino acid sequence into a binding pocket of its receptor protein. In fact, many protein-protein-interactions are mediated by protein-binding domains, such as PDZ, WW, SH2 or SH3 domains, that accommodate binding of relatively short peptide sequences in their respective binding pockets (Pawson and Scott, 1997; Pawson *et al*., 2002; Pawson and Nash, 2003).

The generation of peptide arrays usually relies on the SPOT synthesis, the chemical synthesis of peptides on cellulose membranes which was originally developed to enable the mapping of linear antibody epitopes (Frank, 1992). The key amino acid residues of linear binding sites which mediate the contacts to the antibody are located within one stretch of the primary structure, usually not exceeding 15 amino acids in length. The entire sequence of a protein antigen is synthesized as overlapping 8- to 15-mer peptides covalently bound to a solid phase such as cellulose membranes. These solid phase-bound peptides are subsequently tested for binding of an antibody directly on the membrane (reviewed in Reineke *et al*., 2001). In a typical peptide scanning experiment the peptide array is first overlayed with an antibody or a protein probe under

investigation. Detection of binding between a peptide and the protein probe is then mediated by a labeled antibody directed against the probe protein.

A frequent application of peptide arrays is the identification of linear peptide epitopes that bind to a certain protein domain (Weiergräber *et al*., 1996; Reineke *et al*., 1998; Ball *et al*., 2000; Rüdiger *et al*., 2001; Töpert *et al*., 2001). Once a protein-protein-interaction has been discovered, such screening can be knowledge-based and use overlapping peptides delineated from the appropriate protein sequence to map the critical region for association with an interaction partner.

1.3 Other *in vitro* techniques with recombinant proteins

There are many systems available for the production of recombinant proteins. The most commonly used protein fusion systems include: polyhistidine-tags (Hochuli, *et al*., 1987), maltose binding protein fusions (Bedouelle and Duplay, 1988; Guan *et al*., 1988, Maina *et al*., 1988), Glutathione-S-transferase fusion proteins (Smith and Johnson, 1988) as well as tags for defined antigenic regions such as the Myc epitope or the flag-tag. Glutathione-S-transferase (GST) fusion proteins and His-tagged proteins have had the most widespread range of applications since their introduction as tools for the synthesis of recombinant proteins in *E. coli*. The development of vectors and antibodies that support applications with GST fusion proteins as well as His-tagged proteins has made them a popular choice to study protein-protein-interactions *in vitro*.

1.3.1 Far Western overlay assays

This technique is similar to Western blot techniques developed for antibody detection of proteins immobilized on membranes by gel blotting or spotting. In a Far Western overlay, the antibody is replaced by a recombinant protein and the interaction of this protein with a protein on the membrane is assayed by a subsequent overlay with an antibody that detects the recombinant protein probe. Through this antibody, bands or spots of proteins which the probe protein has bound can be visualized on the membrane, thereby identifying an interaction with an immobilized protein.

In a proof-of-concept experiment it was shown that an antibody can be used as a probe to screen a λgt11 cDNA expression library to identify a gene encoding a protein antigen (Young and Davis, 1983). The library typically uses an IPTG-inducible promoter

to express proteins fused to ß-galactosidase in *E. coli*. The proteins to be probed are prepared by plating the λgt11 cDNA expression library on agar dishes. After induction of expression, proteins from the bacteriophage plaques are transferred to nitrocellulose filters, incubated with antibody, and washed to remove nonspecifically bound antibody (Young and Davis, 1991).

Macgregor *et al*. (1990) used the leucine zipper and DNA-binding domain of the proto-oncogene Jun as a biotinylated probe to screen a λgt11 cDNA expression library. An interaction with the cAMP response element-binding protein type 1 (CRE-BP1) was detected with a streptavidin-alkaline phosphatase conjugate. Blackwood and Eisenman (1991) used a direct labeling technique to identify a c-Myc interacting protein. A C-terminal fragment of c-Myc containing the basic-region helix-loop-helix leucine zipper (bHLH-zip) domain was expressed as GST fusion protein and radioactively labeled with ^{125}I. This probe detected a new bHLH-zip protein termed Max. Ayer *et al.* (1993) extended the results of Blackwood and Eisenman by using Max as a labeled probe protein to identify another member of bHLH-zip domain containing proteins.

Blanar and Rutter (1992) described a λgt11 cDNA expression library screen with a GST fusion protein probe that contains a recognition site for cAMP-dependent protein kinase A. This allowed a site-specific phosphorylation with ^{32}P. Radioactively labeling the probe protein is rapid and generally has little impact on the subsequent activity of the protein because the kinase target site for phosphorylation is in the fusion portion of the protein.

1.3.2 *in vitro* pulldown assays

The GST pull down, as originally described (Kaelin *et al*., 1991), is an affinity purification of an unknown protein from a complex mixture of proteins in a soluble cell extract. The examined GST fusion protein is coupled to glutathione agarose beads and used to isolate proteins which compete successfully with all other proteins in the mixture for binding. Sedimentation of the beads through centrifugation is used to collect the GST fusion protein and any associated protein. The complexes are washed to remove nonspecifically adhering proteins. Excess free glutathione is used to elute the complexes from the beads, or they are boiled directly in SDS loading buffer. Constituents of a protein complex are then resolved by SDS-PAGE and processed for Coomassie staining or Western blotting.

Such *in vitro* pulldown experiments can be adapted to detect binary protein-protein-interactions between two recombinantly expressed proteins bearing different fusion tags. The first fusion tag is used for immobilization to the agarose beads. A protein-protein-interaction can then be assayed by incubating the immobilized protein with a second protein tagged differentially. An interacting protein can then be detected by means of a Western blot using an antibody directed against the second tag.

A number of important points have to be considered for a pulldown experiment. Being able to synthesize the fusion protein without excess degradation and insolubility is the most important consideration. Because different preparations of fusion protein may result in variable degradation it is important to monitor the status of the fusion protein after purification. Second, it is important to control and minimize aggregation, because aggregated proteins are frequently pulled down together with the bead-bound proteins unspecifically. Thus, it is difficult to distinguish specific from unspecific binding.

1.4 Objective

Most of the genome-wide protein-protein-interaction data published so far has been produced with *in vivo* pulldowns or with the yeast two-hybrid system. Both techniques are known to generate false positives and require independent confirmation of potential interaction partners. Furthermore, both techniques have their limitations and fail to discover particular protein-protein-interactions. An integration of data with different experimental approaches will be essential to obtain a reliable understanding of protein-protein-interactions on a genome-wide scale. Protein arrays are a promising tool to complement existing large-scale techniques. High-density protein arrays from cDNA expression libraries produce thousands of human proteins without the need to clone and express individual proteins. To date, only a few studies based on high-density protein array screens have been undertaken to identify antigens (Büssow *et al*., 1998, Holt *et al*., 2000), binding proteins (Mahlknecht *et al*., 2001), substrates for methyltransferases (Lee and Bedford, 2002) or kinase substrate proteins (de Graf *et al*., 2004). As of now, little is known about the use of high-density protein arrays in search for novel protein-protein-interactions. A screening method has to be developed which allows for a sensitive detection of binding partners on protein arrays. The screening results have to be confirmed or disproved with an independent technique, such as an *in vitro* pulldown assay, to obtain reliable and quick yes or no answers for the examined interactions. The src homology 3 (SH3) domain of endophilin-1, a protein that regulates synaptic vesicle formation in the cell and participates in a number of intracellular signaling pathways (Reutens and Begley, 2002), was chosen to evaluate the developed screening method based on high-density protein arrays. Endophilin-1 is a multifunctional protein that has different functions depending on its interaction partners and its location within the cell. A number of interaction partners have been described previously for the SH3 domain of endophilin-1. Furthermore, SH3 domains recognize unstructured peptides in its binding partners and the SH3 recognition mode has been studied extensively. These are good prerequisites to study SH3 domain-dependent binding of endophilin-1 on a genome-wide scale.

2 Materials

2.1 Bacterial strains

E. coli XL-I Blue
(Stratagene)
Genotype: *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac*[F'*pro*AB l*acI*^{q}ZΔ*M15* Tn*10*(tet^r)]

E. coli SCS1
(Stratagene)
Genotype: *hsd*R17(r_K^- m_K^+) *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1 *sup*E44

E. coli DH10B
(Invitrogen)
Genotype: F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ*80*l*ac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(ara, leu)7697 *gal*U *gal*K λ$^-$ *rps*L *nup*G *ton*A

2.2 Plasmids

Table 1: *E. coli* cloning vectors

vector name	description	resistance	source
Lafmid BA	promotor: P(lac), host: *E. coli*	ampicillin	RZPD
pBluescriptR	promotor: T3, T7, host: *E. coli*	ampicillin	Stratagene
PBluescriptSK+	promotor: T3, T7, host: *E. coli*	ampicillin	Stratagene
pBTM117c	promotor: ADH1, lexA, host: *E. coli*, yeast	ampicillin	AG Wanker
pCMV-SPORT6	promotor: SP6, T7, host: E. *coli*, mammalian	ampicillin	Invitrogen
pOTB7	promotor: SP6, T7 host: E. *coli*	chloramph.	RZPD
pT7T3D-PacI	promotor: T3, T7 (mod), host: E. *coli*	ampicillin	RZPD

Table 2: *E. coli* expression vectors

vector name	description	resistance	source
pGEX-6P-1	production of GST fusion proteins, tac promotor, lacI^q repressor, BamHI/NotI digest	ampicillin	Amersham
pGEX-6P-2	production of GST fusion proteins, tac promotor, lacI^q repressor, SalI/NotI digest	ampicillin	Amersham
pGEX-2T-K	production of GST fusion proteins, tac promotor, lacI^q repressor, cAMP kinase site	ampicillin	Amersham
pQE30NST	production of N-terminal His-tagged proteins, T5 promotor, lac operator	ampicillin	(Büssow et al., 1998)
pQE31N	production of N-terminal His-tagged proteins, T5 promotor, lac operator, BamHI/NotI digest	ampicillin	QIAGEN
pQE32N	production of N-terminal His-tagged proteins, T5 promotor, lac operator, SalI/NotI digest	ampicillin	QIAGEN
pSE111	helper plasmid, T7 promotor, lacI^q repressor, argU gene	kanamycin	E. Scherzinger (unpublished)

2.3 Template clones

Table 3: Template clones for PCR amplification of DNA fragments

clone ID	description of insert	cloning vector
DKFZp586H1320	alpha-adaptin	pSPORT1
IMAGp958E125	syntaxin 1A (brain)	pOTB7
IMAGp998A0810138	syntaxin 12	pCMV-SPORT6
IMAGp998A101942	syntaxin 5A	pT7T3D-PacI
IMAGp998A12676	vamp 5 (myobrevin)	pT7T3D-PacI
IMAGp998A2110040	ARP-1 actin-related protein 1 (centractin)	pBluescriptR
IMAGp998B036270	SH3P7 (SH3 domain-containing protein HIP-55)	pBluescript SK+
IMAGp998B119638	trafficking protein particle complex 4 (synbindin)	pCMV-SPORT6
IMAGp998B2410085	p41-Arc (actin related protein 2/3 complex, subunit 1B)	pCMV-SPORT6
IMAGp998C136535	tyrosine kinase p59fyn	pT7T3D-PacI
IMAGp998D2310100	cortactin (Oncogene EMS1)	pCMV-SPORT6
IMAGp998E117174	tyrosine protein-kinase LYN	pT7T3D-PacI
IMAGp998E169703	calmodulin 2	pCMV-SPORT6
IMAGp998E20139	endophilin 1	Lafmid BA
IMAGp998F029996	filamin A (actin binding protein 280)	pCMV-SPORT6
IMAGp998F079995	ubiquitin-conjugating enzyme UbcH7	pCMV-SPORT6
IMAGp998G019562	Chip (STIP1 and U-box containing protein 1)	pCMV-SPORT6
IMAGp998H0910243	dynamitin	pCMV-SPORT6
IMAGp998I112838	syntaxin 8	pT7T3D-PacI
IMAGp998I166882	amphiphysin 1	pBluescript SK+
IMAGp998K016273	synaptosomal-associated protein SNAP25	pBluescript SK+
IMAGp998K0410067	auxilin (DnaJ homolog)	pCMV-SPORT6
IMAGp998K089740	cathepsin B	pCMV-SPORT6
IMAGp998K226887	synaptotagmin 1	pBluescript SK+
IMAGp998L0510082	Tim44 (translocase of inner mitochondrial membrane)	pCMV-SPORT6
IMAGp998L086887	phocein	pBluescript SK+
IMAGp998L1010008	ubiquitin-conjugating enzyme UbcH8	pCMV-SPORT6
IMAGp998L1110101	Hsc70 (heat shock cognate 70 kDa protein 8)	pCMV-SPORT6
IMAGp998M209732	p20-Arc (actin related protein 2/ complex, subunit 4)	pCMV-SPORT6
IMAGp998N195197	tyrosin-protein kinase src-1 (c-src)	pT7T3D-PacI
IMAGp998P099558	amphiphysin 2	pCMV-SPORT6
IMAGp998P159641	heat shock 60 kDa protein 1 (p60 lymphocyte protein)	pCMV-SPORT6
IMAGp998P246272	synapsin 1A (brain protein 4.1)	pBlueskript SK+
IRAKp961A1513	syntenin (syndecan binding protein)	pCMV-SPORT6
IRAKp961E0321	syntaxin 7	pCMV-SPORT6
IRAKp961G0813	Rab4	pCMV-SPORT6
IRAKp961K0931	Psd95	pCMV-SPORT6
IRALp962G089	Hop (Hsp70/Hsp90 organizing protein)	pOTB7
pBTM117c-parkin wt	parkin	pBTM117c

2.4 Oligonucleotides

Table 4: PCR primer (forward)
underlined: restriction site, bold/underlined: first matching codon

primer name	primer sequence	restriction site
adap F2	5´- GACACTCGAGC **CCG** GCC GTG TCC AAG GGC -3´	XhoI
adap F319	5´- CACAGTCGACC **CAC** TAT GAC AGT GAG CCC AAC -3´	SalI
amphi1 F291	5´- GACAGTCGACC **CCT** CGG TCA CCT TCA CAG A -3´	SalI
amphi1 F560	5´- GAGAGTCGACC **ACT** ATA GGT GCA GAG CCC AAG -3´	SalI
amphi2 F497	5´- CAGAGTCGACC **ACC** TTC CCA GCA ACT GTG A -3´	SalI
arp1 F2	5´- GAGAGTCGACC **GAG** TCC TAC GAC ATC ATC GCC -3´	SalI
aux F550	5´- GACAGTCGACC **CCT** AGT GGA CCT GCG TCT -3´	SalI
aux F816	5´- GAGAGTCGACC **GAT** CCT GAG AAA TTA AAG ATT CTG GAA TGG -3´	SalI
c-src F83	5´- CAGAGTCGACC **GCC** GGT GGA GTG ACC ACC -3´	SalI
calmod F2	5´- CAGGATCC **GCT** GAC CAA CTG ACT GAA GAG -3´	BamHI
cathep F2	5´- GAGAGTCGACC **TGG** CAG CTC TGG GCC TCC -3´	SalI
chip F143	5´- GAGAGTCGACC **AAG** AAG AAG CGC TGG AAC AG -3´	SalI
chip F198	5´- CAGAGTCGACC **GCC** TGC ATT GAG GCC AAG -3´	SalI
chip F2	5´- GAGAGTCGACC **AAG** GGC AAG GAG GAG AAG G -3´	SalI
cort F2	5´- GAGAGTCGACC **TGG** AAA GCT TCA GCA GGC -3´	SalI
cort F350	5´- CTGAGTCGACC **ATC** AGA GCT AAC TTT GAA AAC CTC G -3´	SalI
cort F85	5´- GAGAGTCGACC **GGA** GGG AAA TTT GGT GTG GAA -3´	SalI
dynmt F2	5´- GAGAGTCGACC **GCG** GAC CCT AAA TAC GCC -3´	SalI
endoph F2	5´- GAGACTCGAGC **TCG** GTG GCC GGC CTC AAG -3´	XhoI
endoph F292	5´- GACAGTCGACC **CAG** CCC TGC TGC CGA GCT -3´	SalI
fila F1644	5´- GAGGATCC **AAG** TGC ACT GTC ACA GTG TC -3´	BamHI
fila F1788	5´- GAGGATCC **CTG** AGG CCC TTT GAC CTT G -3´	BamHI
fila F2015	5´- GAGGATCC **GAG** ACG GGG GAG CAC CTG -3´	BamHI
fila F2041	5´- GAGGATCC **CAG** TCG GAA ATT GGG GAT GC -3´	BamHI
fyn SH3 F2	5´- GAGGATCC **GGC** TGT GTG CAA TGT AAG GA -3´	BamHI
fyn SH3 F82	5´- CAGGATCC **ACA** GGA GTG ACA CTC TTT GTG G -3´	BamHI
hop F2	5´- GAGACTCGAGC **GAG** CAG GTC AAT GAG CTG AAG -3´	XhoI
hsc70 F2	5´- CAGAGTCGACC **TCC** AAG GGA CCT GCA GTT -3´	SalI
hsc70 F373	5´- GAGAGTCGACC **GCA** GCT GTC CAG GCA GCC -3´	SalI
hsc70 F541	5´- GAGAGTCGACC **TCA** CTT GAG TCC TAT GCC TTC A -3´	SalI
lyn SH3 F2	5´- GAGGATCC **GGA** TGT ATA AAA TCA AAA GGG AAA GAC AGC -3´	BamHI
lyn SH3 F62	5´- GAGGATCC **GAG** GAA CAA GGA GAC ATT GTG G -3´	BamHI
p20-arc F2	5´- GAGTGTCGACC **ACT** GCC ACT CTC CGC CCC -3´	SalI
p41-arc F2	5´- CAGAGTCGACC **GCC** TAC CAC AGC TTC CTG G -3´	SalI
p60 F2	5´- GTCAGTCGACC **CTT** CGG TTA CCC ACA GTC TTT C -3´	SalI
park F2	5´- GTGAGTCGACC **ATA** GTG TTT GTC AGG TTC AAC TCC A -3´	SalI
park F220	5´- GAGAGTCGACC **AAG** GAA ACA CCA GTA GCT TTG C -3´	SalI
park F304	5´- CTGAGTCGACC **TTC** AGG ATT CTG GGA GAA GAG C -3´	SalI
park F32	5´- GAGAGTCGACC **AAG** CGA CAG GGG GTT CCG -3´	SalI
park F395	5´- GAGAGTCGACC **GAA** AGA GCC GCC GAG CAG -3´	SalI
phocein F4	5´- GAGGATCC **GAG** GGG ACG GCA GTG CTG -3´	BamHI
psd95 F108	5´- GAGAGTCGACC **GAA** ATC ACA TTG GAA AGG GGT AAC T -3´	SalI
rab4 F2	5´- CAGAGTCGACC **TCC** GAA ACC TAC GAT TTT TTG TTT AAG T -3´	SalI
sbd F2	5´- GAGAGTCGACC **GCG** ATT TTT AGT GTG TAT GTG GTG A -3´	SalI

Table 4: PCR primer (forward), continued
underlined: restriction site, bold/underlined: first matching codon

primer name	primer sequence	restriction site
sh3p7 F2	5´- GAGAGTCGACC **GCG** GCG AAC CTG AGC CGG -3´	SalI
sh3p7 F282	5´- GAGAGTCGACC **AGG** AGC CCC TTC CTG CAG -3´	SalI
snap25 F2	5´- CAGAGTCGACC **GCC** GAA GAC GCA GAC ATG -3´	SalI
stn F2	5´- GAGAGTCGACC **TCT** CTC TAT CCA TCT CTC GAA GAC T -3´	SalI
stx1A F2	5´- GAGAGTCGACC **AAG** GAC CGA ACC CAG GAG -3´	SalI
synt5 F2	5´- CAGACTCGAGC **TCC** TGC CGG GAT CGG ACC -3´	XhoI
synt7 F2	5´- GAGAGTCGACC **TCT** TAC ACT CCA GGA GTT GGT G -3´	SalI
synt8 F2	5´- GAGAGTCGACC **GCA** CCG GAC CCC TGG TTC -3´	SalI
syps1A F113	5´- CAGAGTCGACC **TCC** AGG GTG CTG CTG GTC -3´	SalI
sytag F135	5´- GAGAGTCGACC **GAA** GAA CCC AAA GAA GAG GAG AAA C -3´	SalI
sytag F265	5´- GAGAGTCGACC **AGT** GCT GAG AAG GAA GAG CA -3´	SalI
tim44 F39	5´- CACACTCGAGC **CGC** CGG CCG GGC GGA GAG -3´	XhoI
ubch7 F2	5´- GAGAGTCGACC **GCG** GCC AGC AGG AGG CTG -3´	SalI
ubch8 F2	5´- GAGAGTCGACC **GCG** AGC ATG CGA GTG GTG -3´	SalI
vamp5 F2	5´- GAGAGTCGACC **GAG** GAA GCC AGT GAA GGT GG -3´	SalI

Table 5: PCR primer (reverse)
underlined: restriction site, bold/underlined: last matching codon

primer name	primer sequence	restriction site
adap R478	5´- GACTGCGGCCGC TCA **GCC** CTG GAC GTC ATC ACG -3´	NotI
amphi1 R559	5´- GACTGCGGCCGC TCA **TAT** TTC GTT TTC TCC TTC CTC TTC ATG -3´	NotI
amphi1 R695	5´- GAGAGCGGCCGC CTA **ATC** TAA GCG TCG GGT GAA GT -3´	NotI
amphi2 R593	5´- GACTGCGGCCGC TCA **TGG** GAC CCT CTC AGT G -3´	NotI
arp1 R376	5´- GAGAGCGGCCGC CTA **GAA** AGT TTT GCG ATG AAT AGC ACG -3´	NotI
aux R913	5´- GAGTGCGGCCGC TTA **ATA** TAA GGG CTT TTG GCC TTG G -3´	NotI
c-src R144	5´- GACTGCGGCCGC TCA **GTC** GGA GGG CGC CAC GTA -3´	NotI
c-src R218	5´- GACTGCGGCCGC TCA **GGT** GAT GTA GAA GCC GCC -3´	NotI
calmod R149	5´- CACTGCGGCCGC TCA **CTT** TGC TGT CAT CAT TTG TAC AA -3´	NotI
cathep R339	5´- GAGTGCGGCCGC TTA **GAT** CTT TTC CCA GTA CTG ATC GG -3´	NotI
chip R142	5´- CACTGCGGCCGC TCA **CTG** CTG GGC CCG GAC GTG -3´	NotI
chip R197	5´- CACTGCGGCCGC TCA **CGC** GAT TCG AAG AGC GCT -3´	NotI
chip R303	5´- GACTGCGGCCGC TCA **GTA** GTC CTC CAC CCA GC -3´	NotI
cort R330	5´- CACTGCGGCCGC TCA **CAC** CTG GGT GAC ATC CTC A -3´	NotI
cort R550	5´- CAGAGCGGCCGC CTA **CTG** CCG CAG CTC CAC -3´	NotI
dynmt R406	5´- CACTGCGGCCGC TCA **CTT** TCC CAG CTT CTT CAT CC -3´	NotI
dynmt R87	5´- GACTGCGGCCGC TCA **TTC** ATA TCC TGT CCT CTT GGT TTT TCC -3´	NotI
endoph R352	5´- GAGAGCGGCCGC CTA **ATG** GGG CAG GGC AAC -3´	NotI
fila R2136	5´- GACTGCGGCCGC TCA **GCC** TGT CAC CTT CAC AGA GA -3´	NotI
fila R2198	5´- GACTGCGGCCGC TCA **GTA** GGT GTG GTT CTC CCC T -3´	NotI
fyn SH3 R142	5´- GACGGCGGCCGC TCA **GTC** AAC TGG AGC CAC ATA ATT GC -3´	NotI
hop R543	5´- CACTGCGGCCGC TCA **CCG** AAT TGC AAT CAG ACC -3´	NotI
hsc70 R540	5´- GACTGCGGCCGC TCA **ATT** CTT GGA TGA CAC CTT GTC CC -3´	NotI
hsc70 R646	5´- GAGTGCGGCCGC TTA **ATC** AAC CTC TTC AAT GGT GGG C -3´	NotI
lyn SH3 R121	5´- GACTGCGGCCGC TCA **GAG** TTT GGC CAC ATA GTT GCT -3´	NotI

Table 5: PCR primer (reverse), continued
underlined: restriction site, bold/underlined: last matching codon

primer name	primer sequence	restriction site
lyn SH3 R141	5´- CACTGCGGCCGC TCA **CCT** TTC TGC GTC CTT CCT G -3´	NotI
p20-arc R168	5´- GAGTGCGGCCGC TTA **AAA** ATT CTT AAG GAA CTC TTC AGC CAC -3´	NotI
p41-arc R372	5´- GACTGCGGCCGC TCA **TTT** GAT CTT GAG GTC CTT CAA GG -3´	NotI
p60 R573	5´- GAGTGCGGCCGC TTA **GAA** CAT GCC ACC TCC CA -3´	NotI
park R103	5´- GACTGCGGCCGC TCA **AGT** CAA GCT CTG GGG CTC -3´	NotI
park R318	5´- GACTGCGGCCGC TCA **ATA** CTG CTG GTA CCG GTT GT -3´	NotI
park R404	5- GACTGCGGCCGC TCA **TTC** CCA ACG AGC CTG CTC -3´	NotI
park R465	5´- CAGAGCGGCCGC CTA **CAC** GTC GAA CCA GTG GT -3´	NotI
phocein R216	5´- CACTGCGGCCGC TCA **CTG** TAC TTC CTC TTC TAA AAT TGG TAC AAT -3´	NotI
psd95 R541	5´- CACTGCGGCCGC TCA **CTC** TCG TCG CTC AAC CCG -3´	NotI
rab4 R213	5´- GAGAGCGGCCGC CTA **ACA** ACC ACA CTC CTG AGC -3´	NotI
sbd R219	5´- GAGAGCGGCCGC CTA **TGA** CCC AGG TCC AAA AGT TC -3´	NotI
sh3p7 R281	5´- CACTGCGGCCGC TCA **CAG** CTT GCC AGG CTG AGG –3´	NotI
sh3p7 R430	5´- CACTGCGGCCGC TCA **CTC** AAT GAG CTC CAC GT -3´	NotI
snap25 R206	5´- GAGTGCGGCCGC TTA **ACC** ACT TCC CAG CAT CTT TG -3´	NotI
stn R273	5´- GACTGCGGCCGC TCA **AAA** AGC AGG CAT GAT TGT AAT AGT AAC T -3´	NotI
stn R298	5´- GAGTGCGGCCGC TTA **AAC** CTC AGG AAT GGT GTG GT -3´	NotI
stx1A R265	5´- GACTGCGGCCGC TCA **TTT** CTT CCG GCG CGC CTT -3´	NotI
stx1A R288	5´- GAGAGCGGCCGC CTA **GGC** GAA GAT GCC CCC -3´	NotI
synt12 R269	5´- CACTGCGGCCGC TCA **CTT** CGT TTT ATA AAC TAG CCA GAT AAT -3´	NotI
synt5 R284	5´- GACTGCGGCCGC TCA **TTT** GAC CAT GAG CCA CCG -3´	NotI
synt5 R301	5´- GACTGCGGCCGC TCA **AGC** AAG GAA GAC CAC AAA G -3´	NotI
synt7 R236	5´- GACTGCGGCCGC TCA **TTT** TCT GGA TTT GCG CTG ATA ATC T -3´	NotI
synt7 R261	5´- GACTGCGGCCGC TCA **GTG** GTT CAA TCC CCA TAT GA -3´	NotI
synt8 R236	5´- GACTGCGGCCGC TCA **GTT** GGT CGG CCA GAC -3´	NotI
syps1A R420	5´- CACTGCGGCCGC TCA **CCG** GGG CAG GGC CTG AGC -3´	NotI
sytag R422	5´- CAGTGCGGCCGC TTA **CTT** CTT GAC GGC CAG CA -3´	NotI
tim44 R452	5´- GACTGCGGCCGC TCA **GAG** AAT CTG CTC GGT GC -3´	NotI
ubch7 R154	5´- GAGTGCGGCCGC TTA **GTC** CAC AGG TCG CTT TTC C -3´	NotI
ubch8 R152	5´- GAGTGCGGCCGC TTA **GGA** GGG CCG GTC CAC -3´	NotI
vamp8 R75	5´- CACTGCGGCCGC TCA **CTT** CAC GTT CTT CCA CCA GAA TT -3´	NotI

Table 6: Vector primers

primer name	primer sequence	description
pQE 65	5´- TGA GCG GAT AAC AAT TTC ACA CAG -3´	forward primer for vectors of the pQE series
pQE 276	5´- GGC AAC CGA GCG TTC TGA AC -3´	reverse primer for vectors of the pQE series
pGEX-5´	5´- GGG CTG GCA AGC CAC GTT TGG TG -3´	forward primer for vectors of the pGEX series
pGEX-3´	5´- CCG GGA GCT GCA TGT GTC AGA GG -3´	reverse primer for vectors of the pGEX series
pKA3F	5´- GAA GAT CTC GTC GTG CAT CTG TTG G -3´	primer matching pKA site of pGEX-2T-K

All oligonucleotides were obtained from EUROGENTEC S.A., Belgium.

2.5 PCR fragments

Table 7: PCR fragments created in this work

PCR fragment	forward primer	reverse primer	template clone
adaptin $_{2-478}$	adap F2	adap R478	DKFZp586H1320
adaptin $_{319-478}$	adap F319	adap R478	DKFZp586H1320
amphiphysin 1 $_{291-559}$	amphi1 F291	amphi1 R559	IRAKp961B0450
amphiphysin 1 $_{560-695}$	amphi1 F560	amphi1 R695	IMAGp998I166882
amphiphysin 2 $_{497-593}$	amphi2 F497	amphi2 R593	IMAGp998P099558
centractin $_{2-476}$	arp1 F2	arp1 R376	IMAGp998A2110040
auxilin $_{550-913}$	aux F550	aux R913	IMAGp998K0410067
auxilin $_{816-913}$	aux F816	aux R913	IMAGp998K0410067
calmodulin $_{2-149}$	calmod F2	calmod R149	IMAGp998E169703
cathepsin $_{2-339}$	cathep F2	cathep R339	IMAGp998K089740
CHIP $_{143-303}$	chip F143	chip R303	IMAGp998G019562
CHIP $_{198-303}$	chip F198	chip R303	IMAGp998I079706
CHIP $_{2-142}$	chip F2	chip R142	IMAGp998I079706
CHIP $_{2-197}$	chip F2	chip R197	IMAGp998I079706
CHIP $_{2-303}$	chip F2	chip R303	IMAGp998I079706
cortactin $_{2-330}$	cort F2	cort R330	IMAGp998D2310100
cortactin $_{2-550}$	cort F2	cort R550	IMAGp998D2310100
cortactin $_{350-550}$	cort F350	cort R550	IMAGp998P159751
cortactin $_{85-330}$	cort F85	cort R330	IMAGp998P159751
c-src $_{83-144}$	c-src F83	c-src R144	IMAGp998N195197
c-src $_{83-218}$	c-src F83	c-src R218	IMAGp998N195197
dynamitin $_{2-406}$	dynmt F2	dynmt R406	IMAGp998H0910243
dynamitin $_{2-87}$	dynmt F2	dynmt R87	IMAGp998H0910243
endophilin 1 $_{2-352}$	endoph F2	endoph R352	IMAGp998E20139
endophilin 1 $_{292-352}$	endoph F292	endoph R352	IMAGp998E20139
filamin A $_{1644-2136}$	fila F1644	fila R2136	IMAGp998F029996
filamin A $_{1788-2136}$	fila F1788	fila R2136	IMAGp998F029996
filamin A $_{2015-2198}$	fila F2015	fila R2198	IMAGp998F029996
filamin A $_{2041-2136}$	fila F2041	fila R2136	IMAGp998F029996
Fyn SH3 $_{2-142}$	fyn SH3 F2	fyn SH3 R142	IMAGp998C136535
Fyn SH3 $_{82-142}$	fyn SH3 F82	fyn SH3 R142	IMAGp998C136535
HOP $_{2-543}$	hop F2	hop R543	IRALp962G089
Hsc70 $_{2-540}$	hsc70 F2	hsc70 R540	IRALP962I1026
Hsc70 $_{373-540}$	hsc70 F373	hsc70 R540	IMAGp998L1110101
Hsc70 $_{373-646}$	hsc70 F373	hsc70 R646	IMAGp998L1110101
Hsc70 $_{541-646}$	hsc70 F541	hsc70 R646	IMAGp998O119580
Lyn SH3 $_{2-121}$	lyn SH3 F2	lyn SH3 R121	IMAGp998E117174
Lyn SH3 $_{2-141}$	lyn SH3 F2	lyn SH3 R141	IMAGp998E117174

Table 7: PCR fragments, continued

PCR fragment	forward primer	reverse primer	template clone
Lyn SH3 $_{62-121}$	lyn SH3 F62	lyn SH3 R121	IMAGp998E117174
myobrevin $_{2-75}$	vamp5 F2	vamp5 R75	IMAGp998A12676
p20-arc $_{2-168}$	p20-arc F2	p20-arc R168	IMAGp998M209732
p41-arc $_{2-372}$	p41-arc F2	p41-arc R372	IMAGp998B2410085
p60 $_{2-573}$	p60 F2	p60 R573	IMAGp998P159641
parkin $_{2-103}$	park F2	park R103	pBTM117c-parkin wt
parkin $_{220-465}$	park F220	park R465	pBTM117c-parkin wt
parkin $_{2-318}$	park F2	park R318	pBTM117c-parkin wt
parkin $_{304-404}$	park F304	park R404	pBTM117c-parkin wt
parkin $_{32-465}$	park F32	park R465	pBTM117c-parkin wt
parkin $_{395-465}$	park F395	park R465	pBTM117c-parkin wt
phocein $_{4-216}$	phocein F4	phocein R216	IMAGp998L086887
Psd95 $_{108-541}$	psd95 F108	psd95 R541	IRAKp961K0931
Rab4 $_{2-213}$	rab4 F2	rab4 R213	IRAKp961G0813
SH3p7 $_{2-281}$	sh3p7 F2	sh3p7 R281	IMAGp998B036270
SH3p7 $_{2-430}$	sh3p7 F2	sh3p7 R430	IMAGp998B036270
SH3p7 $_{282-430}$	sh3p7 F282	sh3p7 R430	IMAGp998B036270
SNAP25 $_{2-206}$	snap25 F2	snap25 R206	IMAGp998K016273
synapsin1A $_{113-420}$	syps1A F113	syps1A R420	IMAGp998P246272
synaptotagmin $_{135-422}$	sytag F135	sytag R422	IMAGp998K226887
synaptotagmin $_{265-422}$	sytag F265	sytag R422	IMAGp998K226887
synbindin $_{2-219}$	sbd F2	SBD R219	IMAGp998B119638
syntaxin12 $_{2-269}$	synt12 F2	synt12 R269	IMAGp998A0810138
syntaxin1A $_{2-265}$	stx1A F2	stx1A R265	IMAGp958E125
syntaxin1A $_{2-288}$	stx1A F2	stx1A R288	IMAGp958E125
syntaxin5 $_{2-284}$	synt5 F2	synt5 R284	IMAGp998A101942
syntaxin5 $_{2-301}$	synt5 F2	synt5 R301	IMAGp998A101942
syntaxin7 $_{2-236}$	synt7 F2	synt7 R236	IRAKp961E0321
syntaxin7 $_{2-261}$	synt7 F2	synt7 R261	IRAKp961E0321
syntaxin8 $_{2-236}$	synt8 F2	synt8 R236	IMAGp998I112838
syntenin $_{114-273}$	stn F114	stn R273	IRAKp961A1513
syntenin $_{114-298}$	stn F114	stn R298	IRAKp961A1513
syntenin $_{2-273}$	stn F2	stn R273	IRAKp961A1513
syntenin $_{2-298}$	stn F2	stn R298	IRAKp961A1513
Tim44 $_{39-452}$	tim44 F39	tim44 R452	IMAGp998L0510082
UbcH7 $_{2-154}$	ubch7 F2	ubch7 R154	IMAGp998F079995
UbcH8 $_{2-152}$	ubch8 F2	ubch8 R152	IMAGp998L1010008

2.6 Antibodies

Table 8

antibody	description	supplier
goat anti-GST antibody	horseradish peroxidase (HRP) conjugated to goat anti-GST polyclonal antibody	Amersham
mouse anti-RGS-His antibody	mouse IgG1, polyconal	QIAGEN
Rabbit anti-mouse antibody	horseradish peroxidase (HRP) conjugated to rabbit anti-mouse polyclonal antibody	Dako/Cytoma-tion

2.7 Enzymes, proteins and molecular weight standards

- Restriction enzymes — New England Biolabs, Frankf.
- Shrimp alkaline phosphatase — Roche Diagnostics, Mannheim
- T4-DNA-Ligase — New England Biolabs, Frankf.
- Lysozyme — Roche Diagnostics, Mannheim
- Benzonase — Merck, Darmstadt
- Trypsin — Promega, Mannheim
- Protein kinase A — Sigma, Deisenhofen
- Bovine serum albumine — Sigma, Deisenhofen
- Precision protein Standard, prestained — Bio-Rad Laboratories, Münch.
- Precision plus protein Standard, unstained — Bio-Rad Laboratories, Münch.
- 1 kB DNA marker — Invitrogen, Karlsruhe

A preparation of Taq DNA polymerase was a gift from Roman T. Pawlik (MPIMG, Dahlem).

2.8 Reaction systems and kits

- Plasmid Miniprep Kit — QIAGEN, Hildesheim
- Gel extraction Kit — QIAGEN, Hildesheim
- PCR purification Kit — QIAGEN, Hildesheim
- Expand High Fidelity PCR System — Roche Diagnostics, Mannh.
- Western Lightning ECL Plus — PerkinElmer, Rodgau
- BM Chemiluminescence Blotting Substrate (POD) Kit — Roche Diagnostics, Mannh.
- Gel Drying Kit for SDS polyacryleamide gels — Promega, Mannheim

2.9 Reagents an chemicals

- amonium persulfate — Bio-Rad, München
- ampicillin — Sigma, Deisenhofen
- ATP — Sigma, Deisenhofen
- Bradford protein assay reagent, 5x dye — Bio-Rad, München
- Brij 58 (Polyoxyethylene 20 cethyl ether) — Sigma, Deisenhofen
- Coomassie Brillant blue R-250 — Bio-Rad Laboratories, Münch.
- D+ glucose monohydrate — Merck, Darmstadt
- Dimethyl pimelimidate (DMP) — Pierce, Rockford
- dNTP´s, PCR-grade — Roche Diagnostics, Mannh.
- DTT — Merck, Darmstadt
- EDTA — Merck, Darmstadt

▪ ethidium bromide, 1% solution	Fluka, Darmstadt
▪ glycerol	Merck, Darmstadt
▪ glycogen	Boehringer, Mannheim
▪ imidazole	Sigma, Deisenhofen
▪ IPTG	Sigma, Deisenhofen
▪ kanamycin	Sigma, Deisenhofen
▪ L-glutathione, reduced	Sigma, Deisenhofen
▪ N-octyl-ß-D-glucopyranoside	Sigma, Deisenhofen
▪ Non-fat dry milk powder	Roth, Karlsruhe
▪ Nonidet P-40	Sigma, Deisenhofen
▪ PCR buffer, 10X	MBI Fermentas, St. Leon-Rot
▪ PMSF	Roche Diagnostics, Mannh.
▪ Redivue adenosine 5´-[γ-32P]- triphosphate, 250 µCi	Amersham Pharmacia, Freib.
▪ Rotiphorese Gel 30 (29% acrylamide, 0.8% bisacrylamide)	Roth, Karlsruhe
▪ TEMED	Bio-Rad, München
▪ Thiamine	Sigma, Deisenhofen
▪ Triton X-100	Sigma, Deisenhofen
▪ Tween 20	Sigma, Deisenhofen

Anorganic salts, acids, bases and alcohols were from Merck, Darmstadt.

2.10 Buffers and media

▪ Benzonase mix	50 mM Tris-HCl, pH 8.0 12 mM $MgCl_2$ 10 U/mL Benzonase added
▪ Blot transfer buffer	25 mM Tris-HCl 192 mM Glycine 10 % methanol
▪ Coomassie staining solution	45 % methanol 45 % ddH_2O 10 % acetic acid 0.2 % Coomassie blue R-250
▪ Denaturing solution	1.5M NaCl 0.5M NaOH
▪ DNA loading buffer 4X	10 mM Tris-HCl, pH 8.0 30 % glycerol 0.01 % bromphenolblue
▪ Freezing Mix 10X	0.27M KH_2PO_4 0.13M K_2HPO_4 68 mM $(NH_4)_2SO_4$ 4 mM $MgSO_4$ 15 mM Na_3-citrate 36 % glycerol
▪ Glucose 40 % (w/v)	40 g glucose-monohydrate dissolve in ddH_2O to 100 mL

- Kinase buffer 10X
  200 mM Tris-HCl, pH 7.5
  1 M NaCl
  120 mM $MgCl_2$
  10 mM DTT
- KPB 20X
  46 g KH_2PO_4
  243 g K_2HPO_4
  dissolve in ddH_2O to 1 liter
- Lysis buffer
  50 mM Tris-HCl, pH 8.0
  300 mM NaCl
  0.1 mM EDTA
- Lysozyme mix
  50 mM Tris-HCl, pH 8.0
  0.5 % Brij58 (w/v)
  0.2 % lysozyme (w/v)
- MALDI matrix solution
  50 % acetonitrile
  1 % CHCA
  0.1 % trifluoracetic acid
- NiNTA elution buffer
  50 mM Tris-HCl, pH 8.0
  300 mM NaCl
  250 mM imidazole
- NiNTA wash buffer
  50 mM Tris-HCl, pH 8.0
  300 mM NaCl
  20 mM imidazole
- PBS 10X
  1.37 M NaCl
  27 mM KCl
  100 mM Na_2HPO_4
  20 mM KH_2PO_4
  adjust pH to 7.4
- PBS-T
  PBS 1X
  0.1% Tween 20
- SDS loading buffer 4X
  200 mM Tris-HCl, pH 6.8
  40 % glycerol (w/v)
  8 % SDS
  0.004 % bromphenolblue
  100 mM DTT (separately)
- Sodiumborate buffer
  200 mM sodiumborate, pH 9
- SSC 20X
  3M NaCl
  0.3M Na_3-citrate
- TBE buffer
  90 mM Tris-borate, pH 8.0
  1 mM EDTA
- TBS 10X
  0.5M Tris-HCl, pH 8.0
  1.5M NaCl

▪ TBS-T	TBS 1X 0.05 % Tween 20
▪ TE buffer	10 mM Tris-HCl, pH 8.0 1 mM EDTA
▪ 2YT broth	1.6 % (w/v) Bacto-Tryptone 1.0 % (w/v) Bacto-Yeast extr. 0.5 % (w/v) NaCl adjust pH to 7.0
▪ 2YT agar	1.6 % (w/v) Bacto-Tryptone 1.0 % (w/v) Bacto-Yeast extr. 1.0 % (w/v) Bacto-Agar 0.5 % (w/v) NaCl adjust pH to 7.0
▪ LB medium	1.0 % (w/v) Bacto-Tryptone 0.5 % (w/v) Bacto-Yeast extr. 1.0 % (w/v) NaCl adjust pH to 7.0
▪ SOB medium	2.0 % (w/v) Bacto-Tryptone 0.5 % (w/v) Bacto-Yeast extr. 10 mM NaCl 2.5 mM KCl adjust pH to 7.0

Bacto-Tryptone, Bacto-Yeast extract and Bacto-Agar were from Difco Laboratories (Detroit, USA). Media, Freezing mix and KPB were autoclaved. Buffers and solutions were sterilfiltrated, except Coomassie staining and Loading buffers.

2.11 Diverse materials

▪ 3MM Blotting paper	Whatman, Göttingen
▪ Agarose, ultra pure	Gibco/BRL, Eggenstein
▪ Cellulose acetate filter, 0.45 µM pore size	Sartorius, Göttingen
▪ Chroma Spin columns	Clontech, Palo Alto
▪ Electroporation cuvettes, 0.1 cm gap	Bio-Rad Laboratories, Münch.
▪ Glutathione agarose	Sigma, Deisenhofen
▪ Microtitre plates, 96-well	Nunc, Wiesbaden
▪ NiNTA agarose	QIAGEN, Hildesheim
▪ Petri dishes	Nunc, Wiesbaden
▪ Polycarbonate centrifugal tubes, 10 mL	Oak Ridge Nalgene, USA
▪ Polypropylen Falcon tubes, 15 mL and 50 mL	Becton-Dickinson, Heidelberg
▪ PVDF filters, Immobilon P	Millipore, Eschborn
▪ Reaction vials, 0.5 mL, 1.5 mL and 2 mL	Eppendorf, Hamburg
▪ Sephadex G-50 (medium grade)	Amersham Pharmacia, Freibg
▪ V-bottom 96-well PCR plates, Thermo-Fast	Rapidozym, Berlin

2.12 Laboratory hardware equipment

▪ Freezer, -80°C, Forma	Thermoquest GmbH, Egelsb.
▪ Incubator WB 120	Mytron, Heiligenstadt
▪ Shaker incubator Innova 4430	New Brunswick Scie., Nürt.
▪ Bench top centrifuge 5804R	Eppendorf, Hamburg
▪ Table top centrifuge 4K15	Sigma, Osterode
▪ Table top centrifuge Labofuge 1	Heraeus-Christ, Osterode
▪ High Performance centrifuge Avanti J-20XP	Beckman Coulter, USA
▪ Ultracentrifuge L7-55	Beckman Coulter, USA
▪ Mass Spectrometer Reflex II MALDI-TOF	Bruker Franzen, Bremen
▪ DNA Engine Thermal Cycler PTC-200	MJ Research, USA
▪ Power supply	Bio-Rad Laboratories, Münch.
▪ Cell electroporator Gene Pulser II	Bio-Rad Laboratories, Münch.
▪ Gel electrophoresis equipment Hoefer Mighty Small	Amersham, Freiburg
▪ Semidry blotting device	Hoelzel, Wörth/Hörlkofen
▪ DNA electrophoresis equipment	MPIMG workshop, Dahlem
▪ PhosphorImager SI	Molecular Dynamics, Krefeld
▪ CCD camera Fuji LAS-1000	Raytest, Straubenhardt
▪ UV/Visible spectrophotometer, Vary 50Scan	Varian Inc., Darmstadt
▪ Microplate spectrophotometer Mikro-Tek DS	Bio-Tek, Bad Friedrichshall
▪ Thermoblock Techne DB-2D	Labtech, Jahnsdorf
▪ Orbital platform shaker, Type 3015	Ges. f. Labortechnik, Burgw.
▪ Table top roller RM5	Hecht, Sondheim
▪ Membrane vacuum pump	KNF Neuberger, Switzerland
▪ Vortex mixer Genie II	Scientific Industries, USA
▪ Magnetic stirrer RCT basic	IKA Labortechnik, Staufen
▪ Precision balance MPX-200	Denver Instrument, USA
▪ Pipettes, adjustable	Abimed, Langenfelden
▪ Pipettes, adjustable, 8-channel	Dunn Labortechnik, Asbach
▪ Sterile filtration device, filter holder with receiver	Nalgene, USA

3 Methods

3.1 DNA techniques and cloning

3.1.1 Electrophoretic DNA separation

Agarose gel electrophoresis is used for separation, purification and identification of plasmid DNA and DNA fragments. Because of the negative charge of its phosphate groups, the DNA moves towards the plus-pole in an electric field. Under constant voltage the migration speed of linear, double-stranded DNA in agarose gels is proportional to the logarithm of its molecular weight. The size of a DNA fragment can be determined by comparison with standard DNA marker fragments of identified size. Rough estimation of the DNA concentration can be made by comparing the band intensity of the sample and a reference marker DNA band upon staining with ethidium bromide. Depending on the size of the DNA molecules, the agarose concentration chosen was between 0.8 % and 1.2 % (w/v). DNA-samples were mixed with 4x DNA loading buffer and applied into the wells of the gel. In parallel, a marker was loaded. TBE buffer was used for agarose solution and as electrophoretic buffer. The electrophoretic separation was done at 5 volts/cm. Separated fragments were visualized by UV-light.

3.1.2 Plasmid DNA preparation

Plasmid extractions from 2 mL *E. coli* overnight cultures in LB medium were performed with the Plasmid miniprep Kit from QIAGEN according to the manufacturer's recommendations. This procedure uses alkaline lysis according to Birnboim and Doly (1979). The purified plasmid DNA was concentrated through a precipitation step by adding 0.7 volumes of isopropanol. Precipitated plasmid DNA was washed once in 75 % ethanol and resuspended in a suitable volume of ddH_2O. Purity and quantity of the isolated plasmid DNA was checked by agarose gel electrophoresis.

3.1.3 Enzymatic manipulation of plasmid DNA

3.1.3.1 Restriction enzyme digestion

Restriction endonucleases catalyze a sequence specific cleavage of doublestranded DNA, thereby generating cohesive ends or blunt ends. The amounts of restriction enzyme and DNA, the buffer and ionic concentrations, and the temperature and duration of the reaction varies depending upon the specific application. The conditions were chosen as proposed in the manufacturer’s manual. In a typical reaction, 1-5 units of restriction enzyme per µg plasmid DNA were mixed in a final volume of 50 µL. Doubledigested DNA fragments were seperated from released

polylinkers either by gel electrophoresis followed by gel extraction (QIAGEN Kit), or by size exclusion using a Chromaspin column (Clontech).

3.1.3.2 Dephosphorylation

Shrimp alkaline phosphatase (SAP) catalyzes the dephosphorylation of 5´-end phosphat groups of cleaved DNA. Reaction conditions reflected to the manufacturer´s recommendation. SAP was heat-inactivated for 20 min. at 65°C and seperated from dephosphorylated DNA with a Chromaspin column (Clontech).

3.1.4 PCR amplification of DNA fragments

The Polymerase Chain Reaction (PCR) selectively amplifies a DNA sequence with two flanking oligonucleotide primers. PCR is an iterative process, consisting of three steps: denaturation of the template DNA by heat, annealing of the oligonucleotide primers to the single-stranded target sequence, and extension of the annealed primers by a thermostable DNA polymerase. A typical 50 µl reaction mixture consisted of 1-10 ng plasmid DNA as a template, 20 pmol of each primer (forward and reverse), 200 µM of each dNTP and one unit of Taq DNA polymerase in 1x PCR buffer. The reaction was incubated in a thermocycler device. Usually there was a preheating step of 5 min. at 94°C during which the template DNA was denatured. This was followed by 20-28 cycles of denaturing (60 sec. at 94°C), annealing (60 sec. at 55-65°C) and elongation (120 sec. at 72°C). The last cycle was followed by an extra elongation step of 5 min. at 72°C. The correct size and the purity of amplified PCR fragments was monitored by agarose gel electrophoresis.

3.1.5 Purification of DNA

PCR products were purified by using a PCR purification kit (QIAGEN) in accordance with the manufacturer's instructions.

3.1.6 Ligation of DNA fragments

The T4 DNA ligase catalyzes the repair of single-stranded nicks in duplex DNA and joins duplex DNA restriction fragments having either blunt or cohesive ends. Cohesive end ligations were carried out for 1hr. at 37°C or overnight at 4°C using a molar 1:4 ratio of vector to insert. A typical 20 µl reaction mixture contained 10-200 ng total DNA mass, 1-2 units of T4 DNA ligase in T4 DNA Ligase buffer.

3.1.7 Precipitation of ligated DNA

The ligated DNA must be liberated from high salt concentrations for an efficient transformation of *E. coli* by electroporation. This was achieved by precipitation of DNA with

ammonium acetate. DNA was precipitated by adding 0.5 volumes of 7.5M ammonium acetate (pH 5.0) and 2.5 volumes of icecold ethanol. The DNA pellet was washed once in 75 % ethanol and resuspended in a suitable volume of ddH_2O.

3.1.8 Preparation of electrocompetent *E. coli* cells

For preparing cells competent for electroporation the bacterial culture was grown to an OD_{600nm} of 0.6-0.7 in 500 ml LB medium in a two liter flask at 160 rpm and 37°C. The culture was then chilled for 20 min. in an ice-water bath and centrifuged 15 min. at 1,600 x *g* and 4°C. The cell pellet was subsequently washed 3 times with icecold sterile 10 % glycerol. It is important to keep cells on ice at all steps. To freeze competent cells for electroporation, a volume of ice-cold 10 % glycerol roughly equal to the one of the cell pellet was added and mixed thoroughly. Aliquots of 50 µl of cell suspensions were then transferred into prechilled 0.5 mL Eppendorf tubes and frozen in liquid nitrogen. The cells could then be stored at –80°C for several months without losing their transformation competence.

3.1.9 Transformation of *E. coli*

Electroporation with high voltage is currently the most efficient method for transforming *E. coli* with plasmid DNA. The exposure of a dense suspension of bacterial cells and plasmid DNA to a high strength electric field of short duration induces DNA uptake and elicits competence for stable plasmid transformation. With freshly grown cells, electroporation routinely gives more than 10^9 bacterial transformants per microgram of plasmid DNA. Approximately 50 µl of electrocompetent cells were thawed on ice and mixed with a volume of plasmid DNA not exceeding 3 µl, which usually contained 10 to 100 ng of plasmid DNA. The mixture of cells and DNA were placed subsequently into a prechilled electroporation cuvette with 0.1 cm gap and subjected to a pulse of a strong electric field in the electroporation apparatus (Gene Pulser II, Bio-Rad). The setting was 2,5 kV, 25 µF and 200-400 ohms. Immediately after the electroporation, cells were recovered with 0.5 mL SOB medium supplemented with glucose (20 mM) and $MgCl_2$ (20 mM) incubated for 30 min. at 37°C and than plated onto petri dishes containing 2YT agar and appropriate antibiotics for the selection of transformants.

3.1.10 Colony picking and long-term storage of *E. coli* clones

For long-term storage, *E. coli* transformants were picked into microtitre plates containing 2YT medium supplemented with freezing mix and appropriate antibiotics. After growth overnight at 37°C microtitre plates were frozen at -80°C. Bacterial cells will only survive a limited number of freezing and thawing cycles. Therefore a sufficient number of replicate microtitre plates was stored frozen at any time.

3.1.11 DNA sequencing

Sequencing of DNA relies on the dideoxynucleotide-method developed by Sanger (1992). This technique utilizes 2',3'-dideoxynucleotide triphospates (ddNTP´s), molecules that differ from deoxynucleotides by having a hydrogen atom attached to the 3' carbon rather than an OH group. These molecules terminate DNA chain elongation because they cannot form a phosphodiester bond with the next deoxynucleotide. A sequencing reaction contains a mixture of fluorescently labeled ddNTP´s with the normal dNTP´s. The logic behind this ratio is that after DNA polymerase is added, the elongation will take place and will terminate whenever a ddNTP is incorporated into the growing strand. If the ddNTP is only 1 % of the total concentration of dNTP, a whole series of labeled strands will result. The DNA mixture can be separated by cappilary gel electrophoresis and is analysed base-specifically due to the four differently labeled ddNTP´s on an Abi Prism Genetic Analyser. DNA sequencing was done in house by the service group of our department and at AGOWA (Berlin, Adlershof).

3.2 Recombinant protein expression in *E. coli*

3.2.1 Microtitre plate cultures

Ninety-six-well microtitre plates with 2 ml cavities were filled with 100 µl 2YT medium supplemented with appropriate antibiotics and 2 % glucose. Cultures were inoculated with bacteria from 96-well plates that had been stored at -80ºC. For inoculation, replicating devices carrying 96 steel pins, were used. After overnight growth in an orbital shaker at 37ºC and 220 rpm, 900 µl of 2YT medium prewarmed to 37ºC and supplemented with antibiotics and 20 µg/mL thiamine were added to the cultures, and incubation was continued for 3 hrs. For induction of protein expression, IPTG was added to a final concentration of 1 mM, and incubation continued for 3 hrs. Cells were harvested by centrifugation at 1,900 x *g* for 10 min., washed by resuspension in Lysis buffer, centrifuged for another 5 min. and cell pellets were stored at -80ºC.

3.2.2 Flask cultures

Twenty mL 2YT medium supplemented with appropriate antibiotics and 2 % glucose was inoculated with the desired expression clone from a frozen stock. After overnight growth in an orbital shaker at 37ºC and 220 rpm, the 20 mL preculture was transferred to 200 mL of 2YT medium prewarmed to 37ºC supplemented with appropriate antibiotics and 20 µg/mL thiamine, and incubation was continued until an OD_{600} of 0.7-0.8 was reached. For induction of protein expression, IPTG was added to a final concentration of 1 mM, and incubation was continued for 3

hrs. Cells were harvested by centrifugation at 1,900 x *g* for 20 min., washed by resuspension in Lysis buffer, centrifuged for another 5 min. and cell pellets were stored in four aliqouts at -80ºC.

3.3 Purification of recombinant proteins

3.3.1 Protein extraction

3.3.1.1 Protein extraction in microtitre plates

Bacterial pellets of 1 mL cultures in microtitre plates were thawed and resuspended in 100 µL icecold lysis buffer containing 0.5 mM PMSF. Twenty-five µL Lysozyme mix was added to each well, the resuspension was briefly vortexed and incubated for 30 min. on ice. DNA was decomposed by adding 25 µL Benzonase mix and incubation was continued for 30 min. at room temperature. The whole cell lysate preparation was cleared by centrifugation in a table top centrifuge for 30 min. at 6,200 rpm. Approximately 150 µL of the supernatant representing the soluble protein extract was used for further purification.

3.3.1.2 *Protein extraction from flask cultures*

Typically, bacterial pellet from a 50 mL culture was thawed and resuspended in 5 mL icecold Lysis buffer containing 0.5 mM PMSF. Cells were lysed by adding 1.25 mL Lysozyme mix and incubated for a duration of 30 min. on ice. DNA was decomposed by adding 1.25 mL Benzonase mix and incubation was continued for 30 min. at room temperature. The whole cell lysate was then transferred to a polycarbonate tube. To remove cellular debris the lysate was cleared by ultracentrifugation for 45 min. at 40,000 rpm and 4 °C. The supernatant was taken for subsequent purifications.

3.3.2 Purification of recombinant proteins

3.3.2.1 Purification of GST fusion proteins

Typically, 2 mL of a 50 % (v/v) glutathione agarose beads suspension equilibrated in Lysis buffer were added to 10 mL cleared lysate containing the desired GST fusion protein. The mixture was incubated on a table roller for 1 hr. at 4°C to enable binding of the GST fusion protein to the glutathione agarose beads. All of the following steps were performed at 4°C. The glutathione agarose beads were transferred to a 2 mL purification column and washed extensively with TBS. Bound GST fusion protein was eluted into 200 µL fractions using TBS with 20 mM reduced glutathione peptide and the protein containing fractions were pooled.

3.3.2.2 NiNTA purification of His-tagged proteins

Typically, 2 mL of a 50 % (v/v) NiNTA agarose beads suspension equilibrated in Lysis buffer were added to 10 mL cleared lysate containing the desired His-tagged protein. The mixture was incubated on a table roller for 1 hr. at 4°C to enable binding of the His-tagged protein to the NiNTA agarose beads. All following steps were performed at 4°C. The agarose beads were transferred to a 2 mL purification column and washed extensively with NiNTA wash buffer. Bound His-tagged protein was eluted into 200 µL fractions using NiNTA elution buffer and the protein containing fractions were pooled.

3.3.2.3 NiNTA purification of His-tagged proteins in 96-well microtitre plates

Approximately 150 µL of soluble protein extract were transferred to a V-bottom plate and 25 µL of a 50 % (v/v) NiNTA agarose beads suspension were added to each well. All of the following steps were performed at 4°C. His-tagged proteins were allowed to bind to the NiNTA beads by incubating the mixtures for 1 hr. by end over end rotation. The beads were washed extensively with NiNTA washing buffer containing 20 mM imidazole. Bound His-tagged proteins were used in subsequent pulldown experiments or assayed for purified His-tagged protein by adding 4X SDS loading buffer to the NiNTA beads for SDS-PAGE.

3.4 Protein-protein-interaction studies

3.4.1 High-density protein array screening

3.4.1.1 High-density spotting of expression clones onto filter membranes

The human fetal brain cDNA expression library MPMGp800 is enriched for inframe cDNA inserts and consists of 36,864 individual clones. The expression library is stored at –80°C in 96 separate 384-well microtitre plates containing 2YT medium and 2 % glucose supplemented with Freezing mix. Bacteria were grown in microtitre plates overnight and clones were gridded onto 222 mm x 222 mm PVDF filter membranes as an orthogonal equidistant array. The PVDF filter membrane was wetted in ethanol for at least 5 minutes, then washed twice in distilled water and finally in 2YT broth. The PVDF filter membrane was placed onto 3MM Whatman blotting paper soaked with 2YT broth. Prior to spotting, air bubbles and excess liquid was removed by rolling with a long glass pipet. The spotting robot divides the membrane area into six equal fields. The size of each field corresponds to one 384-well microtitre plate and consists of 384 blocks with a 5 x 5 spotting pattern. 12 individual clones were spotted in duplicates in each block with black ink as guide spot in the center of each block. After spotting, the PVDF filter membrane was

placed onto a square 2YT agar plate (23 cm x 23 cm, Bio Assay Dish, Nunc) supplemented with antibiotics and 2 % glucose. Colonies grew on the filter overnight at 30ºC to a size of approximately 1 mm diameter. After growing overnight, the PVDF filter membrane was transferred onto prewarmed 2YT agar plates supplemented with antibiotics, 2 % glucose and 1 mM IPTG to induce protein expression for 3 hrs. at 37ºC.

3.4.1.2 Denaturing release of cellular proteins

This procedure uses alkaline conditions to release cellular proteins on a PVDF filter membrane. A sheet of 3MM Whatman blotting paper was placed into the lid of a square agar plate dish and Denaturing solution was added. Excess liquid was removed and the PVDF filter membrane was carefully transferred onto the 3MM Whatman blotting paper with forceps. After a 10 min. incubation with Denaturing solution the PVDF filter membrane was placed twice for 5 min on 3MM Whatman blotting paper soaked with Neutralising solution and finally on 2X SSC for 15 min. The PVDF filter membrane was transferred on a dry blotting paper, allowed to air-dry and stored at room temperature between two sheets of blotting paper.

3.4.1.3 Radioactive labeling of GST fusion proteins

Two hundred units of lyophilized cAMP-dependent protein kinase A from heart muscle were reconstituted in 20 µl 40 mM DTT for 15 min. at room temperature. 500 µg of purified GST fusion protein containing the recognition site for protein kinase A were diluted in 160 µl Kinase buffer and added to 20 µL of the reconstituted protein kinase A. The phosphorylation reaction was started by adding 20 µl of [γ-^{32}P]ATP (20 µl 1 mM ATP, 20 dpm/nmol). After 1 hr. at 25°C the phosphorylated fusion protein was separated from unincorporated [γ-^{32}P]ATP by gel filtration chromatography. The reaction mix (200 µl) was loaded onto a Sephadex G-50 (medium grade) gel filtration column with 2.5 mL bed volume equilibrated in 20 mM Hepes-KOH, 50 mM KCl, 0.1 mM EDTA, 2.5 mM $MgCl_2$, pH 7.4. A total of 10 fractions each of 200 µl equilibration buffer were eluted from the gel filtration column by spinning the column 5 min. at 1,800 rpm in a Labofuge 1. The Cerenkov counts in 1 µL aliquots of each fraction were monitored in a liquid scintillation counter. Two peaks of radioactivity usually eluted from the column and were well separated from each other. The first peak contained the phosphorylated fusion protein, whereas the second peak contained the unincorporated [γ-^{32}P]ATP.

3.4.1.4 Blocking and probing filter membranes

The dried PVDF filter membrane carrying the protein array was wetted in ethanol for at least 5 min., then washed twice for 5 min. each in TBS-T. Washing steps were performed by shaking the PVDF filter membrane in a plastic box on a rocker submerged in a large volume, approximately one liter, of TBS-T (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05 % (v/v) Triton

X-100). The PVDF filter membrane was blocked in blocking buffer (20 mM Hepes-KOH, 5 mM $MgCl_2$, 5 mM KCl, 0.1 mM EDTA, pH 7.4, 0.05 % (v/v) Nonidet P-40, 4 % (w/v) non-fat dry milk powder) in the cold room for 3-4 h on a rocker and then equilibrated in hybridization buffer (20 mM Hepes-KOH, 50 mM KCl, 0.1 mM EDTA, 2.5 mM $MgCl_2$, pH 7.4, 0.05 % (v/v) Nonidet P-40, 1 % (w/v) non-fat dry milk powder) for 15 min. The radioactively labelled GST fusion protein was diluted in 20 ml hybridization buffer and added to the blocked PVDF filter membrane. Incubation was done in the cold room for at least 12 hrs. After incubation the PVDF filter membrane was washed three times, each for 15 min. with 50 ml hybridization buffer.

3.4.1.5 Image analysis and clone identification

The air-dried PVDF filter membrane was covered with Saran wrap and exposed to a storage phosphor screen. Images of the probed PVDF filter membrane were generated from the storage phosphor screen by scanning with a phosphor imager. Subsequently, images of probed PVDF filter membranes were analysed with the software Xdigitise (Wruck *et al*. 2002). Xdigitise can be used to score positive clones on the image and to retrieve their microtitre plate position. First, x-y coordinates of the detected doublet signals were calculated with Xdigitise. The microtitre plate position of the corresponding clone can be retrieved with the calculated x-y coordinates from the website of the RZPD. Clones of the human fetal brain expression library MPMGp800 are available upon request at the distribution service of the RZPD.

3.4.2 Pulldown assays

3.4.2.1 Pulldown assays with NiNTA agarose beads

Bacterially expressed His-tagged proteins were affinity-purified with NiNTA beads without elution from the beads. 50 µL of a 50 % (v/v) NiNTA beads slurry containing 10-50 µg of purified His-tagged protein were incubated with 100 µg of purified GST fusion protein or GST as a control in a total volume of 200 µl TBS containing 0.1 % (w/v) BSA. After 2 hrs. of incubation under continuous end over end rotation at 4°C the NiNTA beads were washed 5 x 5 min. with the same buffer and 2 x 5 min. in TBS without BSA. NiNTA beads were mixed with an equal volume of 2X Loading buffer and bound proteins were eluted by incubation for 3 min. at 90°C. Ten µl aliquots of eluted proteins were subsequently separated on SDS-PAGE gels.

3.4.2.2 Pulldowns assays with glutathione agarose beads

GST fusion proteins were purified on glutathione agarose without elution. For covalent crosslinking, beads with bound GST fusion protein were washed twice in PBS and then twice in sodiumborate buffer. After that, beads were incubated for 30 min. in sodium borate buffer containing 20 mM dimethyl pimelimidate (DMP) crosslinker on a roller at room temperature. The

crosslinking reaction was stopped by washing the beads once in 0.2M ethanolamine solution (pH 8.0). The incubation was continued for 2 hrs. at room temperature. Then the beads were washed once in PBS and twice in PBS including 20 mM reduced glutathione. After equilibration in Lysis buffer, a 50 µL aliquot of a 50 % (v/v) glutathione agarose beads slurry in Lysis buffer with crosslinked GST fusion protein was incubated with 10-50 µg of purified His-tagged protein in a total volume of 200 µL. After 2 hrs. of incubation under continuous end over end rotation at 4°C the beads were washed 5 x 5 min. in Lysis buffer. To elute bound proteins, beads were mixed with an equal volume of 2X SDS loading buffer without DTT and bromphenolblue. Eluates without beads were mixed with 4X SDS loading buffer and incubated for 3 min. at 90°C. Ten µl aliquots of eluted proteins were subsequently separated on SDS-PAGE gels.

3.4.3 Peptide scans

3.4.3.1 Preparation of peptide scans

Scans of overlapping peptides derived from amino acid sequences of binding proteins were used to map linear epitopes responsible for protein-protein-interactions. Series of tetradecapeptides were generated, covering partial protein sequences of identified binding proteins with an overlap of ten amino acids. For the preparation of the peptides, SPOT synthesis on membrane supports was applied (Frank, 1992; Kramer, 1999). The peptides were covalently bound to a Whatman 50 cellulose support by the C-terminus and had a free N-terminus. The solid phase-bound peptides were used for binding studies directly on the membrane. A customized peptide array consisting of 12 individual peptide scans was prepared at Jerini AG, Berlin.

3.4.3.2 Blocking and probing of peptide arrays

After washing the cellulose membrane once with ethanol and twice with TBS, the peptide array was blocked for 4 hrs. with TBS containing 10 % (v/v) blocking reagent (Roche Applied Science) and finally incubated with 20 µg/mL of GST-fusion protein or GST in blocking buffer at 4°C for 16 hrs. After threefold washing of the peptide array with TBS, the bound GST protein was immunodecorated with horseradish peroxidase conjugated to goat anti-GST polyclonal antibody at 1:5000 in blocking buffer followed by detection with BM chemiluminescence substrate POD. Images were taken with a Fujifilm LAS-1000 CCD camera.

3.5 Protein analysis

3.5.1 SDS-PAGE

One dimensional gel electrophoresis under denaturing conditions in presence of 0.1 % SDS separates proteins according to their molecular size. The polyacrylamide gel is cast as a separating gel topped by a stacking gel. Sample proteins were solubilized by boiling in SDS loading buffer and individual proteins were separated electrophoretically. The mobility of the proteins is inversely proportional to the logarithm of their molecular mass. SDS was employed to effect denaturation of the proteins, to dissociate protein complexes and to impart upon the polypeptide chains net negative charge densities proportional to the length of the molecule. The reducing agent DTT was used to reduce any existing disulphide bond. Coomassie brilliant blue R-250 binds nonspecifically to almost all proteins, which allows detection of protein bands in polyacrylamide gels. Gels are soaked in Coomassie staining solution with gentle shaking for 1 hr. at room temperature. The background was subsequently reduced by soaking the gel in Destaining solution. After that, gels were dried and permanent records of dried gels were taken with a Fujifilm LAS-1000 CCD camera.

3.5.2 Western Blotting

3.5.2.1 Electroblotting from polyacrylamide gels on membranes

Electroblotting of proteins from polyacrylamide gels onto retentive membranes is usually performed for immunoblotting (or Western blotting). Proteins were separated by SDS-PAGE and transferred from the polyacrylamide gel to a PVDF membrane with a constant current of 1 mA per cm^2 of membrane for 1 hr. by semi-dry blotting.

3.5.2.2 Immunodetection of blotted proteins

Antibodies and dilution were as follows: For the detection of GST fusion proteins, membranes were first blocked with PBS-T containing 5 % (w/v) non-fat dried milk. Then horseradish peroxidase (HRP) conjugated to goat anti-GST polyclonal at 1:5000 in PBS-T was used for immunodetection. Washing procedures were according to the manufacturer´s instructions. To detect His-tagged proteins, membranes were first blocked with TBS-T containing 2 % (w/v) BSA, then incubated with primary mouse anti-RGS-His antibody 1:2000 diluted in blocking buffer and finally with secondary rabbit anti-mouse antibody conjugated to HRP 1:4000 diluted in TBS-T containing 5 % (w/v) non-fat dried milk. All signals were visualized with the ECL Plus Western Lightning Kit and detected with a Fujifilm LAS-1000 CCD camera.

3.5.3 Measurement of protein concentration

The color change produced when the Coomassie dye binds to proteins provides a measure of total protein. The absorbance for the protein-specific dye, Coomassie brilliant blue G-250, shifts from 465 nm to 595 nm when binding to protein occurs. Therefore, the absorbance at 595 nm yields a good linear concentration dependence for most soluble proteins. The Bradford Mini assay (Bio-Rad Laboratories, München) adapted to the microtitre plate format was used and the measured absorbance at 595 nm was blotted against a reference curve obtained with known concentrations of BSA.

3.5.4 Mass spectrometry

3.5.4.1 Tryptic digestion

Coomassie Brillant blue stained protein samples were excised from the gel. For each protein band only the center of staining intensity was excised. The excised gel pieces were destained by three subsequent incubations (each 30 min.) at 37 °C with 50 µL of 25 % 2-propanol and 75 % 20 mM Tris-HCl, pH 7.5. The destained gel material was shrunk for 1 hr. with 50 µL 100 % acetonitrile (ACN) and dried by vacuum centrifugation in a speedvac for 5 min. 5 µL of trypsin solution containing 12 ng/µL of modified porcine trypsin (sequencing grade, Promega), 5 mM *n*-OGP, 5 mM DTT, and 20 mM Tris-HCl, pH 7.8 were added to the dried gel pieces. After overnight incubation at 37 °C, 10 µL of 0.5 % TFA solution with 5 mM *n*-OGP were added to each reaction vial to stop tryptic digestion. After 30 min. of incubation at an ambient temperature, peptide samples were stored at –20 °C until mass spectrometric analysis.

3.5.4.2 External peptide calibration standard

Table 9: External peptide calibration standard

No.	Peptide	$[M+H]^+$	$[M(^{13}C)+H]^+$
1	Angiotensin II (human)	1046.5423	1047.5457
2	Angiotensin I (human)	1296.6853	1297.6887
3	Substance P-methylester	1362.7356	1363.7390
4	Neurotensin (clip 1-11)	1446.7494	1447.7527
5	Neurotensin	1672.9175	1673.9202
6	ACTH (clip 1-17)	2093.0867	2094.0901
7	ACTH (human clip 18-39)	2465.1989	2466.2022
8	Somatostatin-28	3147.4715	3148.4749

The peptides were purchased from Bachem (Germany). A stock solution containing 1 pmol/µL of each of the peptides 1-6 and 2 pmol/µL of peptides 7 and 8 in 35 % acetonitrile (ACN) and 0.1 % trifluoracetic acid (TFA) was prepared following the quantity specifications recommended by the

manufacturer. This solution was diluted 1 in 20 with a saturated solution of α-cyano-4-hydroxycinnamic acid in 50 % ACN and 0.1 % TFA (CHCA matrix solution) and used as a peptide calibration mixture.

3.5.4.3 Sample preparation

The α-cyano-4-hydroxycinnamic acid (CHCA) surface affinity preparation of peptide samples (Gobom *et al*., 2001) was applied for MALDI-TOF mass spectrometric analysis. A Bruker Scout-384 MALDI sample support with 384 target positions arranged according to the microtiter plate format was used. The sample support consists of a stainless steel surface. 0.5 µL of a MALDI matrix solution saturated with α-cyano-4-hydroxycinnamic acid were loaded onto each target position and immediately mixed with 0.5 µL aliquots of peptide samples. The mixture was gently scratched with a pipet to initialize crystallization. After 3 min. of incubation excess of solvent was carefully drawn apart by a tissue tip. The dry crystalline deposit was rinsed carefully with a small volume of 0.1 % TFA. All samples were deposited onto central positions of the MALDI support. In addition, 0.5 µL of the external peptide calibration standard was placed between the sample positions.

3.5.4.4 MALDI-TOF-MS fingerprints

All mass spectra were recorded automatically on a Bruker Scout 384 Reflex II mass spectrometer (Bruker-Franzen Analytik GmbH, Bremen, Germany) equipped with a SCOUT ion source in reflector mode. Ion acceleration voltage was 25 kV. Only positively charged ions were detected, and approximately 120 single-shot spectra were accumulated from each sample. Ions below m/z 800 were excluded from the analysis. External spectra calibration was performed using a standard peptide mixture. All further processing including baseline correction and peak assignment was achieved using the software package XMASS 5.0, provided by the manufacturer. Peak labeling was performed with the software routine SNAP. This algorithm matches the peaks in the spectrum with expected peptide ion isotopic patterns and, in case a match is obtained, extracts the corresponding monoisotopic mass per charge values automatically. Detected peptide signals were matched with expected protein sequences by the software package GPMAW 4.21 (Lighthouse Data, Denmark). Peptides matching within 200 ppm were considered identified. Protein identification by MALDI-TOF-MS peptide mapping was accomplished using the search engine MASCOT (Matrix Science Ltd., UK) to query a comprehensive nonredundant sequence database such as the combined SWISS-PROT and TrEMBL protein database (Boeckmann *et al*., 2003). Databases were searched with tryptic peptide molecular mass maps determined by MALDI-TOF-MS analysis.

4 Results

4.1 Expression and phosphorylation of GST fusion proteins

To detect protein-protein-interactions on a PVDF membrane carrying a protein array, a suitable probe protein has to be generated. The protein array can then be overlayed with the probe and spots of proteins which have bound the probe can be detected on the membrane. To detect and amplify signals associated with a binding event, the probe protein can be directly labeled. Radiolabeling of a purified probe protein at a single site has a minimal impact on the structural integrity of the protein. Therefore, GST fusion proteins with a unique phosphorylation site for protein kinase A (PKA) interposed between GST and the fusion portion were used as probes in this study. A number of GST fusion proteins with a PKA site were expressed from *E. coli*, purified and examined for the incorporation of [γ-^{32}P].

4.1.1 Construction of a vector for GST fusion proteins with a PKA site

The vector pGEX-2TK contains a PKA site but could not be used in this study for cloning of inserts with *SalI/NotI* overhangs. Therefore, a vector for expression of GST fusion proteins with a PKA site was generated which was suitable for cloning of inserts with *SalI/NotI* overhangs from vector pGEX-6P-2. A 120 bp fragment including the PKA site was PCR amplified using pGEX-2TK as a template. The forward primer pKA3F contains a *BglII* restriction site and a sequence coding for the amino acids RRASV which represent the phosphorylation site of PKA. The primer pGEX-3´ matched a sequence downstream of the multiple cloning site of pGEX-vectors and was used as a reverse primer. The PCR product was digested with *BglII/EcoRI* and the resulting 26 bp fragment was cloned into the corresponding *BamHI/EcoRI* sites of pGEX-6P-2, thereby deleting the *BamHI* site of the vector (*Figure 1*).

```
     GCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCG
1    ---------+---------+---------+---------+---------+---------+  60
     A  W  P  L  Q  G  W  Q  A  T  F  G  G  G  D  H  P  P  K  S

     GATCTGGAAGTTCTGTTCCAGGGGCCCCTGGGATCTCGTCGTGCATCTGTTGGATCCCCG
61   ---------+---------+---------+---------+---------+---------+  120
     D  L  E  V  L  F  Q  G  P  L  G  S  R  R  A  S  V  G  S  P
      EcoRI      SalI       NotI
      |          |          |
     GGAATTCCCGGGTCGACTCGAGCGGCCGCATCGTGACTGACTGACGATCTGCCTCGCGCG
121  ---------+---------+---------+---------+---------+---------+  180
     G  I  P  G  S  T  R  A  A  A  S  *

     TTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTG
181  ---------+---------+---------+---------+---------+---------+  240
```

Figure 1. Map of pGST-PKA. A 26 bp fragment from pGEX-2T-K was inserted into pGEX6P-2 (bp 96-121). Amino acid sequence representing the phosphorylation site for protein kinase A (bold underlined), forward primer pGEX3´ and reverse primer pGEX5´ (underlined), deleted *BamHI* restriction site (bold underlined), restriction sites for *SalI*, *NotI* and *EcoRI* are indicated.

The proper integrity of the 26 bp fragment from pGEX-2TK was controlled using the primers pKA3F and pGEX3´ (*Figure 2*). A 120 bp fragment was amplified exclusively from transformants harbouring the newly generated pGST-PKA construct. A positive clone was sequenced and used to prepare pGST-PKA for expression cloning of GST fusion proteins with a PKA site between GST and the fusion portion.

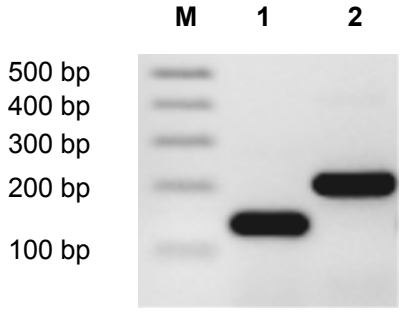


Figure 2. Insert integrity of pGST-PKA. The integrity of the insert from pGEX-2T-K (see *Figure 1*, bp 96-121) was checked by PCR using the primers pKA3F and pGEX-3´. The 120 bp fragment was exclusively amplified from plasmids containing the insert (lane 1). PCR products of 210 bp were amplified with primers pGEX5´and pGEX3´ (lane 2).

4.1.2 Expression and purification of GST fusion proteins with a PKA site

A total of 75 PCR products representing 39 different proteins were generated (see *Table 7*, materials section). The PCR products code for full length proteins as well as protein fragments. All 39 proteins are believed to play a role in neurodegenerative disorders, among them several proteins which have previously been demonstrated to interact with each other. For expression of GST fusion proteins with a PKA site, *E. coli* SCS1 was transformed with the vector pGST-PKA that carries a tac promoter for IPTG-inducible recombinant protein expression. In addition, *E. coli* SCS1 contained the helper plasmid pSE111 with the *argU* gene for the rare arginine tRNA which was shown to improve the expression of eukaryotic genes with multiple arginine codons (Schenk *et al*., 1995). Twenty-two of the 75 generated pGST-PKA constructs were chosen randomly for expression (for details see *Table 10*).

Table 10. Expression of randomly chosen GST fusion proteins with PKA site.

GST fusion protein	note	aa	protein name	GenBank accession
amphiphysin-1	C-terminus	560-695	amphiphysin (Stiff-Man syndrome128kD autoantigen)	NP_001626
auxilin*	C-terminus 1	550-913	KIAA0473 protein	BAA32318
auxilin (C-term.)	C-terminus 2	816-913	KIAA0473 protein	BAA32318
CHIP	full length	2-303	carboxy terminus of Hsp70-interacting protein	AAD33400
cortactin*	full length	2-550	cortactin; oncogene EMS1	NP_005222
endophilin-1	full length	2-352	SH3-domain GRB2-like 2	NP_003017
endophilin-1 (C-term.)*	C-terminus	292-352	SH3-domain GRB2-like 2	NP_003017
HOP*	full length	2-543	Hsp70/Hsp90-organizing protein	NP_006810
Hsc70*	C-terminus	373-646	heat shock cognate protein 70 kDa	NP_006588
myobrevin	N-terminus	2-75	myobrevin	AAC08434
p20arc	full length	2-168	Arp2/3 protein complex subunit p20-Arc	AAB64192
p60*	full length	2-573	heat shock 60kD protein 1 (chaperonin)	NP_002147
parkin (C-term.)*	C-terminus	304-404	parkin	BAA25751
parkin (N-term.)	N-terminus	2-103	parkin	BAA25751
Rab4*	full length	2-213	GTP-binding protein (RAB4)	AAA60244
SH3P7	full length	2-430	src homology 3 domain-containing protein HIP-55	NP_054782
SNAP25*	full length	2-206	synaptosomal-associated protein (25kD)	XP_045655
synapsin 1A	N-terminus	113-420	synapsin I	XP_013120
synaptotagmin*	C-terminus	265-422	synaptotagmin 1	NP_005630
syntaxin 1A	full length	2-265	syntaxin 1A (brain)	NP_004594
syntaxin 12	N-terminus	2-245	syntaxin 12	XP_039018
syntenin	full length	2-298	syndecan binding protein (syntenin)	XP_044627

*subsequently chosen for incorporation of [γ-^{32}P].

After growing overnight and induction of expression with IPTG, bacteria were harvested by centrifugation. Bacterial pellets were lysed and the clarified supernatant containing the soluble protein extract was incubated with glutathione agarose beads. Bound GST fusion protein was eluted and aliquots of eluates were analysed by SDS-PAGE (*Figure 3*).

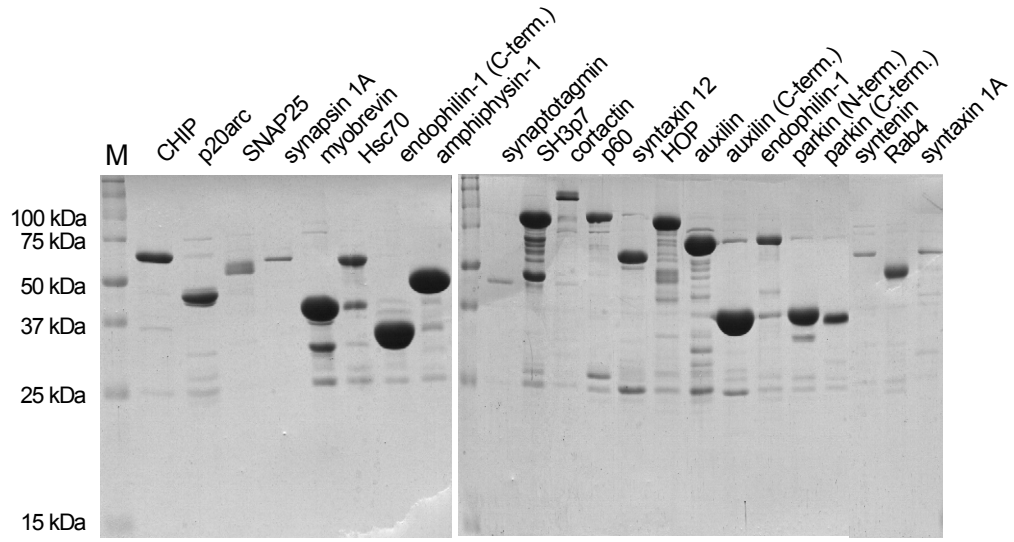


Figure 3. Purification of GST-PKA fusion proteins. Proteins were purified from 1 mL bacterial cultures using 50 µL of a 50 % (v/v) glutathione agarose beads suspension. Aliquots of eluates were analysed by SDS-PAGE (for details see *Table 10*).

Coomassie-stainable bands were evaluated for size, degradation by-products and purity of the GST fusion protein. The protein concentration of purified protein ranged from approximately 1 µg to 25 µg per mL of bacterial culture as determined in a Bradford microtitre assay. For all proteins a band of expected size was observed which was accompanied in most cases by degradation products. Large proteins such as the heat shock organizing protein HOP (*Figure 3*, lane 16) and a C-terminal fragment of auxilin (*Figure 3*, lane 17) exhibited an increased degradation leading to multiple bands down to approximately 26 kDa which represented the stable degradation end product GST.

4.1.3 Phosphorylation of GST fusion proteins

A random subset of ten proteins out of the 22 tested for expression (see indicated proteins in *Table 10*) was chosen to monitor the incorporation of [γ-^{32}P]. Two µg of purified fusion protein were incubated with 10 units of protein kinase A in the presence of 6.5 µCi [γ-^{32}P]ATP. The phosphorylation was stopped after 60 min. by adding TCA to

a final concentration of 10 % (v/v). Phosphorylated proteins were precipitated quantitatively with BSA by a short centrifugation and aliquots of precipitates were analysed by SDS-PAGE. The dried SDS-PAGE gel was placed on a phosphor screen for 30 min. and the screen was then scanned with a phosphor imager (*Figure 4*).

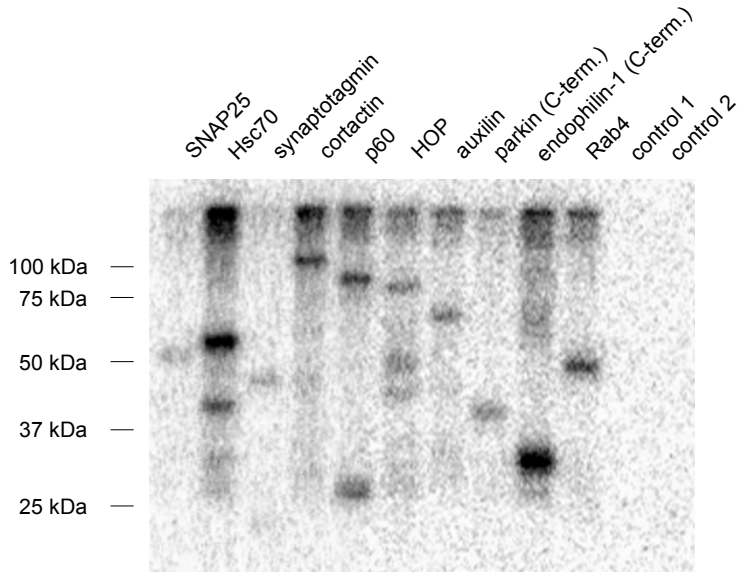


Figure 4. Phosphorylation of GST-PKA fusion proteins. Purified proteins were labeled with [γ-^{32}P]ATP. Aliquots of radiolabeled proteins were separated by SDS-PAGE. The dried gel was subjected to a phosphor screen, which was subsequently scanned by a phosphor imager. GST fusion proteins without a PKA site were used as control: SNAP25 (control 1), synaptotagmin (control 2).

Incorporation of [γ-^{32}P] varied for individual proteins, but in all cases a band of expected size was detected. The typical degradation pattern of Hsc70 (*Figure 4*, lane 3), p60 (*Figure 4*, lane 6) and HOP (*Figure 4*, lane 7) as observed in the SDS-PAGE analysis of purified proteins (*Figure 3*) was also present in the phosphorylated protein samples. A high molecular weight fraction of phosphorylated material did not migrate into the polyacrylamide gel. The signal intensities of these fractions correlated with the signal intensities of expected protein size, suggesting that the phosphorylated proteins aggregated partially, probably due to temperature effects during the phosphorylation reaction and subsequent sample treatment. The efficiency of phopshorylation varied among the tetsted proteins. Similar amounts of auxilin, endophilin-1 (C-term.) and Hsc70 were subjected to phosphorylation (see *Figure 3*) but the signal intensity obtained in the autoradiogramm was considerably lower for auxilin compared to endophilin-1 (C-term.) and Hsc70 (*Figure 4*). To discriminate non-specific phosphorylations of protein on

recognition sites other than the vector-encoded PKA site, two GST fusion proteins without a PKA site were subjected to phosphorylation. A radioactive signal was not observed for GST-SNAP25 (*Figure 4*, control 1) and GST-synaptotagmin (*Figure 4*, control 2), indicating that protein kinase A phosphorylated exclusively the vector-encoded PKA site.

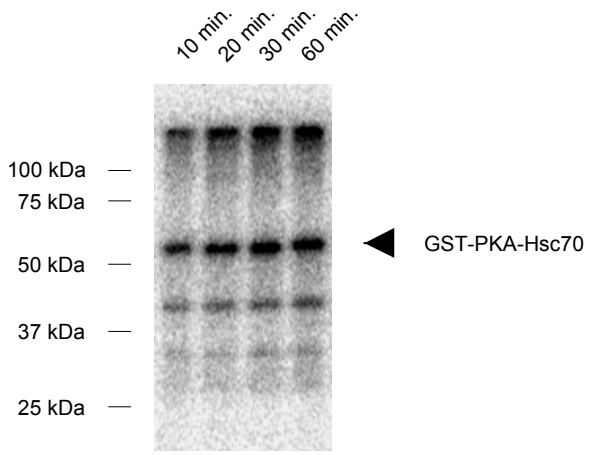


Figure 5. Time course of [γ-^{32}P]-incorporation for Hsc70. Phosphorylation of Hsc70 (aa 373-646) was stopped after 10, 20, 30 and 60 min. by adding 10% TCA. After quantitative precipitation of proteins, aliquots were separated by SDS-PAGE. The dried gel was exposed to a phosphor screen and phosphorylated proteins were visualized by a phosphor imager.

A time course of [γ-^{32}P]-incorporation was monitored for GST-PKA-Hsc70 over a period of 60 min. (*Figure 5*). The phosphorylation was stopped after 10, 20, 30 and 60 min. and aliquots of precipitated samples were analysed. The radioactive signal intensity did not increase significantly after 30 min. which led to the conclusion that a reaction time of 30 min. was suitable for the phosphorylation of GST-PKA-Hsc70.

4.2 Expression and purification of His-tagged proteins

The 75 PCR fragments (see section 4.1.2) were cloned in pQE32N or pQE31N expression vectors to generate His-tagged proteins which could be arrayed on PVDF membranes and tested for protein-protein-interactions with radioactively labeled probe proteins. PCR products with *SalI/NotI* overhangs were directionally cloned into the corresponding *SalI/NotI* sites of pQE32N. Several PCR products contained *BamHI/NotI*

overhangs which were directionally cloned into the corresponding *BamHI/NotI* sites of pQE31N (detailed in *Table 4*, materials section).

E. coli SCS1 cells harbouring the helper plasmid pSE111 were transformed with various pQE32N or pQE31N constructs for the expression of His-tagged proteins. A set of 96 bacterial clones was arrayed into a microtitre plate. Plasmids of 48 clones were sequenced. All clones contained an insert in the correct reading frame (detailed in Appendix A). Protein expression was induced in these clones and expressed proteins were purified with NiNTA agarose beads. Bound His-tagged proteins were eluted with imidazole and eluates were analysed by SDS-PAGE (*Figure 6*). All clones expressed a protein of expected size (detailed in Appendix A).

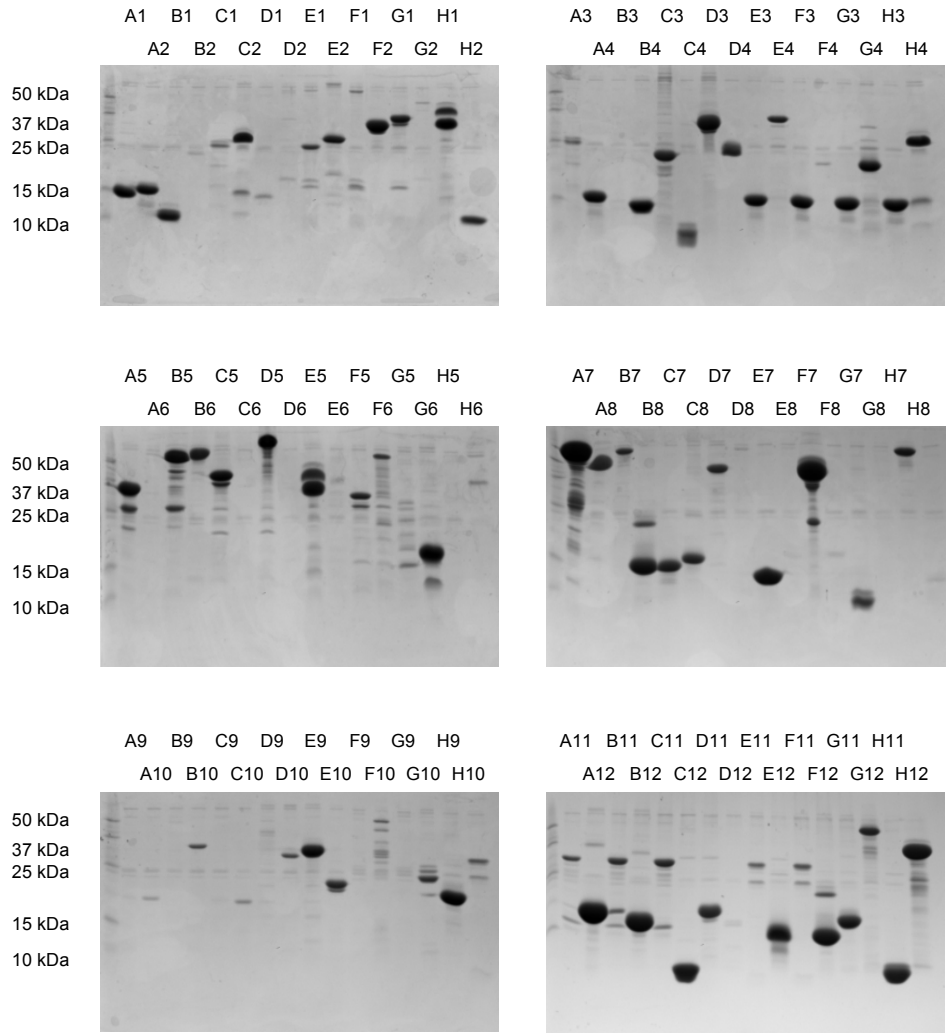


Figure 6. Immobilized metal affinity chromatography of His-tagged proteins. His-tagged proteins were purified from 1 mL bacterial cultures with 50 µL of a 50 % (v/v) NiNTA agarose beads suspension. Aliquots of purified protein were separated by SDS-PAGE. Lanes are numbered according to the microtitre plate positions of clones (for details see Appendix A).

Several clones expressed identical proteins. Seven clones contained a fragment of Hsc70 and expressed a protein of 14 kDa including the His-tag (*Figure 6*; A2, E3-H3, A4, C7). Two clones contained a fragment of filamin A and expressed a protein of 21 kDa accompanied by a prominent degradation product of 14 kDa (*Figure 6*; B8, F12).

All purified proteins were further analysed by MALDI-TOF mass spectrometry. Of the 96 attempted proteins 60 were identified by their MALDI mass finger print. The results obtained from DNA sequencing of 48 clones were confirmed by mass spectrometry in 30 cases. DNA sequences and protein mass finger prints remained unidentified in 18 cases (detailed in Appendix A).

4.3 Development of a protein array screen

Selected GST-PKA fusion proteins (section 4.1.3) were used as radiolabeled probes to overlay a protein array consisting of His-tagged proteins immobilized on a PVDF membrane. The chosen probes represent proteins or protein fragments which have been reported to interact with other proteins also present on the array. His-tagged proteins were immobilized on the membrane applying two different techniques. In the first approach soluble denatured protein extracts were prepared from bacterial cultures and used for spotting. In the second approach, clones were spotted on the PVDF membrane, induced for protein expression and whole cellular proteins were directly released and immobilized on the membrane.

4.3.1 Protein array screens using spotted protein extracts

The 96 bacterial clones, each expressing a His-tagged protein (see section 4.3, *Figure 6*), were grown in a micotitre plate and protein expression was induced with IPTG. To prepare a soluble protein extract for spotting on PVDF membranes, pellets of 1 mL bacterial culture were solubilized in 250 µL SDS loading buffer and cells were lysed at 90°C. After removal of cell debris, clarified supernatant containing denatured soluble protein was used for spotting. Approximately 0.1 µL of each extract was spotted manually on a PVDF filter membrane. The spotting coordinates of extracts were according to the microtitre plate positions of the corresponding clones (see *Figure 6* and Appendix A). The so prepared PVDF filter membrane carrying the protein array was dried and stored at room temperature. To probe the protein array, 4 µg of phosphorylated

GST fusion protein were desalted on a Sephadex G25 column to remove unincorporated [γ-^{32}P]ATP and added to the blocked PVDF filter membrane.

```
2    EQVNELKEKGNKALSVGNIDDALQCYSEAIKLDPHNHVLYSNRSAAYAKKGDYQKAYEDG  61
62   CKTVDLKPDWGKGYSRKAAALEFLNRFEEAKRTYEEGLKHEANNPQLKEGLQNMEARLAE  121
122  RKFMNPFNMPNLYQKLESDPRTRTLLSDPTYRELIEQLRNKPSDLGTKLQDPRIMTTLSV  181
182  LLGVDLGSMDEEEE IATPPPPPPPKK ETKPEPMEEDLPENKKQALKEKELGNDAYKKKDF  241
242  DTALKHYDKAKELDPTNMTYITNQAAVYFEKGDYNKCRELCEKAIEVGRENREDYRQIAK  301
302  AYARIGNSYFKEEKYKDAIHFYNKSLAEHRTPDVLKKCQQAEKILKEQERLAYINPDLAL  361
362  EEKNKGNECFQKGDYPQAMKHYTEAIKRNPKDAKLYSNRAACYTKLLEFQLALKDCEECI  421
422  QLEPTFIKGYTRKAAALEAMKDYTKAMDVYQKALDLDSSCKEAADGYQRCMMAQYNRHDS  481
482  PEDVKRRAMADPEVQQIMSDPAMRLILEQMQKDPQALSEHLKNPVIAQKIQKLMDVGLIA  541
542  IR
```

Figure 7. Amino acid sequence of HOP. The clone contains the full length amino acid sequence of heat shock organizing protein HOP (GenBank accession NP_006810, aa 2-543). The tetratricopeptide repeat (TPR) domains are underlined. TPR1, N-terminal TPR domain (aa 4-104); TPR2A, first central TPR domain (aa 225-333); TPR2B, second central TPR domain (aa 360-467). A proline-rich amino acid stretch is depicted in bold letters (aa 195-206).

Phosphorylated GST-PKA fusion proteins of Hsc70 and HOP were used as probes to screen the protein array for their interaction partners. The interaction of Hsc70 with HOP has been reported before and it has been shown that HOP functions as an adaptor protein for Hsc70 and the 90 kDa heat shock protein Hsp90 to form a multichaperone complex (Smith *et al*., 1993; Chen *et al*., 1996). HOP has three tetratricopeptide repeat (TPR) domains (*Figure 7*, underlined) and it interacts with Hsc70 via its N-terminal TPR domain TPR1. The binding involves electrostatic interactions between charged amino acid residues in the TPR1 domain of HOP and the C-terminal EEVD motif of Hsc70 (*Figure 8*, underlined) as well as hydrophobic interactions between residues in the TPR1 domain with isoleucine upstream of the EEVD motif (Scheufler *et al*., 2000; Brinker *et al*., 2002).

```
541  SLESYAFNMKATVEDEKLQGKINDEDKQKILDKCNEIINWLDKNQTAEKEEFEHQQKELE  600
601  KVCNPIITKLYQSAGGMPGGMPGGFPGGGAPPSGGASSGPTIEEVD  646
```

Figure 8. Partial amino acid sequence of Hsc70. The clones detected on the array with GST-PKA-HOP as a probe contain a C-terminal fragment of Hsc70 (GenBank accession NP_006588, aa 541-646). The last five amino acids (underlined) mediate the interaction with the TPR1 domain of HOP.

A spot with high signal intensity was detected with GST-PKA-Hsc70 (aa 373-646) as a probe (*Figure 9A*, spot A7). The spot contained full length heat shock organizing protein His-HOP (aa 2-543) indicating that the C-terminal fragment of Hsc70 recognized its interaction partner HOP on the array. Reversly, full length GST-PKA-HOP detected all but one C-terminal fragments of His-Hsc70 spotted on the protein array (*Figure 9B*, spots D3-H3, A4, C7).

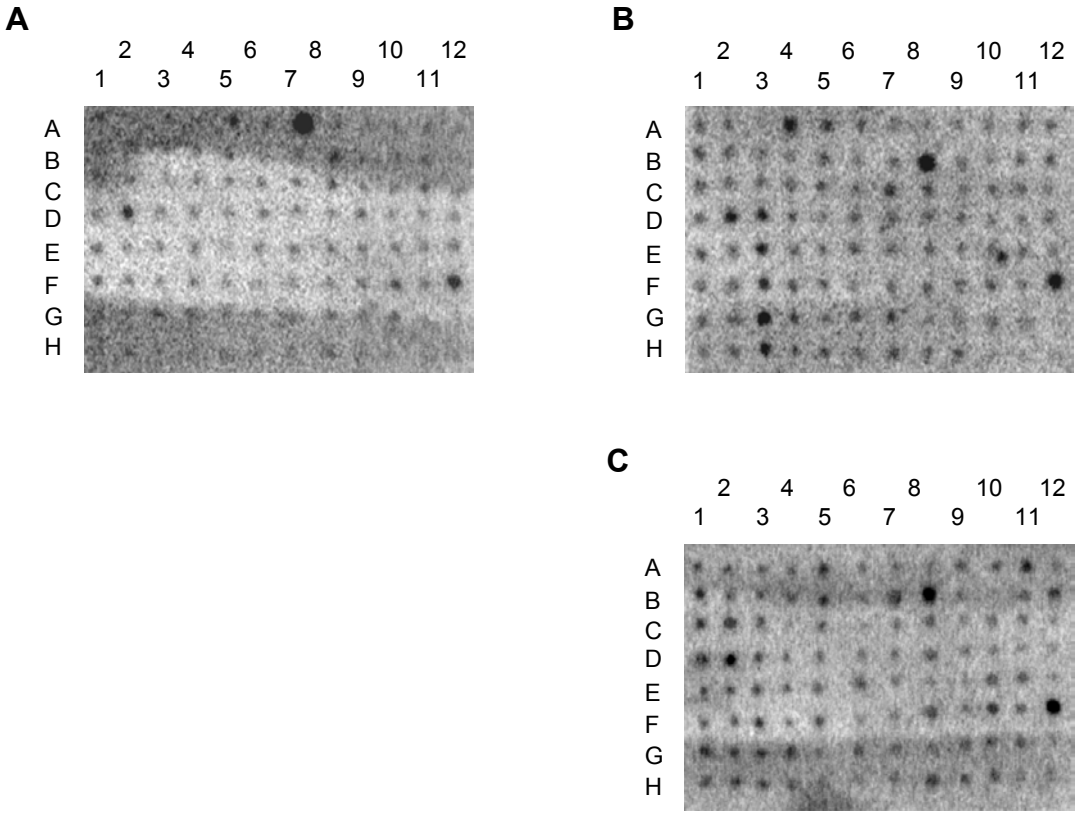


Figure 9. Filter overlays with spotted soluble protein extracts. Bacterial pellets of 1 mL cultures were lysed in 250 µL SDS loading buffer at 90°C and 0.1 µL aliquots were spotted manually on PVDF membranes. For a detailed description of clones according to the array positions see Appendix A. A) Overlay with radiolabeled GST-PKA-Hsc70 (aa 373-646). B) Overlay with radiolabeled GST-PKA-HOP (aa 2-543). C) Control overlay with a phosphorylation mix containing no radiolabeled protein probe.

A phosphorylation reaction that contained no recombinant protein has been used as a negative control (*Figure 9C*). Two spots (B8, F12) were detected which have also been detected on arrays overlayed with GST-PKA-Hsc70 and GST-PKA-HOP. The corresponding clones both express a fragment of filamin A (aa 2015-2198). Even though three other spots with fragments of filamin A (aa 2041-2136, aa 1644-2136, aa 1788-2136) were present on the array, they were not detected in the negative control overlay.

This observation suggests that the amino acid residues 2136-2198 of filamin A must be responsible for the signals at B8 and F12. Sequence analysis revealed that filamin A possesses a the amino acid Ser2152 which is part of the sequence RRRAPSV. This sequence is highly similar to the recognition motif RRASV of protein kinase A.

4.3.2 Protein array screens using spotted clones

To improve the signal-to-noise ratio on PVDF membranes overlayed with radiolabeled GST fusion protein, a protein array was generated at the German Resource Center RZPD by a robot. Each of the 96 clones (see section 4.2.1) was spotted in duplicate on a PVDF filter membrane. Thus each clone was represented on the array by one pair of spots. The spotting coordinates of clones were according to their microtitre plate positions (see *Figure 6* and Appendix A). Following growth overnight and induction of protein expression with IPTG, whole cellular proteins were released and immobilized directly on the PVDF filter membrane under denaturing alkaline conditions.

The so prepared protein array was overlayed with radiolabeled GST-PKA-HOP, GST-PKA-SH3E1 and GST-PKA (*Figure 10*, performed by H. Weiner). In accordance with results obtained from the protein array screen using spotted extracts, GST-PKA-HOP detected its interaction partner Hsc70 on the array (*Figure 10A*, spots D3-H3, A4, C7). The signal-to-noise ratio was drastically increased compared to that obtained from arrays with spotted extracts. All pairs of spots had similar signal intensities which demonstrated the robustness of the assay. The signal intensity of duplicate D3 was significantly lower, even though the expression strength of clone D3 was slightly higher (see *Figure 6*, lane D3) than that of the other clones.

GST-PKA-HOP additionally detected a pair of spots representing a 16 kDa N-terminal fragment of protein-tyrosine kinase fyn (*Figure 10A*, spot G11). A src homology 3 (SH3) domain resides in the N-terminal fragment of fyn which binds to proline-rich motifs of its interacting proteins. HOP contains such a proline-rich motif spanning the amino acid sequence 195-206 (see *Figure 7*, bold letters) which probably has been recognized by the SH3 domain of fyn. A 9 kDa fragment of fyn (aa 82-142) at position H11 (see figure 6, lane H11) was not detected with GST-PKA-HOP as a probe.

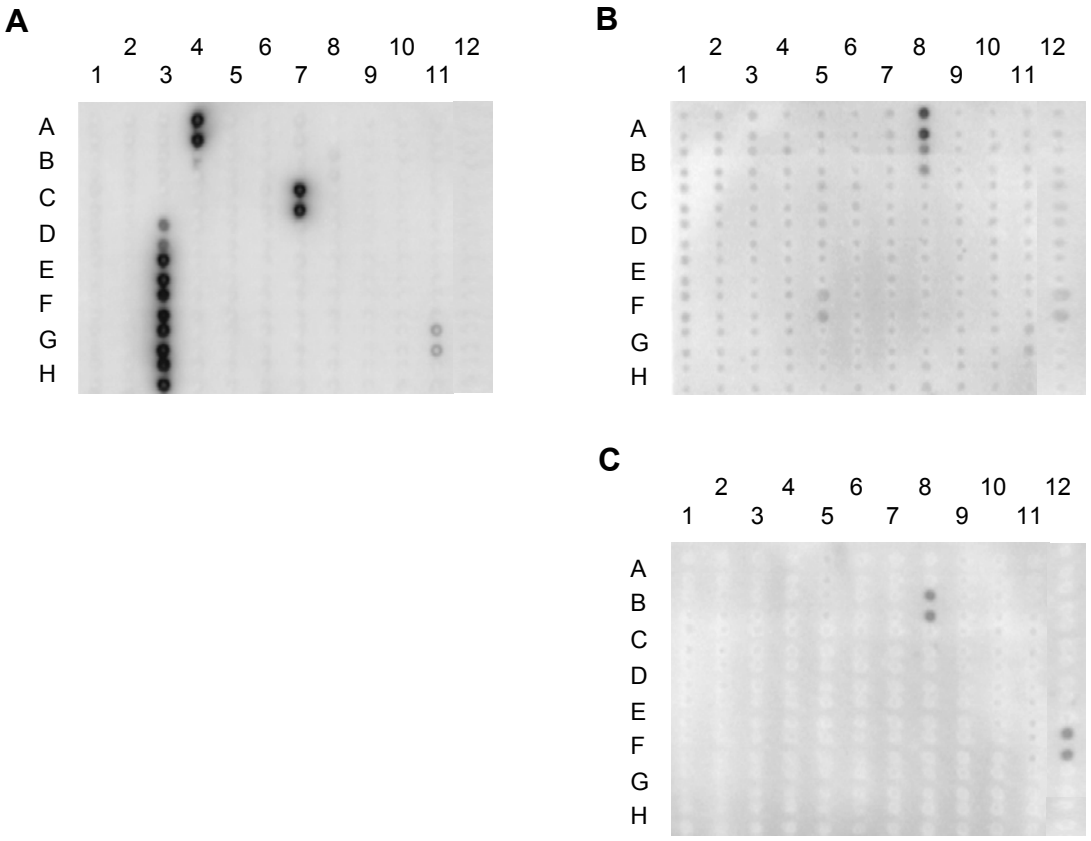


Figure 10. Filter overlays with spotted clones. Clones were spotted on PVDF filter membranes in duplicates. After overnight growth and induction of protein expression bacteria were lysed by alkali treatment. For a detailed description of clones according to the array positions see Appendix A. A) Overlay with radiolabeled GST-PKA-HOP (aa 2-543). B) Overlay with radiolabeled GST-PKA-SH3E1 (aa 292-352). C) Overlay with radiolabeled GST-PKA as a control.

A C-terminal fragment coding for the SH3 domain of endophilin-1 (aa 292-352) was used as a radiolabeled probe, termed GST-PKA-SH3E1, to screen the protein array. A spot with high signal intensity was detected with GST-PKA-SH3E1 (*Figure 10B*, spot A8). The duplicate clones expressed a central fragment of amphiphysin-1 (aa 291-559). It has been reported that the SH3 domain of endophilin-1 interacts with amphiphysin-1 via a proline-rich motif spanning the amino acid sequence 298-309 of amphiphysin-1 (Micheva *et al*., 1997). The proline-rich motif of the central fragment of amphiphysin-1 is shown (*Figure 11*, underlined). It can be concluded that this motif was recognized by GST-PKA-SH3E1 and mediated binding of the SH3 domain of endophilin-1.

```
291  PRSPSQTRKGPPVPPLPKVTPTKELQQENIISFFEDNFVPEISVTTPSQNEVPEVKKEET  350
351  LLDLDFDPFKPEVTPAGSAGVTHSPMSQTLPWDLWTTSTDLVQPASGGSFNGFTQPQDTS  410
411  LFTMQTDQSMICNLAESEQAPPTEPKAEEPLAAVTPAVGLDLGMDTRAEEPVEEAVIIPG  470
471  ADADAAVGTLVSAAEGAPGEEAEAEKATVPAGEGVSLEEAKIGTETTEGAESAQPEAEEL  530
531  EATVPQEKVIPSVVIEPASNHEEEGENEI  559
```

Figure 11. Partial amino acid sequence of amphiphysin-1. The clone contains the central amino acid sequence of amphiphysin-1 (GenBank accession NP_001626, aa 291-559). The proline-rich motif which mediates binding to the SH3 domain of endophilin-1 is underlined.

Radiolabeled GST-PKA was used as a control in a filter overlay (*Figure 10C*). Signals of two duplicate clones (*Figure 10C*, spots B8, F12) were detected as already observed for a control overlay in the absence of radiolabeled GST fusion protein (see *Figure 9C*, spots B8, F12). No other signals were observed in the control filter overlay that could be deduced to a binding of GST-PKA.

4.4 High-density protein array screens with the SH3 domain of endophilin-1 and heat shock protein p60

As demonstrated on a small scale, the method of a filter overlay screen with a radiolabeled probe protein was successful for the detection of interaction partners of GST-PKA-Hsc70, GST-PKA-HOP and GST-PKA-SH3E1. The signal-to-noise ratio obtained from filters with spotted clones was superior compared to the signal-to-noise ratio obtained from filters with spotted extracts. In a next step, the screening method using spotted clones was extended to the large scale. To find novel protein-protein-interactions for the SH3 domain of endophilin-1 on a genome-wide scale, high-density protein arrays were screened.

4.4.1 Spotting of clones and high-density protein filter overlays

High-density protein arrays were obtained from the German Resource Center RZPD. A total of 36,864 individual clones of a human fetal brain cDNA library enriched for expression of His-tagged proteins were gridded on two PVDF membranes. Individual clones were spotted from 384-well microtitre plates by a robot as duplicates in blocks with a 5 x 5 pattern along with a guide spot in the center of each block (*Figure 12*). The spotting pattern allowed assignment of duplicate clones to their corresponding microtitre plates and well positions.

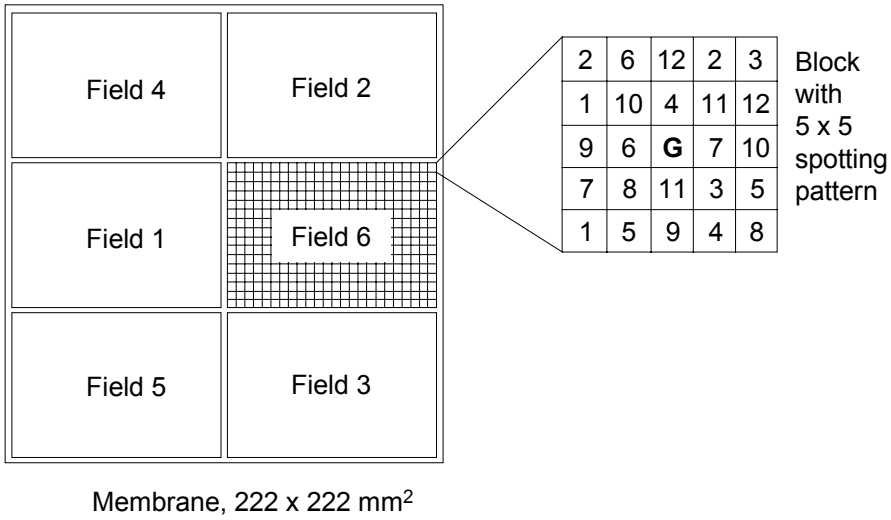


Figure 12. Spotting pattern of a high-density protein array on PVDF filter membranes. Clones were spotted from 384-well microtitre plates on PVDF membranes. The spotting pattern permits an assignment of individual clones to their corresponding microplate and well position. Each clone is spotted twice on the membrane within a 5 x 5 spotting pattern with a guide spot (G) in the center of each block. Clones from 12 individual 384-well microplates are spotted in each field (taken from Weiner *et al*., 2004).

The radiolabeled SH3 domain of endophilin-1 (GST-PKA-SH3E1) was used to screen the so prepared high-density protein array for novel protein-protein-interactions (*Figure 13* and *14*; performed by H. Weiner). A full length heat shock protein of 60 kDA (GST-PKA-p60) served as a control to distinguish false positive signals. Approximately 500 µg of radiolabeled GST fusion protein were added to the blocked PVDF filter membranes and incubation was done overnight at 4°C. Filter membranes were washed and the dried membranes were placed on phosphor screens for 30 min. The phopshor screens were then scanned with a phosphor imager to obtain a permanent record of radioactive signals.

4.4.2 Image analysis

Radioactive signals were analysed with the software Xdigitise which can be used to score positive clones on the protein array and to retrieve their microtitre plate positions. A total of 401 duplicate signals were detected on the two PVDF filter membranes probed with GST-PKA-SH3E1 (*Figure 13*, first PVDF filter membrane of

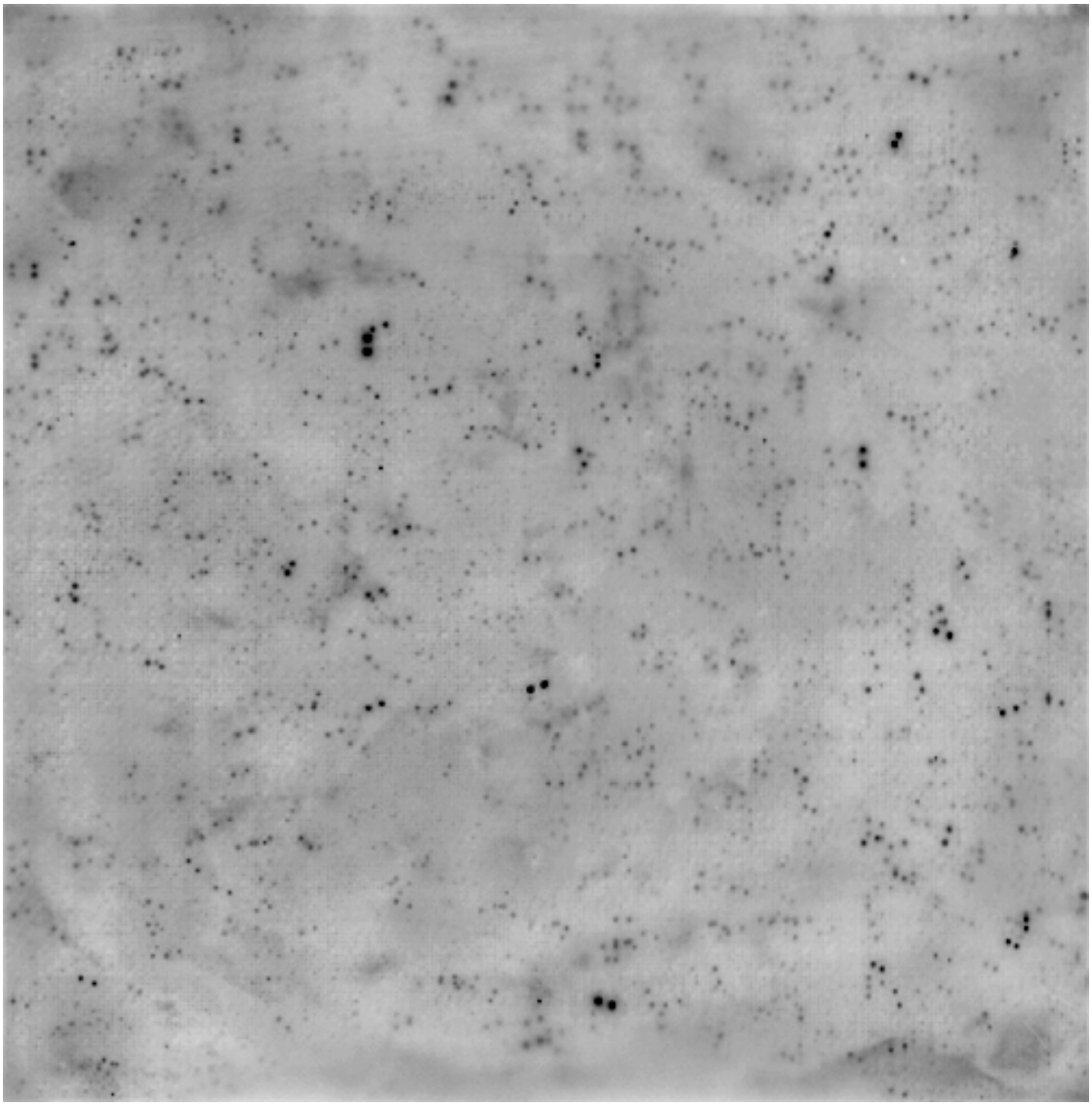


Figure 13. High-density protein filter overlay with GST-PKA-SH3E1. 36,848 individual DNA expression clones were gridded on two PVDF filter membranes. After denaturing release of cellular proteins the filters were probed with purified, radiolabeled GST-PKA-SH3E1. An image of the first PVDF filter membrane is depicted.

the array). Only duplicate signals of equal intensity which corresponded to the same clone were considered to be positives. It has been generally observed that pairs of spots corresponding to the same clone have an equally strong signal intensity.

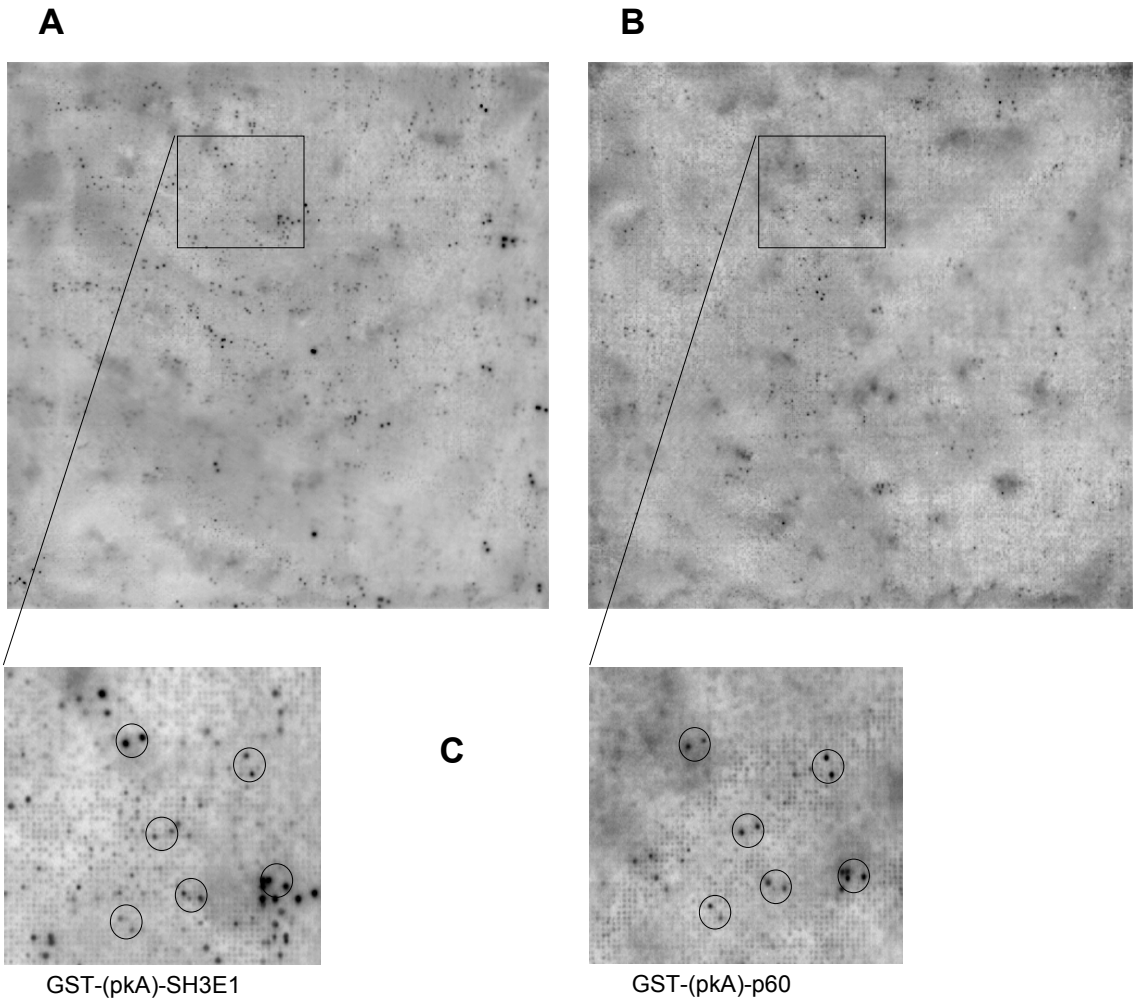


Figure 14. Comparison of high-density protein filters probed with GST-PKA-SH3E1 and GST-PKA-p60. A) Image of the second PVDF filter membrane probed with GST-PKA-SH3E1. B) Image of the second PVDF filter membrane probed with GST-PKA-p60. C) Magnified sections of the filters corresponding to each other.

To distinguish false positive signals the protein array overlayed with GST-PKA-SH3E1 has been compared to a protein array overlayed with GST-PKA-p60 (*Figure 14A* and *14B*, second PVDF filter membrane of the array). A total of 89 duplicate signals generated with GST-PKA-SH3E1 as a probe were determined to be false positives since they had also been detected with GST-PKA-p60 as a probe. Such false positives are clearly distinguishable as demonstrated in the magnified sections of two corresponding PVDF filter areas (*Figure 14C*).

4.4.3 Identification of clones positive with GST-PKA-SH3E1

The microtitre plate positions of all 401 clones positive with GST-PKA-SH3E1 were determined and clones were ordered at the German Resource Center RZPD for a detailed analysis. Clones were arrayed in 96-well microtitre plates to purify all plasmids on a minipreparation-scale. Plasmids were partially sequenced with the vector primer pQE65 and DNA sequences were matched against the ENSEMBL database to assign an ENSEMBL transcript number to each DNA sequence. A total of 184 unique transcripts were determined which were exclusively positive for GST-PKA-SH3E1 as a probe. Thirty-two percent of the transcripts coded for a protein inframe with the vector-encoded His-tag. Twenty-nine percent of the transcripts had a frameshift of one amino acid and thirty-nine of the transcripts had a frameshift of two amino acids relative to the His-tag.

Even though numerous positive clones code for out-of-frame translations, the resulting artificial polypeptides were selected by the SH3 domain of endophilin-1. Most strikingly, two cDNA´s were detected multiple times. Obviously, both cDNA species were present in the library at a high abundance. The first cDNA contained the ORF of myristylated alanine-rich protein kinase C substrate (*MARCKS*) with a partial stretch of the 5´-UTR and was detected 84 times. The second cDNA contained the ORF of creatine kinase B (*CKB*) also including a stretch of the 5´-UTR and was detected 40 times.

An out-of-frame translation in clones containing the cDNA of *MARCKS*, lead to a stop codon and an ORF coding for an artificial polypeptide (detailed in Appendix B). The calculated molecular weight of 10,057 Dalton for the artificial polypeptide including the His-tag was in good agreement with the molecular weight of 11 kDa estimated by SDS-PAGE analysis. For half of the clones a band of 11 kDa was detected by SDS-PAGE analysis (Figure 15A). For the rest of clones harbouring a cDNA of *MARCKS* no coomassie-stainable bands were visible. The artificial polypeptide contained two proline-rich sequences PPPRAPAPPLRRPGRP and PPSRPPSWAARAPRLP one of which is supposed to mediate binding to GST-PKA-SH3E1.

SDS-PAGE analysis of purified protein from clones containing the cDNA of *CKB* revealed several bands up to 48 kDa (*Figure 15B*). DNA sequencing turned out that all clones possessed the full ORF of *CKB* along with a partial stretch of the 5´-UTR. According to the obtained DNA sequence, insertion of the 5´-UTR between the vector sequence and the ORF lead to a frameshift translation of a truncated polypeptide with a calculated molecular weight of 16,633 Dalton (detailed in Appendix C). A band of

approximately 17 kDa estimated from SDS-PAGE analysis agreed well with the calculated molecular weight of the artificial polypeptide. The 5′-UTR is GC-rich and codes preferentially for the amino acids proline, arginine and alanine. An amino acid stretch translated from the 5′-UTR resembles the proline-rich sequence RSAPAPARPPARRR which was supposed to be responsible for the binding of GST-PKA-SH3E1. Mass spectrometric analysis of the 48 kDa band was in contrast to results obtained from DNA sequencing. Mass fingerprints of 48 kDa expression products of three individual clones matched 11 peptides (clone E19575), 16 peptides (clone J11512) or 10 peptides (clone M02542) of the protein CKB (SwissProt accession P12277). The molecular weight of full length CKB expressed inframe with the vector-encoded His-tag was predicted to be 47.7 kDa which agreed well with a molecular weight of 48 kDa estimated from SDS-PAGE analysis. Therefore, the results obtained from mass spectrometric analysis support the findings from SDS-PAGE analysis. However, the proline-rich sequence encoded by the 5′-UTR must also be included in the 48 kDa protein.

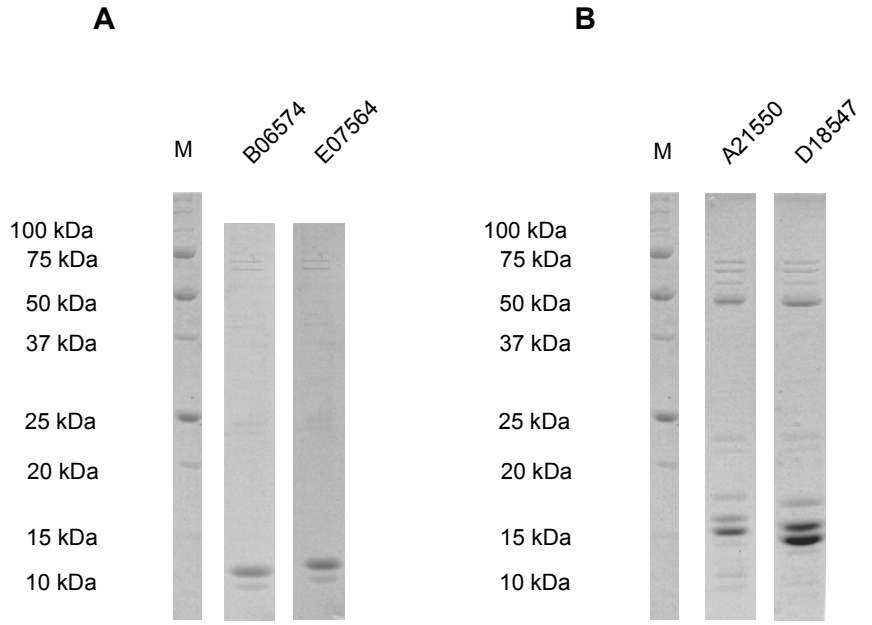


Figure 15. SDS-PAGE analysis of purified proteins from clones harbouring cDNAs for CKB and MARCKS. His-tagged proteins were purified from 1 mL bacterial cultures. Aliquots of purified proteins were separated by SDS-PAGE. M, protein marker. Lanes are designated according to the unique clone identifiers. A) Purified proteins from two clones harbouring a cDNA for MARCKS. B) Purified proteins from two clones harbouring a cDNA for CKB.

DNA sequencing recovered 43 clones harbouring ORF´s inframe with the vector-encoded His-tag (detailed in appendix D), among them 8 clones which in addition contained a 5´-UTR. Three clones contained only the 3´-UTR of a gene. The DNA sequence of one clone coded for a C-terminal fragment of (ALG-2)-interacting protein 1, also named Alix (clone O01523). It has been reported recently, that Alix, a 868-amino acid long cytoplasmic protein involved in apoptosis, interacts with endophilins (Chatellard-Causse *et al*., 2002).

4.4.4 Purification of expression products from positive clones

The solubility and yield of proteins found to bind GST-SH3E1 on high-density protein arrays is a prerequisite for a suspension binding assay, like an *in vitro* pulldown. To examine the solubility and yield of proteins expressed from clones positive for GST-PKA-SH3E1, a subset of 96 clones enriched for inframe expression of His-tagged protein was selected and proteins were purified with NiNTA agarose beads from 1 ml bacterial cultures. Purified expression products were analysed for yield, purity, size and proteolytic degradation by SDS-PAGE (*Figure 16*). Of the attempted 96 bacterially expressed proteins, 82 achieved purification levels detectable in Coomassie-stained polyacrylamide gels. Fourteen proteins were not detected in coomassie-stained polyacrylamide gels. For several proteins multiple proteolytic fragments were visible, indicating that the recombinant protein has been degraded step-by-step to a stable end product or that the transcription/translation was incomplete.

The protein content of purified samples was determined and used to calculate the yield of His-tagged protein per mL of bacterial culture (*Figure 17A*). Fifty percent of the clones expressed 1-10 µg pure recombinant protein per mL of bacterial culture. Seventeen percent of clones contained no purified protein and 33 % of clones expressed more than10 µg pure recombinant protein per mL bacterial culture. The size of the largest protein band detectable in each SDS-PAGE lane was estimated (*Figure 17B*). Fourty-two percent of purified proteins exhibited a size of 20-40 kDa, 32 % were smaller than 20 kDa and 26 % had a size above 40 kDa.

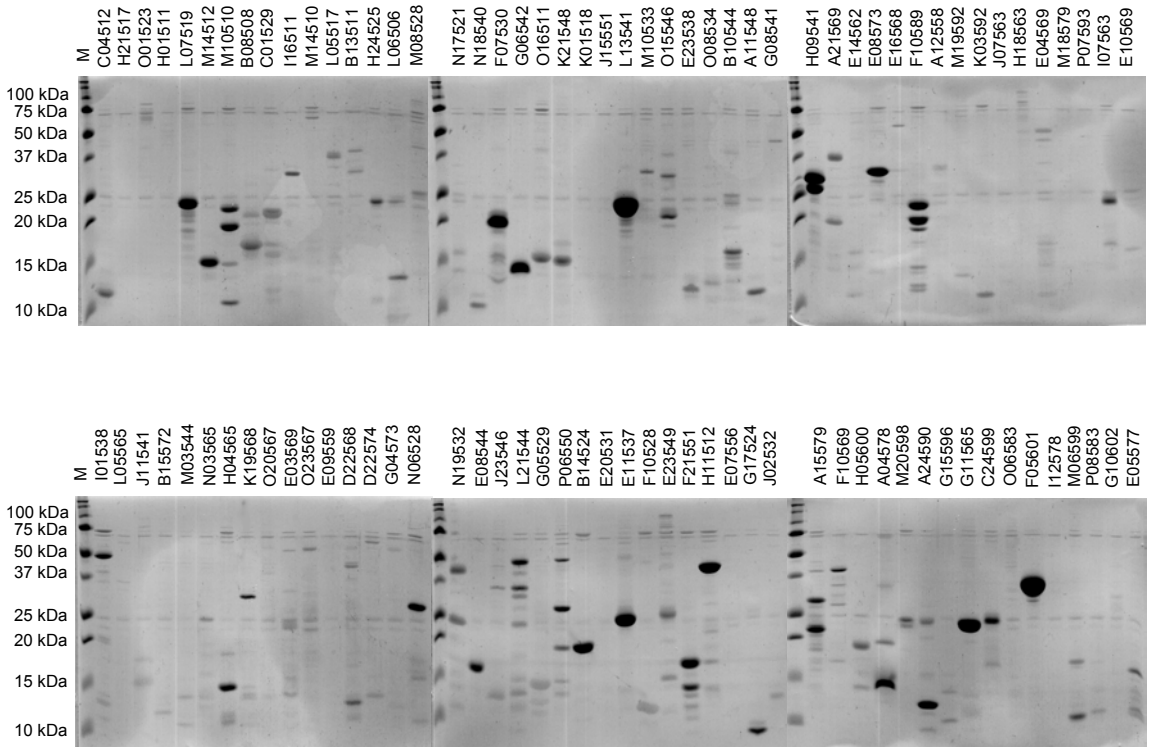


Figure 16. Immobilized metal affinity chromatography of putative SH3-binding proteins. A subset of 96 bacterial cDNA clones enriched for inframe expression of His-tagged protein were cultured in deep well microtitre plates. Soluble proteins were purified from cleared bacterial lysates on NiNTA agarose and eluted with 250 mM imidazole. Aliquots of eluate were separated on SDS-PAGE gels.

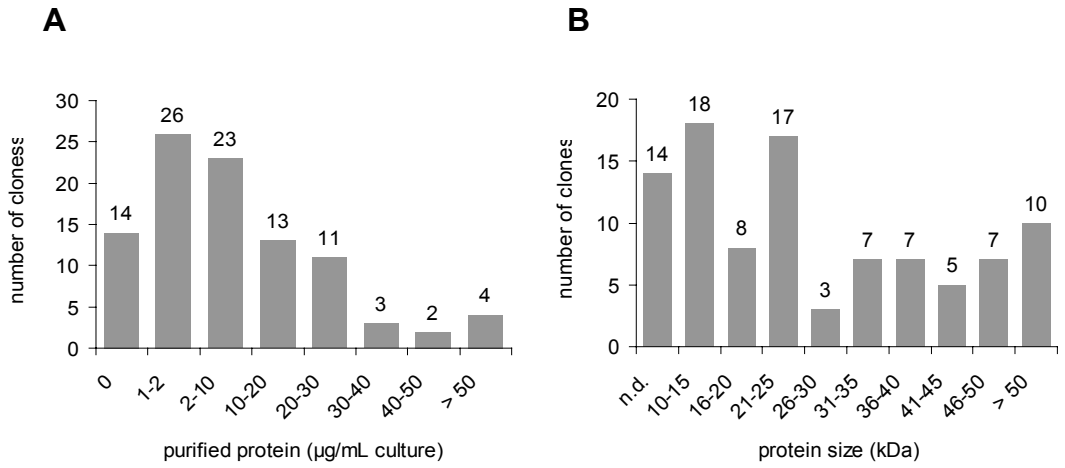


Figure 17. Immobilized metal affinity chromatography of 96 clones positive for GST-PKA-SH3E1. His-tagged proteins were purified from 1 mL bacterial cultures. Protein concentrations were determined with a Bradford mini-assay and the sizes of proteins were estimated from SDS polyacrylamide gels (see *Figure 16*). A) Distribution of protein yields. B) Size distribution of purified proteins. n.d., not detected in coomassie-stainable amounts.

4.5 Confirmation of endophilin-1 interactions

4.5.1 Pulldown assay with glutathione agarose beads

A pulldown assay was developed to independently confirm or disprove the binding of the SH3 domain of endophilin-1 to His-tagged proteins identified on the high-density protein array. The SH3 domain of endophilin-1 (aa 292-352) was expressed as a GST fusion protein (GST-SH3E1) in *E. coli* using the expression vector pGEX-6P-2. GST-SH3E1 was immobilized on glutathione agarose beads and served as a bait to capture His-tagged protein purified from clones positive for GST-PKA-SH3E1. GST served as a negative control. To efficiently pull down His-tagged protein, glutathione agarose beads had to be saturated with GST fusion protein. Since a large fraction of clones expressed a His-tagged protein of 20-40 kDa and GST (26 kDa) or GST-SH3E1 (34 kDa) were of similar size, it was likely that the GST fusion proteins masked the bands of co-purifying His-tagged proteins in a subsequent SDS-PAGE analysis. Covalent crosslinking prevents the GST fusion protein from elution together with the His-tagged protein. Thus, crosslinking should facilitate the analysis of co-purifying proteins by SDS-PAGE without interfering prominent bands of the GST fusion proteins.

To determine the binding capacity of glutathione agarose, different amounts of GST-SH3E1 were applied to a fixed volume of glutathione agarose beads (*Figure 18A*). The amount of bound GST-SH3E1 was determined in a Bradford assay and the concentration of GST-SH3E1 was calculated per µL of settled glutathione agarose beads. The binding capacity of glutathione agarose was 4.5 µg GST-SH3E1 per µL of agarose beads. SDS-PAGE analysis of eluates was consistent with that result (*Figure 18B*).

GST fusion proteins were covalently bound to glutathione agarose using di-methyl-pimelimidate (DMP) as a crosslinking reagent. Glutathione agarose beads were saturated with GST-SH3E1 or GST prior to crosslinking with DMP (*Figure 19A*). After crosslinking and removal of non-covalently bound GST fusion protein with reduced glutathione, the efficiency of crosslinking was examined by incubating aliquots of beads repeatedly at 70°C in SDS loading buffer (*Figure 19B*, lane 1 and 2). A small fraction of GST fusion protein eluted from the beads along with a fraction of higher molecular weight. This may be due to a leaching of crosslinked GST fusion protein (*Figure 19A*).

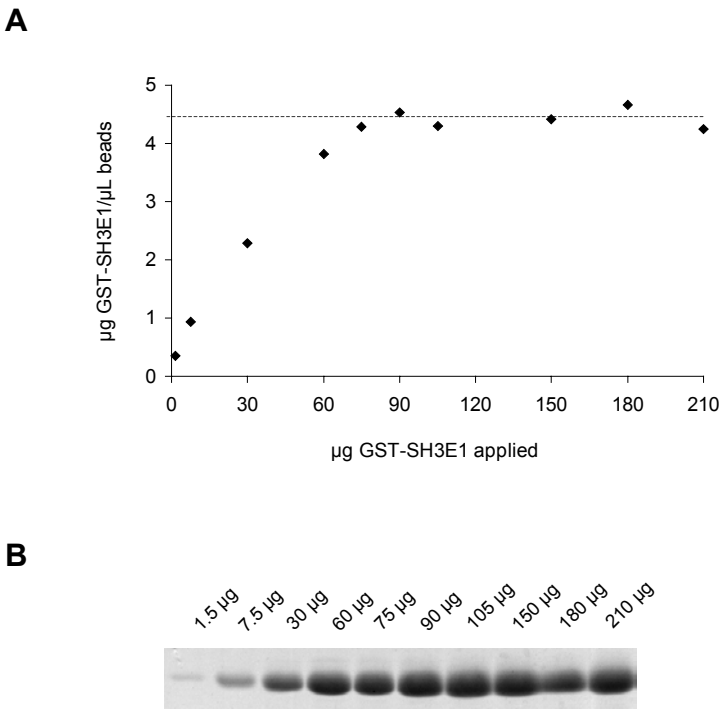


Figure 18. Binding capacity of glutathione agarose. 20 µL of glutathione agarose were incubated in a volume of 1400 µL containing different amounts of GST-SH3E1. A) The concentration of GST-SH3E1 was determined with a Bradford assay and plotted against the amount of GST-SH3E1 applied for purification. B) Aliquots of a 50 % (v/v) beads slurry were analysed by SDS-PAGE.

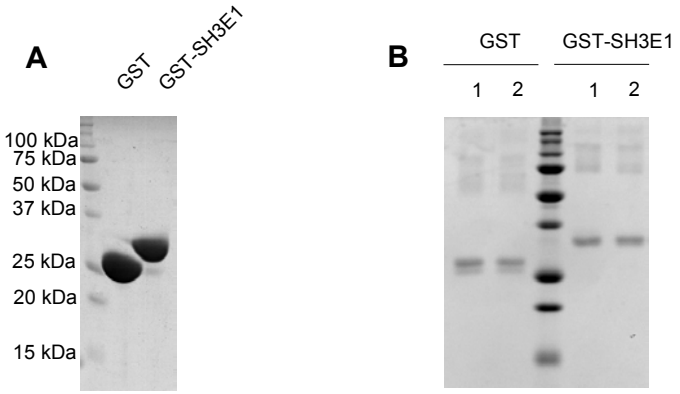


Figure 19. Crosslinking of GST fusion proteins to glutathione agarose beads. A) Aliquots of glutathione agarose beads with purified GST-SH3E1 or GST before crosslinking. B) Aliquots of glutathione agarose beads with purified GST-SH3E1 or GST after crosslinking. (1) incubation of beads in SDS loading buffer at 70°C. (2) repeated incubation of beads in SDS loading buffer at 70°C.

Sixteen clones that express soluble protein and gave a signal in the high-density protein array screen were selected to assay their expression products for interaction with GST-SH3E1 crosslinked to glutathione agarose beads (detailed in *Table 11*). The set of clones included proteins which were translated inframe with the His-tag as well as out-of-frame translated polypeptides and expression products of false positive clones. Eluates of co-purifying proteins were separated by SDS-PAGE (*Figure 20A*). Almost all co-purifying His-tagged proteins were detected in equal coomassie-stainable amounts with GST-SH3E1 as well as with GST as a bait. Therefore, it was not possible to distinguish a specific binding of the SH3 domain of endophilin-1, except for clone C01529. A His-tagged protein of 20 kDa was detected with GST-SH3E1 but the 20 kDa protein was absent with GST as the bait.

Table 11. Pulldown assay with selected clones.

clone	ENSEMBL transcript	encoded protein	note	PD*
B08508	ENST00000322981	hepatocellular carcinoma-associated antigen (HCA25a)		+
B13511	ENST00000224864	CUE domain containing protein 2		-
C01529	ENST00000322981	hepatocellular carcinoma-associated antigen (HCA25a)		+
F07530	ENST00000323233	cysteine and glycine-rich protein 2 (CRP2)		+
G06542	ENST0000219548	STIP1 homology and U-Box containing protein 1	frameshift	-
H24525	ENST0000258415	cytochrome P450/27	frameshift	+
I01538	ENST00000269209	hypothetical protein FLJ21610 (C18ORF11)		+
K19568	ENST0000158302	actin, cytoplasmic 1 (ß-actin)	false positive	-
K21548	ENST00000268989	RUN and TBC1 domain containing protein 1		+
L05577	ENST00000313370	zinc finger protein 106		-
L07519	ENST00000262868	disks large-associated protein 4 (DAP4)		+
L13541	ENST0000013034	nucleoside diphosphate kinase A (NDK)	false positive	-
M06599	ENST0000308639	nuclear factor NF-kappa-B p65 subunit	frameshift	+
M10510	ENST0000294599	MEGF6 (fragment)	frameshift	+
N06528	ENST00000300131	NGFI-A binding protein 2 (EGR1 binding protein 2)		-
O16511	ENST0000278572	40S ribosomal protein S3	false positive	-

[a]RZPD clone IDs without prefix (MPMGp800).

*PD, pulldown. (+) pulldown positive, (-) pulldown negative.

To further analyse co-purifying proteins, aliquots of eluates were subjected to a Western blot and His-tagged proteins were detected with anti-RGS-His antibody (*Figure 20B*). The observations made in the SDS-PAGE analysis were confirmed in Western blots. Except for clone C01529 and to a lesser extent for clones B08508 and K21548, the antibody detected equal amounts of His-tagged protein for GST-SH3E1 and GST. A possible explanation was that most of the His-tagged proteins aggregated during purifica-

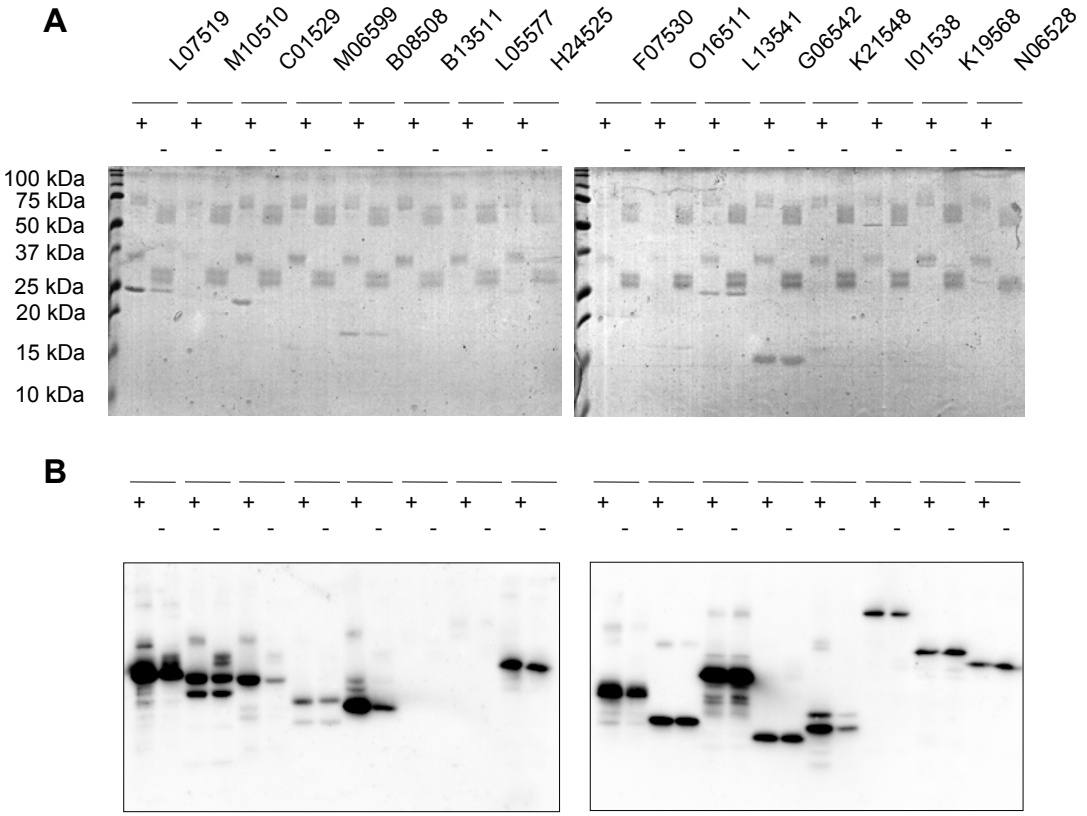


Figure 20. Pulldown assay with glutathione agarose beads. GST-SH3E1 (+) or GST (-) was crosslinked to glutathione agarose and incubated with His-tagged proteins. A) SDS-PAGE analysis of eluates with co-purified His-tagged protein. B) Western blot analysis of co-purified His-tagged proteins with anti-RGS-His antibody.

tion and subsequent incubation with glutathione agarose beads. Since a large fraction of the formed aggregates might be pulled down together with the agarose beads they could not be removed from specifically binding protein by washing. Consequently, such aggregates might also be present in the negative control with GST. Even though for clones C01529, B08508 and K21548 it was possible to discriminate the protein amounts pulled down with GST-SH3E1 from those pulled down with GST, the overall results of the developed pulldown assay were unsatisfactory.

4.5.2 Pulldown assay with NiNTA agarose beads

To circumvent the problems of the initial pulldown experiment, a reverse approach was developed. In this approach His-tagged protein was immobilized on NiNTA agarose beads as a bait to pull down GST-SH3E1. A suitable additive was needed, to prevent proteins from possible aggregation after purification and subsequent steps. Bovine serum albumine (BSA) was found to solve the problem when it was added

to purified proteins at a final concentration of 0.1 % (w/v). Incubation of NiNTA-immobilized protein with GST-SH3E1 or GST and washing of beads was also accomplished in the presence of BSA. Beads were then washed repeatedly without BSA to reduce the concentration of BSA to an acceptable level for SDS-PAGE analysis.

The set of 16 clones already tested in pulldowns with glutathione agarose (see *Table 11* and *Figure 20*) was used and His-tagged proteins were immobilized on NiNTA agarose beads. Incubation with GST-SH3E1 or GST was done in the presence of BSA (*Figure 21A*). Eluates of co-purifying proteins were separated by SDS-PAGE (*Figure 21B*). An additional band of the captured GST-SH3E1 was observed in nine SDS-PAGE lanes, whereas no coomassie-stainable GST was observed in the corresponding control lanes. The results were confirmed in Western blots with anti-GST antibody (*Figure 21C*). GST-SH3E1 was also clearly detectable in Western blots, whereas in the negative control GST was only detected as a very weak band in all samples.

The three clones O16511, L13541 and K19568 were false positive in the screen on high-density protein filters. No binding of GST-SH3E1 to a His-tagged protein from false positive clones was observed in pulldowns. Six out of nine clones which expressed a protein inframe with the His-tag and three out of four clones with a frameshift were able to pull down GST-SH3E1.

Two clones were detected on the high-density protein array which expressed the hepatocellular carcinoma-associated antigen HCA25a (clone B08508 and C01529). Both clones were able to efficiently pull down GST-SH3E1. A detailed analysis of their DNA sequence revealed interspersed repetitive repeats. An estimated third of the human genome consists of interspersed repetitive DNA sequences which are primarily degenerate copies of transposable elements. These are unstable DNA elements that can migrate to different parts of the genome (Smit, 1996). The protein translated from the DNA sequence contained two repeats of proline-rich regions which are assumed to mediate binding to the SH3 domain of endophilin-1 (Appendix E).

DNA sequence analysis of clone K21548 turned out that the reading frame was correct, but the sequence matched the 3´-UTR downstream of the ORF. The coding sequence of the matched mRNA (BC039204) was from bp 178-3333. The inserted sequence of clone K21548 started at bp 3334 and ended at bp 3613 with a stop codon. The insert codes for a stretch of 93 amino acids including a proline-rich motif and

translation yields a polypeptide of 12 kDa which agrees with the calculated molecular weight of 12.082,08 Dalton including the His-tag (Appendix F).

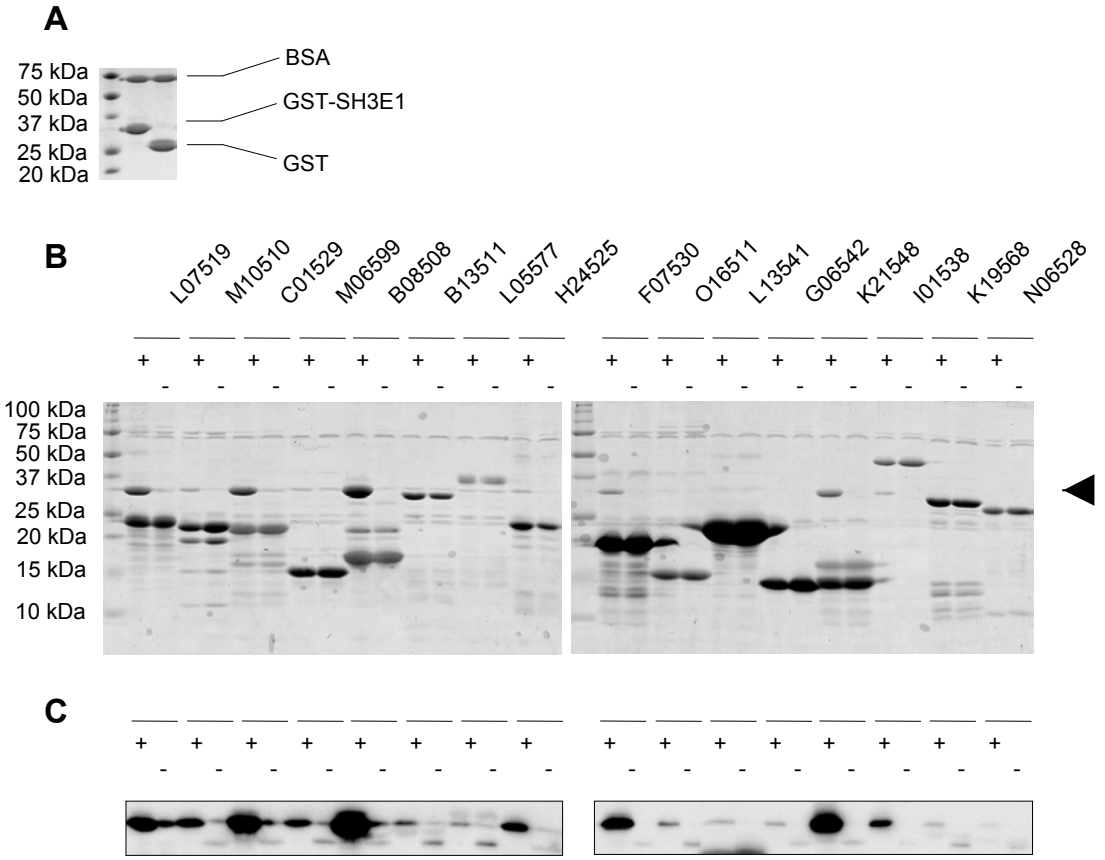


Figure 21. Pulldown assay with NiNTA agarose beads. His-tagged proteins were bound to NiNTA agarose beads and incubated with GST-SH3E1 (+) or GST (-). A) Aliquots of GST-SH3E1 or GST with 0.1 % (v/v) BSA. B) SDS-PAGE analysis of eluates with bound GST-SH3E1 or GST, arrow indicates captured GST-SH3E1. Clones are listed in *Table 11*. C) Western blot analysis of bound GST-SH3E1 or GST with anti-GST antibody.

Expression products from three out of four clones with a frameshift were found to bind GST-SH3E1. The two clones H24525 and M06599 contain the full ORF of a gene with a partial stretch of its 5´-UTR, whereas clone M10510 contains only a partial ORF of a gene. The three clones express an artificial His-tagged polypeptide of expected size with proline-rich motifs which most likely mediated binding of GST-SH3E1. (Appendix G-I).

A total of 30 clones which expressed a protein inframe with the His-tag were assayed for their ability to pull down GST-SH3E1 (*Table 12*). None of the 30 clones includes a 5´-UTR which can otherwise code for a partially artificial amino acid sequence. All of the attempted 30 expressed proteins achieved NiNTA purification levels

detectable in Coomassie-stained polyacrylamide gels. Fourteen of the expressed proteins were able to pull down GST-SH3E1 (*Figure 22*, and indicated in *Table 12*). His-tagged amphiphysin-1 (aa 291-559) was used as a positive control. Amphiphysin-1 efficiently pulled down GST-SH3E1 (see *Figure 22*, control), concurrently confirming the already described interaction of amphiphysin-1 and endophilin-1 (see also *Figure 10B*). The immunodecorated bands were quantified densitometrically to compare the signal intensities of pulldown positive proteins with that of amphiphysin-1 as a positive control (*Figure 22C*).

Table 12. Identification of potential binding proteins for the SH3 domain of endophilin-1.

clone	gene ID	gene symbol	protein	SwissProt ID	aa	PD
A11548	1072	CFL1	cofilin 1	P23528	86-166	-
A12558	51329	ARL6IP4	ADP-ribosylation-like factor 6 interacting protein 4	Q9NR05	1-121	-
A15579	22839	DAP4	disks large-associated protein 4	Q9Y2H0	551-989	+
B13511	79004	CUEDC2	CUE domain containing protein 2	Q9H467	1-287	-
C24599	1466	CSRP2	cysteine and glycine-rich protein 2 (CRP2)	Q16527	1-192	+
E03569	3146	HMGB1	high-mobility group box 1 protein	P09429	27-214	-
E04569	55684	FLJ10101	hypothetical protein FLJ10101	Q8TCL4	224-469	+
E05577	23152	CIC	capicua homolog (Drosophila)	Q96RK0	1314-1608	+
E08544	4809	NHP2L1	NHP2 non-histone chromosome protein 2-like 1	P55769	3-128	-
E23538	777	CACNA1E	calcium channel, voltage-dependent	Q15878	2105-2312	-
E23549	322	APBB1	amyloid beta (A4) precursor protein-binding	O00213	1-710	-
F05601	5372	PMM1	phosphomannomutase 1	Q92871	9-262	-
F07530	1466	CSRP2	cysteine and glycine-rich protein 2 (CRP2)	Q16527	41-192	+
F10569	3178	HNRPA1	heterogeneous nuclear ribonucleoprotein A1	P09651	1-371	+
G05529	349338	CXYORF1	DKFZP434K1323	Q9NSV7	46-250	-
G07549	146909	LOC146909	hypothetical kinesin motor domain containing	Q86Y91	297-410	-
G11565	84557	MAP1LC3A	microtubule-associated protein 1 light chain 3A	Q9H491	4-170	+
H01511	23534	TNPO3	transportin 3	Q9Y3R2	245-923	-
H09541	140465	MLC1SA	myosin light chain 1	P14649	42-208	-
I01538	64762	C18ORF11	hypothetical protein FLJ21610	Q9H706	498-875	+
I07563	1466	CSRP2	cysteine and glycine-rich protein 2 (CRP2)	Q16527	1-192	+
I17574	57522	SRGAP1	SLIT-ROBO Rho GTPase activating protein 1	Q7Z6B7	869-1085	+
L05517	64397	ZFP106	zinc finger protein 106	Q9H2Y7	228-1883	-
L07519	22839	DAP4	disks large-associated protein 4	Q9Y2H0	551-989	+
M08528	2737	GLI3	GLI-Krüppel family member GLI3	P10071	704-1596	+
M10533	9618	TRAF4	TNF receptor-associated factor 4	Q9BUZ4	208-470	-
M14512	667	BPAG1	bullous pemphigoid antigen 1	O94833	5092-5171	-
N06528	4665	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	Q15742	356-525	-
O01523	10015	PDCP6IP	(ALG-2)-interacting protein 1 (Alix)	Q8WUM4	240-868	+
O06583	63974	NEUROD6	neurogenic differentiation 6	Q96NK8	20-337	-
O23567	57517	KIAA1295	hypothetical protein	Q9P2Q1	157-550	+

[a]RZPD clone IDs without prefix (MPMGp800).

*PD, pulldown. (+) pulldown positive, (-) pulldown negative.

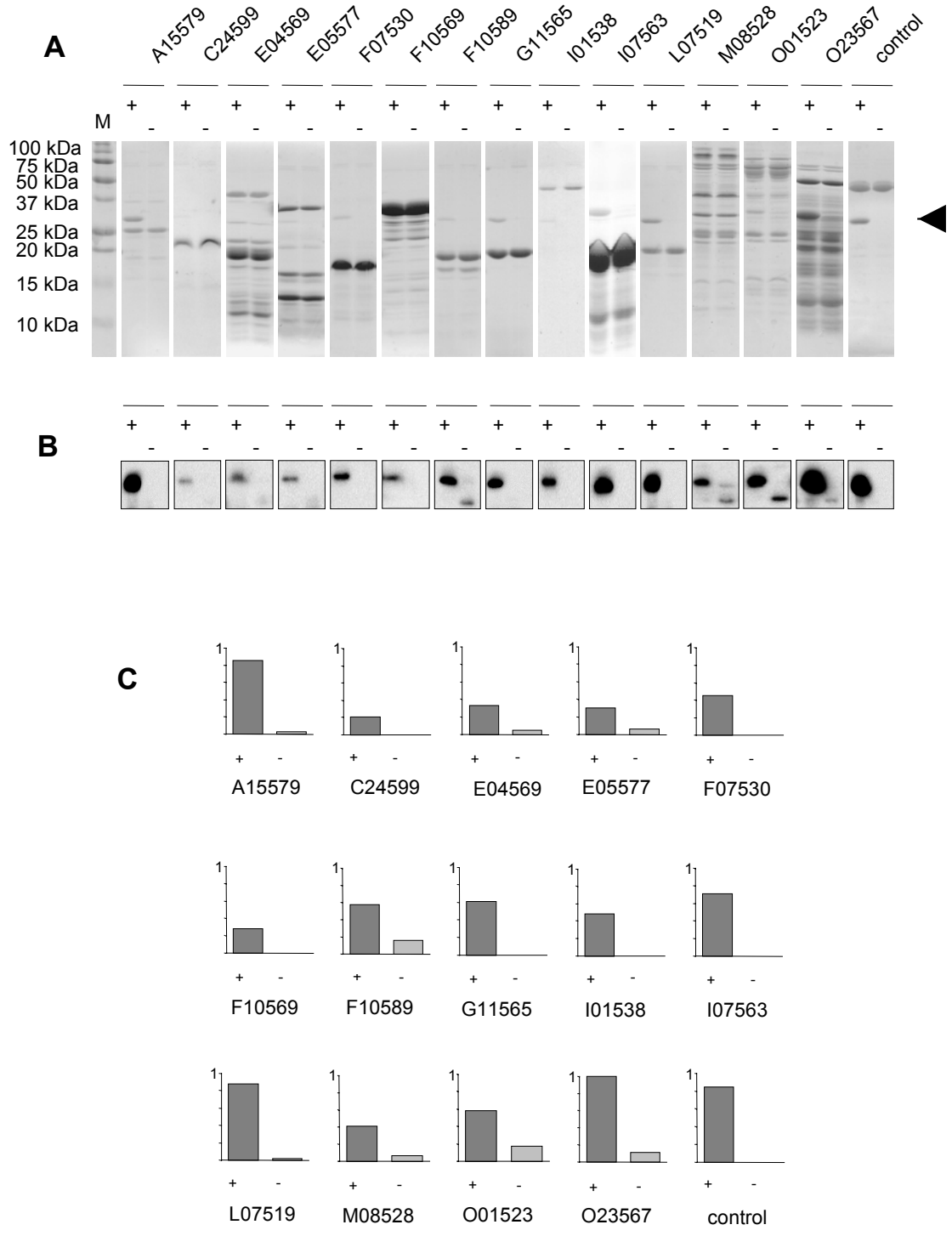


Figure 22. Pulldown assay with NiNTA agarose beads. Purified proteins were incubated with equal amounts of GST-SH3E1 or GST. After a washing step bound proteins were analyzed by SDS-PAGE and Western blotting. A) SDS-PAGE of co-purified proteins. Aliquots of bound proteins were separated by SDS-PAGE. His-tagged amphiphysin-1 served as a positive control, arrow indicates captured GST-SH3E1. B) Bound proteins were further analysed by Western blotting with anti GST antibody. C) Quantification of GST-SH3E1 or GST was done by densitometry of imunodecorated bands. The highest measured value was arbitrarily set as 1 (see O23567).

The previously published interaction of endophilin-1 with Alix found in the screen on high-density protein arrays was confirmed in the pulldown assay. Clone O01523 expresses a 75 kDa C-terminal fragment of Alix. This fragment contains a proline-rich motif at its C-terminus, which is highly homologous to the previously demonstrated endophilin binding sites of synaptojanin 1 and rat germinal center-like kinase rGLK (Chatellard-Causse *et al*., 2002).

The novel interaction between the disks large-associated protein 4 (DAP4) and endophilin-1 was also confirmed in this pulldown assay. DAP4 is an adapter protein that plays a role in the molecular organization of synapses and in neuronal cell signaling (Takeuchi *et al*., 1997). DAP4 contains four proline-rich motifs, indicating that this protein is a potential interaction partner for SH3 domains. Partial DNA sequencing turned out that clone A15579 as well as clone L07519 contained a C-terminal fragment of DAP4.

GST-SH3E1 identified three clones which contained either the full ORF (clones C24599 and I07563) or a partial ORF (clone F07530) of cysteine and glycine-rich protein 2 (CRP2). The protein CRP2 contains no proline-rich motif, but expression products of all three clones were able to bind the SH3 domain of endophilin-1 in pulldowns, albeit to different extents. The predicted size of full length CRP2 including the His-tag is 25.759,17 Dalton, which was in good agreement with the purified 26 kDa proteins of C24599 and I07563, respectively. In contrast to clone I07563, only a minor fraction of CRP2 expressed by clone C24599 was purified (*Figure 22A*) which might explain the difference in the amount of bound GST-SH3E1 detected by SDS-PAGE analysis and Western blotting. The region responsible for the interaction with GST-SH3E1 is supposed to be in the C-terminal part of CRP2, since a fragment of CRP2 as well as full length CRP2 pulled down GST-SH3E1. CRP2 contains two LIM domains, which represent completely independent folded units bridged by a flexible linker region. LIM domains consist of two independent zinc fingers that function as a protein interaction interface capable of recognizing different specific targets. It is evident, that the LIM domain containing protein CRP2 is involved in promoting protein assembly along the actin-based cytoskeleton and is of fundamental importance for distinctive functions in cell differentiation, cytoskeletal remodeling and transcriptional regulation (Weisskirchen and Günther, 2003).

Clone G11565 codes for the microtubule-associated protein 1 light chain 3A (MAP1LC3A) and expressed a 22 kDa protein (MAP1LC3A, aa 4-170). MAP1LC3A contained a stretch of amino acids at its N-terminus consisting of prolines intersected either by the basic amino acid arginine or the hydrophobic amino acid alanin. MAP1LC3A was able to efficiently pull down the SH3 domain of endophilin. Three different light chains with distinctive functions were characterized for MAP1A, which predominantly bind to the heavy chain of MAP1A, but there is strong evidence that MAP1LC´s have additional functions as potential linkers of neuronal microtubules and microfilaments (Noiges *et al*., 2002).

The pulldown experiments confirmed the binding of several hypothetical proteins. Clone I01538 expressed a C-terminal fragment of the hypothetical protein C18ORF11. The strongest signal of copurified GST-SH3E1 detected in Western blots was derived from clone O023567, expressing a C-terminal fragment of the hypothetical protein KIAA1295 which possesses one SH3 domain at either termini.

4.6 Peptide scans of endophilin-1 binding proteins

To map the binding regions of endophilin 1´s SH3 domain and to confirm the found interactions with a second independent *in vitro* method, a subset of twelve proteins identified in the initial filter overlay screen were selected. This subset consisted of pulldown positive as well as pulldown negative proteins. Protein sequences were analysed for the presence of relaxed SH3 binding consensi with the software tool ELM (Puntervoll *et al*. 2003). Two classes of SH3 binding consensi exist (*Figure 23*).

Depending on the position of a positive amino acid relative to the proline core PxxP motif, peptide sequences are recognized in two different orientations. Selected protein sequences were SPOT synthesized as an array consisting of 14mers with an overlap of 10 amino acids on cellulose membranes. Most peptides on the array contained potential SH3 binding motifs of class I ([R/K/Y]xxPxxP), class II (PxxPx[K/R]) or of the non-canonical class I (xxx[P/V]xxP). Peptides were probed with either GST-SH3E1 or GST as a control and bound proteins were detected with anti-GST antibody followed by detection of chemiluminescence signals (*Figure 24*).

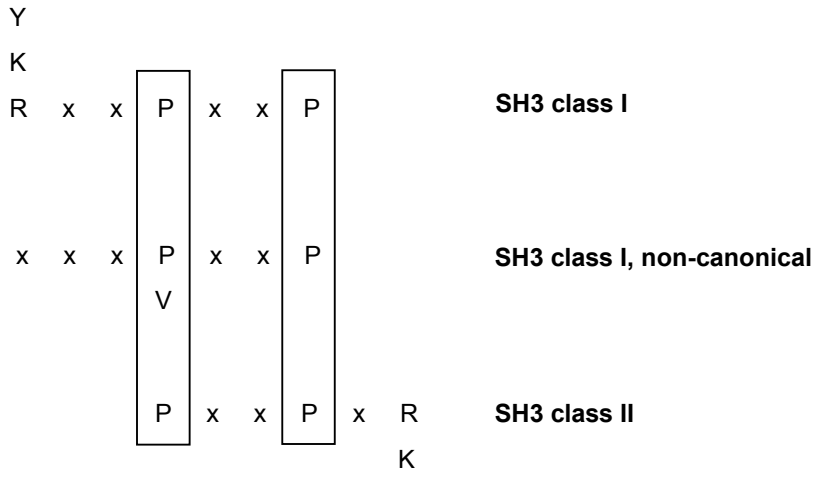


Figure 23. SH3 binding consensus motifs. SH3 domains recognize peptides in two different orientations. Depending on theposition of a positively charged residue relative to the core PxxP motif, SH3 domains can bind in a class I or class II orientation.

For most proteins the signal intensities of peptide scans (*Table 13*) correlated with results from pulldown experiments (*Table 12* and *Figure 22*). Peptides detected with a high signal intensity can be assigned to proteins which were able to efficiently pull down GST-SH3E1. One exception was the protein NAB2 which displayed a high signal intensity for peptide 2 and 3, but was not able to pulldown GST-SH3E1. Peptide scans for the two pulldown negative proteins neurogenic differentiation factor 6 (NEUROD6) and the hypothetical kinesin motor domain containing protein (LOC146909) also yielded no signals on the peptide array. In contrast to results from the pulldown assay, no peptide was detected for the protein CRP2 which specifically recognized the SH3 domain of endophilin-1, although CRP2 was able to bind to GST-SH3E1 in the pulldown assay.

The tightest SH3 binding peptides from the hypothetical protein KIAA1295 (peptides 7, 8) contain an amino acid stretch highly similar to that of the protein Alix (peptides 4, 5). The peptide scan for DAP4 suggests two potential proline-rich motifs spanning the regions from amino acid 565-582 (peptides 1, 2) and 916-937 (peptides 3, 4, 5). The first two peptides were detected by GST-SH3E1 with an equally strong signal, whereas in the second region only peptide 4 gave a strong signal. A number of peptides derived from amino acid sequences of protein KIAA1295 (peptides 1, 2), SRGAP1 (peptides 1-3), MAP1ALC3A (peptide 2), FLJ10101 (peptide 1) and NAB2 (peptide 1) were recognized by GST-SH3E1, even though they did not contain an SH3 consensus motif of class I or class II.

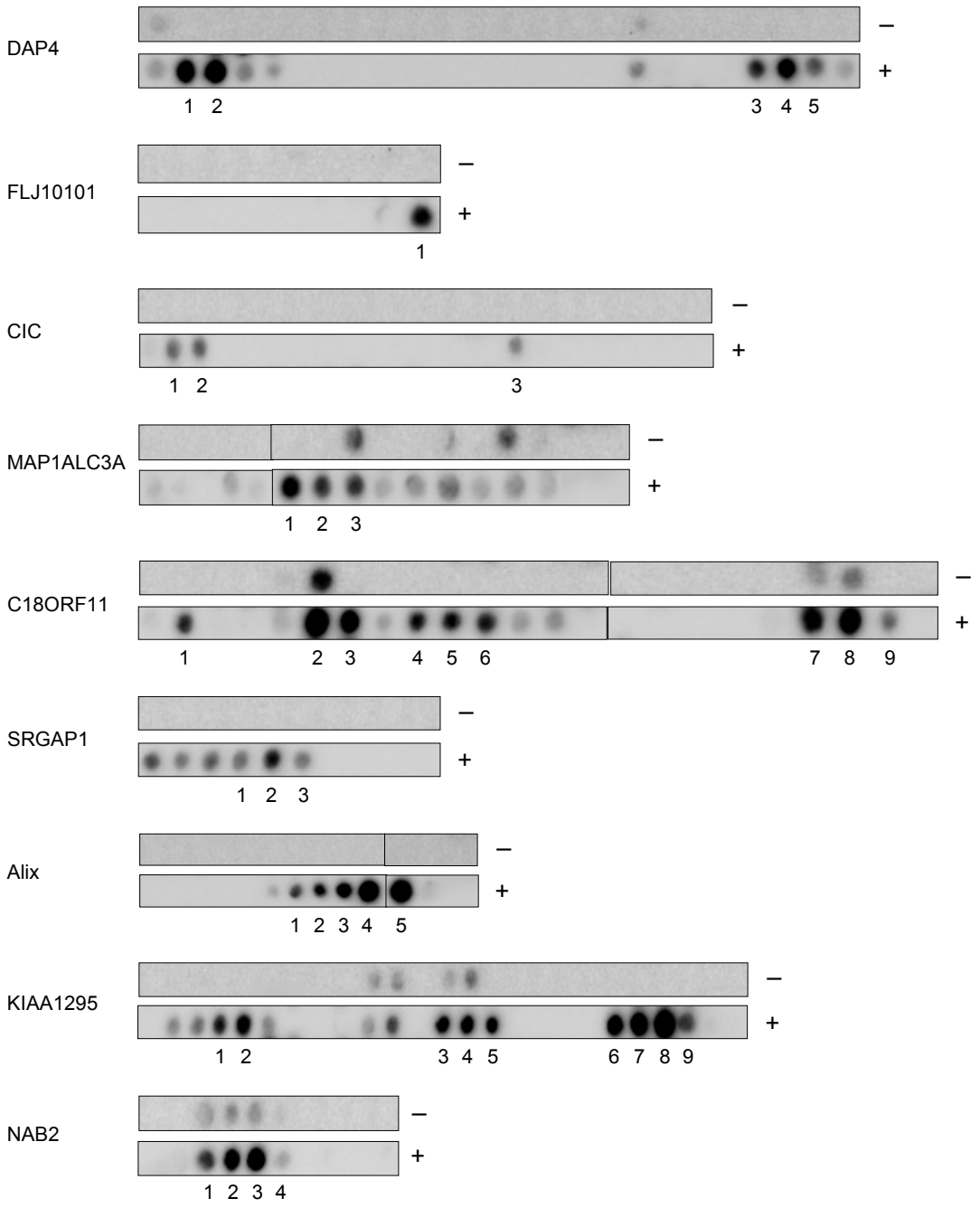


Figure 24. Peptide scans of endophilin-1 binding proteins. Peptide scans of overlapping 14mers derived from protein sequences of endophilin-1 binders were probed with GST-SH3E1 (+) or GST (-). Bound probe protein was detected with anti-GST antibody. Numbered signals indicate peptides which are also listed in *Table 13*.

Table 13. Identification of SH3 domain binding motifs for endophilin-1.

protein	SwissProt ID	peptide no.	aa	peptide sequence	signal intensity	SH3 class I motifs	SH3 class 2 motifs
DAP4	Q9Y2H0	1	565-578	VAYKKTPPPVPPRT	++++	KTPPPVP / TPPPVPP	PPVPPR
		2	569-582	KTPPPVPPRTTSKP	+++++	KTPPPVP / TPPPVPP	PPVPPR
		3	916-929	RKEEKKPPPPVPKK	++	EKKPPPP / KPPPPVP	PPVPPK
		4	920-933	KKPPPPVPKKPAKS	++++	KPPPPVP / PPVPKKP	PPVPPK / PKKPAK
		5	924-937	PPVPKKPAKSKPAV	+	PPVPKKP	PPVPPK / PKKPAK
FLJ10101	Q8TCL4	1	360-373	PPPKLPLPAFRLKN	+++		
CIC	Q96RK0	1	1328-1341	GKKVKVKRPPPLKKT	++	KVKVRPP	
		2	1332-1345	KVRPPPLKKTFDSV	++	KVKVRPP	
		3	1566-1579	SGTAQAAPPLPPPP	+	QAAPPLP / AAPPLPP / PPLPPPP	
MAP1ALC3A	Q9H491	1	33-46	TSPRPRAPACAPSR	++++	SPRPRAP	
		2	37-50	PRAPACAPSRARAM	++		
C18ORF11	Q9H706	1	511-524	LPPPPVPPKSEAVR	++	LPPPPVP / PPPPVPP	PPVPPK
		2	527-540	CRLLNAPPVPPRSA	++++++	LNAPPVP / NAPPVPP	PPVPPR
		3	531-544	NAPPVPPRSAKPLS	+++++	NAPPVPP	PPVPPR
		4	539-552	SAKPLSTSPSIPPR	+++	STSPSIP	PSIPPR
		5	543-556	LSTSPSIPPRTVKP	+++	STSPSIP	PSIPPR
		6	547-560	PSIPPRTVKPARQQ	+++		PSIPPR
		7	714-727	VTKQSTSCPALPPR	++++	TSCPALP	PALPPR
		8	718-731	STSCPALPPRAPKL	+++++	TSCPALP / ALPPRAP	PALPPR
		9	722-735	PALPPRAPKLVEEK	+	ALPPRAP	PALPPR
SRGAP1	Q7Z6B7	1	1047-1060	VQLKPPALRPKPAV	+		
		2	1051-1064	PPALRPKPAVLPKT	+++		
		3	1055-1068	RPKPAVLPKTNPTI	+		
Alix	Q8WUM4	1	739-752	PPTPAPRTMPPTKP	++	TMPPTKP	
		2	743-756	APRTMPPTKPQPPA	+++	TMPPTKP / PTKPQPP	
		3	747-760	MPPTKPQPPARPPP	++++	PTKPQPP / PQPPARP	PQPPAR
		4	751-764	KPQPPARPPPPVLP	++++++	PARPPPP / PPPPVLPP	PQPPAR
		5	755-768	PARPPPPVLPANRA	+++++	PARPPPP / PPPPVLPP	
KIAA1295	Q9P2Q1	1	273-286	PFLKSRPQVRPKPA	+++		
		2	277-290	SRPQVRPKPAPSPK	++++		
		3	359-372	SPKEISCRAPPRPA	+++	CRAPPRP	
		4	363-376	ISCRAPPRPAKTTD	+++	CRAPPRP	PPRPAK
		5	367-380	APPRPAKTTDPVSK	+++		PPRPAK
		6	387-400	QEAPQQRPVVPPRR	++++	QQRPVVP / QRPVVPP	PVVPPR
		7	391-404	QQRPVVPPRRPPPP	+++++	QQRPVVP / VVPPRRP / PRRPPPP	PVVPPR
		8	395-408	VVPPRRPPPPKKTS	++++++	VVPPRRP / PRRPPPP	PPPPKK
		9	399-412	RRPPPPKKTSSSSR	++		PPPPKK
NAB2	Q15742	1	468-481	RLVSHDRVGRLSPC	++		
		2	472-485	HDRVGRLSPCVPAK	++++	RLSPCVP	PCVPAK
		3	476-489	GRLSPCVPAKPPLA	+++++	PCVPAKP / RLSPCVP	PCVPAK
		4	480-493	PCVPAKPPLAEFEE	+	PCVPAKP	PCVPAK
CRP2	Q16527	-					
NEUROD6	Q98NK8	-					
LOC146909	Q86Y91	-					

*Peptides are numbered as in *Figure 24*.

The peptide scan of the hypothetical protein C18ORF11 displayed two peptides each with a high signal intensity for the region from 527-544 (peptides 2, 3) and 714-731 (peptides 7, 8), and three peptides with a moderate signal intensity for the region from 539-560 (peptides 4, 5, 6). A weak signal was detected for peptide 9 of C18ORF11, even though peptide 9 contained the same SH3 consensus motifs like the overlapping peptide 8. A similar observation was made for the proteins DAP4 (peptide 3) and NAB2 (peptide 4). A common feature of the three peptides is the presence of a stretch of consecutive glutamate residues which flanks the SH3 binding motif.

5 Discussion

5.1 Protein expression in *E. coli*

In this study, proteins for *in vitro* binding assays were expressed in *E. coli*. An important prerequisite for *in vitro* binding studies is soluble protein. To express a large fraction of the generated proteins in a soluble form, the design of PCR primers that match suitable domain boundaries for an individual protein was taken into consideration. Consequently, most of the proteins were expressed in a soluble form as GST fusion proteins (*Figure 3*) as well as His-tagged proteins (*Figure 6*). Depending on the expression strength and solubility of an individual protein, varying volumes of a soluble protein extract were purified to obtain the amount of recombinant protein needed for a certain application.

Since the cDNA expression library generated in *E. coli* and used in this study was constructed by a random strategy, various clones express a protein with unappropriate domain boundaries. Thus, a high yield of soluble recombinant protein cannot be expected considering a bacterial cDNA expression library in its entirety. On a case-by-case basis the individuality of proteins demands highly customized procedures for a soluble expression. Bacterial systems provide high yields of recombinant protein in a relatively short time, but expression of eukaryotic proteins may be problematic, since the expression products can form insoluble inclusion bodies and aggregates or can be degraded within the host cell (Hockney, 1994; Makrides, 1996). When eukaryotic expression hosts are used alternatively, they suffer from lower yields of heterologous protein (e.g., *Saccharomyces cerevesiae*; Buckholz and Gleeson, 1991), high demands on sterility (e.g., mammalian systems; Aruffo, 1997; Kingston *et al*., 1997), or time-consuming cloning procedures (e.g., Baculovirus-mediated expression in insect cells; Miller, 1993). In this study, 86 % of the attempted bacterial clones from the cDNA expression library expressed soluble protein that can be detected in Coomassie-stainable amounts (*Figure 16* and *17A*). Therefore, the use of bacterially expressed recombinant protein in a subsequent suspension-binding screen using a pulldown assay was quite reasonable.

5.2 Filter overlay screens

PVDF filter membranes with an array of 96 His-tagged proteins were overlayed with the radiolabeled probes GST-PKA-Hsc70, GST-PKA-HOP and GST-PKA-SH3E1. All probes detected their respective previously known interaction partners on the array. This result clearly demonstrated the reliability of the screening approach. Filter overlaying with a radiolabeled GST fusion protein was extended to high-density protein arrays. A screen with radiolabeled GST-PKA-SH3E1 detected a total of 401 positive clones. Each clone on the high-density protein array was spotted by a robot in duplicate in a defined pattern. The spotting pattern allowed for a reliable scoring of positive duplicates on the array.

Since individual clones are positionally adressable on the PVDF membrane, screens with high-density protein arrays are superior to screens with λgt11 cDNA expression libraries that have been widely used in the past to discover protein-protein-interactions (Blackwood and Eisenman, 19991; Blanar and Rutter, 1992; Ayer *et al*., 1993, Einarson and Chao, 1995). λgt11 cDNA expression libraries are plated on agar dishes and plaques that have bound the probe protein are at random positions. Several rounds of time-consuming selection steps are necessary to isolate a clone and the identitiy of a cloned gene must be unraveled by DNA sequencing in any case. Provided that a clone on the high-density protein array had been sequenced before, the identity of its gene is immediately available and can be directly linked to the microtitre plate position in the cDNA library. In addition, the clone allows direct expression of the recombinant protein for further studies.

5.2.1 Signal-to-noise ratio

The signal-to-noise ratio in filter overlays was drastically increased for PVDF membranes with spotted clones (*Figure 10*) compared to that for PVDF membranes with spotted soluble protein extracts (*Figure 9*). Assuming that an average of 5-20 µg soluble recombinant protein was produced from a 1 mL bacterial culture, each spot area of soluble protein extract contained 2-10 ng of immobilized recombinant protein. Using spotted clones, the whole cellular extract was released on the membrane. Thus, the insoluble fraction of the recombinant protein was additionally immobilized. Moreover, robotic spotting of clones is more precise than manually spotting soluble protein extracts. This leads to a higher local concentration of recombinant protein per spot area than in the

approach with spotted soluble extracts. Superior signals were obtained on membranes with spotted clones, exceeding drastically those of the background. As a consequence, more stringent washing conditions can be applied to reduce the background signals.

5.2.2 Signal intensities

Clones were detected with varying signal intensities. Only duplicates with an equal signal intensity were considered to be true positive hits. The amount of His-tagged protein immobilized per spot varies depending on the expression strength of the clone. Provided that a probe protein binds its ligand with a high affinity but the amount of immobilized ligand per spot area is low, a relative small amount of probe protein remains bound after washing. Consequently, the spot is detected with a low signal intensity, although the binding of the probe protein relies on a strong affinity for the ligand. A second parameter that is not known when the overlay screen is initiated, is the structural integrity of the ligand after a denaturing release on the membrane. A fraction of the recombinant protein may adopt its active conformation in subsequent steps, but denaturation and renaturation is difficult to control if thousands of proteins are treated in parallel. Provided that a correct three-dimensional protein structure is necessary for an interaction with a probe protein, the capability of an immobilized ligand to capture the probe protein also depends on the extent to which a ligand is properly folded and exhibits an active conformation.

5.2.3 False positives

As observed for yeast two-hybrid screens and *in vivo* pulldowns, the identification of protein-protein-interactions on a large scale suffers from a high number of false positive hits. A filter overlay screen with radioactively labeled GST fusion protein can generate a background of false positive signals. In a proof-of-principle approach two signals were scored with radioactively labeled GST as a probe on the protein array with spotted clones (*Figure 10C*). The two signals were also present in the absence of radiolabeled probe protein on the array with spotted soluble extracts (*Figure 9C*). These signals were originated from clones expressing a fragment of filamin A (aa 2015-2198). Even though several clones expressed other fragments of filamin A and were gridded on the array, they were not detected. The amino acid sequence of the two clones contains a serine at position 2152. Ser2152 lies within a potential phosphorylation site highly homologous to

that of protein kinase A (PKA). It has been shown that Ser2152 lies within a phosphorylation site for protein kinase B (Obata *et al*., 2000) and that Ser2152 of filamin A can also be phosphorylated by ribosomal S6 kinase RSK *in vivo* (Woo *et al*., 2004) and protein kinase A *in vitro* (Jay *et al*., 2000).

The fact that the two signals were scored when the filter was overlayed with a phosphorylation reaction in the absence of a radioactively labeled probe, suggests that a residual enzyme activity of PKA and traces of unincorporated [γ-^{32}P]ATP remained in the desalted phosphorylation mix. Overlay screens with radiolabeled probes are more sensitive compared to screens which detect protein-protein-interactions through an antibody directed against the probe protein (Mahlknecht *et al*., 2001). Minute amounts of unincorporated [γ-^{32}P]ATP in the overlay buffer seemed to be sufficient to produce false positive signals. Such signals are most likely originated from proteins that can be phosphorylated by protein kinase A *in vitro* as shown for filamin A in this study. Another explanation would be that background signals can also be generated from unincorporated [γ-^{32}P]ATP which can bind non-covalently to proteins. Therefore, a complete removal of unincorporated [γ-^{32}P]ATP helps to reduce the background signals. In addition, binding of the GST moiety of fusion proteins might generate a background of false positive signals. Thus, scoring of autoradiograms from protein arrays requires an additional overlay with a suitable negative control GST fusion protein (see *Figure 14*).

5.2.4 The SH3 domain of endophilin-1 binds to short artificial polypeptides

A second category of signals is originated from cloning artifacts and cannot simply be discriminated by image analysis of autoradiograms. The cDNA expression library was enriched for clones producing a His-tagged protein with correctly fused reading frame. This was done through a preselection step (Büssow *et al*., 2000). Clones harbouring out-of-frame fusions express artificial polypeptides. These polypeptides are generally short because of a high probability of stop codons in artificial reading frames. Artificial polypeptides are frequently folded into unstable structures that are degraded within the host cell (Gottesman, 1996). Therefore, clones expressing artificial polypeptides cannot be detected with an antibody directed against the His-tag and were removed from the library.

Sixty-eight percent of the transcripts that were detected with the SH3 domain of endophilin-1 (GST-PKA-SH3E1) but not with the negative control protein probe (GST-

PKA-p60) code for out-of-frame fusions. An explanation for this would be that a fraction of artificial polypeptides were still able to fold into stable expression products. Thus, they escaped the preselection step. SH3 domains accommodate binding to short unstructured peptides with proline-rich regions in their binding partners. Therefore, short artificial polypeptides with such proline-rich regions were specifically recognized by the probe GST-PKA-SH3E1.

Most strikingly, many different clones representing two out-of-frame transcripts were recognized by GST-PKA-SH3E1. 84 signals were scored for a transcript of myristylated alanine-rich protein kinase C substrate (*MARCKS*) and 40 signals were scored for creatine kinase B (*CKB*). A partial 5´-UTR upstream of the start codon of both transcripts led to the translation of short artificial polypeptides (see *Figure 15*). These artificial polypeptides contain a proline-rich sequence with a nearly perfect consensus motif for SH3 domains. A 47 kDa polypeptide was expressed from clones with cDNA of *CKB*. Mass spectrometric analysis revealed that this polypeptide was CKB in a correct reading frame. The 5´-UTR of *CKB* contains a GC-rich region which might lead to a slippage during transcription in *E. coli*, thereby correcting the reading frame for *CKB*. This phenomenon has been earlier observed in *E. coli* cells by Wagner *et al*. (1990). Another explanation would be that the cDNA of CKB contains internal start codons. Since signals of MARCKS and CKB were scored on the array overlayed with GST-PKA-SH3E1 but not in the control overlay with the 60 kDa heat shock protein GST-PKA-p60, it is likely that these proline-rich sequences were recognized specifically by the SH3 domain of endophilin-1. Moreover, the GC-rich DNA sequence upstream of the ORF for *CKB* codes preferentially for the amino acids proline, arginine and alanine. These amino acids are the main components of consensus motifs which are preferred by the SH3 domain of endophilin-1. Consecutive prolines intersected by the basic amino acid arginine and the hydrophobic amino acid alanine were also found in the peptide scans for GST-SH3E1 binding proteins (see *Table 13*).

The finding that GST-PKA-SH3E1 recognized many proline-rich motifs of artificial polypeptides was also substantiated in the pulldown assay. Three out of four expression products from clones with an incorrectly fused reading frame were capable to pull down the SH3 domain of endophilin-1 (see *Table 11* and *Figure 21*). The three artificial polypeptides contain proline-rich sequences with SH3 consensus motifs as shown for CKB and MARCKS.

5.3 *In vitro* pulldown assays

Two different pulldown techniques were tested, to confirm protein-protein-interactions found with GST-PKA-SH3E1 in the high-density protein array screen. In the first approach, GST-SH3E1 was crosslinked to glutathione agarose beads. Covalent crosslinking was chosen, since a large excess of eluted GST fusion protein may obscure interacting proteins in the SDS-PAGE analysis. The covalently bound GST-SH3E1 was incubated with purified His-tagged proteins to pull down specifically interacting proteins. Since the results obtained with this pulldown assay were unsatisfactory, a second approach was done. In this approach, His-tagged proteins were bound to NiNTA agarose beads and incubated in turn with GST-SH3E1 to pull down specifically interacting GST-SH3E1. In both cases GST was used as a control. Co-purifying proteins were fractionated by SDS-PAGE. Gel profiles from pulldowns with GST-SH3E1 were compared to those obtained from pulldowns with GST to identify proteins that interact specifically. To increase the sensitivity of detection, Western blots were done in addition.

5.3.1 Minimizing the effect of aggregation

Irreversible aggregation with time or thermal stress is a major problem when using proteins for *in vitro* binding studies. When these aggregates precipitate or form large particles they are easy to detect, but smaller aggregates which are often precursors can be difficult to detect and quantitate. Such aggregates can interfere with the SDS-PAGE analysis following a pulldown assay, since they cannot be removed from specifically binding protein. A pulldown experiment was initiated with GST-SH3E1 crosslinked to glutathione agarose beads. The amount of His-tagged protein observed in pulldowns with GST-SH3E1 did not significantly differ from that of the GST control (see *Figure 20*, Western blotting). This finding suggests that a large fraction of His-tagged protein formed insoluble aggregates which were sedimented together with the agarose beads during the centrifugation step. Consequently, it was not possible to distinguish specifically binding proteins.

A large number of protein stabilizers are available which can prevent aggregation when added to the assay buffer at an appropriate concentration (Bondos and Bicknell, 2003), but each protein might need an individual treatment. To circumvent solving the problem for each protein individually, His-tagged proteins were immobilized on NiNTA agarose beads and purified GST-SH3E1 or GST was added in solution with bovine serum

albumine (BSA) as an additive. It is well known that BSA prevents adhesion of proteins to reaction tubes and stabilizes enzymes during incubation steps. Addition of endogenous BSA may also protect purified recombinant proteins from proteolytic degradation by contaminating proteases. BSA stabilized GST-SH3E1 and GST and prevented them from being pulled down unspecifically. BSA was removed by a washing step after the pulldown and did not interfere with the subsequent SDS-APGE analysis. The amount of protein was dramatically diminished in the negative control compared to the negative controls of pulldowns using glutathione agarose indicating that no aggregates were formed. This enabled a clear and reliable differentiation between specifically binding GST-SH3E1 and GST (see *Figure 21*).

5.3.2 Variables that influence protein binding in a pulldown experiment

Several other variables determine the success of an *in vitro* pulldown experiment and must be taken into consideration when evaluating the results from a pulldown assay.

Amount of His-tagged protein coupled to NiNTA agarose beads

A limitation to identify protein-protein-interactions by a pulldown technique is isolating sufficient quantities of soluble protein for the assay. The extent to which individual proteins were expressed in a soluble form strongly varied in a subset of 96 His-tagged proteins from the cDNA expression library (*Figure 16* and *17*). A fraction of 86% of the proteins can be purified from 1 mL cultures in amounts detectable in a Coomassie stained polyacrylamide gel. But, to efficiently bind GST-SH3E1, the amount of His-tagged protein immobilized to the NiNTA agarose beads should be as high as possible. The influence of the amount of His-tagged protein coupled to NiNTA agarose beads on the amount of pulled down GST-SH3E1 was demonstrated with the protein CRP2. Two clones expressed full length CRP2. In contrast to clone I07563, only a smaller amount of CRP2 was purified from clone C24599 which led to a smaller amount of GST-SH3E1 being pulled down (*Figure 21*).

Incubation time

The incubation time applied for complex formation prior to the pulldown influences the amount of bound protein. The *in vitro* pulldown assay examines the interaction between a His-tagged protein immobilized on the surface of NiNTA agarose

beads and a GST fusion protein free in solution. Depending on the association rate of two proteins and their concentration in the assay buffer, complex formation can be slow. To enable the association of slowly binding proteins the time for complex formation should be sufficient. Contrarily, a prolonged incubation time can increase the probability of aggregate formation. For all proteins examined in the pulldown assay, an incubation time of 2 hrs. was chosen. A prolonged incubation time of 16 hrs. did not increase the amount of bound protein and did not lead to aggregation of GST-SH3E1 or GST (data not shown).

Washing conditions

Nonspecifically associated proteins were removed by washing the NiNTA agarose beads after incubation with GST-SH3E1 or GST. Although extensive washing will remove unspecifically binding proteins, too much washing can result in the loss of specifically associated GST-SH3E1. Several quick washing steps with small volumes of buffer are more efficient than fewer and prolonged washing in a larger volume. The extent of washing should be adjusted so as to maximize the capture of specifically interacting GST-SH3E1 while removing as much of the GST background in the control as possible. Western blot analysis showed that the backgound of GST in the negative control was reduced sufficiently by the washing conditions that were chosen (*Figure 21* and *22*).

5.3.3 Confirmation of protein-protein-interactions with the SH3 domain of endophilin-1

The filter overlay screen of a high-density protein array with radiolabeled GST-PKA-SH3E1 detected 30 expression products which were correctly fused to the His-tag and contained no untranslated region (*Table 12*). To independently confirm their binding to GST-SH3E1, they were assayed in parallel pulldown experiments using NiNTA agarose beads. His-tagged amphiphysin-1, a known interaction partner of endophilin-1, served as a positive control. Amphiphysin-1 efficiently pulled down GST-SH3E1 (see *Figure 25A*, control). Thus, the SH3 domain of endophilin-1 exhibited an active conformation and recognized its SH3-binding motif for amphiphysin-1. Among the 30 proteins tested, 14 were able to pull down GST-SH3E1 (indicated in *Table 12*). Although signal intensities of immunodecorated bands varied, it was clearly possible to distinguish

specific binding of GST-SH3E1 from unspecific binding of GST. 16 interactions, found by the filter overlay screen of the high-density protein array, were not confirmed in the pulldown assay. Probably, relatively weak protein-protein-interactions mediated by the SH3 domain of endophilin-1 are lost due to an extended washing applied in the pulldown assay. These interactions are most likely discovered in the overlay screen with radiolabeled GST-SH3E1, because of the high sensitivity of this screening method.

5.4 Peptide scans

Protein-protein-interactions are believed to be dependent on the structural conformation of proteins. However, there is an increasing number of examples of protein domains that recognize short linear epitopes of their binding partners. These protein-protein-interactions can be reproduced using short synthetic peptides lacking a defined three-dimensional structure. Peptide arrays prepared by the SPOT synthesis, which immobilizes peptides on a cellulose membrane (Reineke *et al*., 2001), provides a very helpful method to locate a binding site in a protein.

5.4.1 SH3 domains bind to small proline-rich peptides

The SH3 (Src homology 3) domain was first described in the Src cytoplasmic tyrosine kinase (Tanaka *et al*., 1987). SH3 domains are small protein modules consisting of approximately 60 amino acids. They are found in a great variety of intracellular or membrane-associated proteins implicated in signal transduction, cytoskeleton reorganization and membrane traffic (Pawson, 1995). SH3 domain-containing proteins comprise a variety of proteins with enzymatic activity, adaptor proteins that lack catalytic sequences and cytoskeletal proteins. The SH3 domain has a characteristic fold which consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets (Musacchio *et al*., 1992). A hydrophobic region constitutes the ligand binding surface. The observation that most SH3 ligands are proline-rich and structure analysis of SH3 domains complexed with their peptide ligands led to the formulation of a general SH3-peptide binding model (Lim *et al*., 1994; Feng *et al*., 1994). SH3 domains consist of two hydrophobic binding pockets followed by a third binding pocket which is lined by negative residues. The third binding pocket can accept a positively charged side chain. Depending on the position of a positive residue SH3 ligands can bind in either of two opposite orientations, termed class I or class II orientation.

To date more than 1500 different SH3 domains can be identified by search algorithms in protein databases. Within the SH3 family, variations in the chemical characteristics of the three binding pockets modulate peptide recognition and determine the recognition specificity of binding partners. Screening of the high-density protein array combined with a pulldown assay has confirmed a number of binding partners for the SH3 domain of endophilin-1 in this study. Identified proteins were analysed for regions with potential SH3 binding sites and peptides delineated from that regions were scanned for binding to the SH3 domain of endophilin-1.

5.4.2 Selectivity of the SH3 domain of endophilin-1

Even though the peptide array contained numerous peptides with potential SH3 binding motifs, only a small subset of peptides was detected by the SH3 domain of endophilin-1 (*Table 13*). The strongest signals were derived from peptides containing a non-canonical SH3 class I motif with basic (Arg, Lys) and hydrophobic (Ala, Val) variable residues. Peptides with acidic residues flanking the SH3 binding motif clearly reduced or even prevented binding of GST-SH3E1, as observed earlier for endophilins (Cestra *et al*. 1999). Two binding events with lower signal intensities were obtained for peptides exclusively containing class II motifs (C18ORF11, peptide 6 and KIAA1295, peptide 5) suggesting that class II motifs might also be recognized by the SH3 domain of endophilin-1.

The peptide scan for DAP4 suggests two potential proline-rich motifs spanning the regions from amino acid 565-582 and 916-937. Peptides 1 and 2 of the first region were detected by SH3E1 with an equally strong signal, whereas in the second region only the signal for peptide 4 was strong. The weak binding of peptide 3 in the C-terminal region can be explained by the presence of two consecutive acidic glutamate residues flanking the class I SH3 binding motif KPPPPVP. Only peptide 4 of the region 916-937 gave a strong signal. This peptide contains the class I motif KPPPPVP. Since region 916-933 of DAP4 contains acidic Glu residues, it can be suggested that the SH3 class I motif KTPPPVP or the class II motif PPVPPR present in the region 565-582 of DAP4 mediates binding of SH3E1.

A number of peptides derived from KIAA1295 (peptides 1, 2), SRGAP1 (peptides 1-3), MAP1ALC3A (peptide 2), FLJ10101 (peptide1) and NAB2 (peptide 1) were

recognised by SH3E1, even though they do not contain any classical SH3 consensus motif. The three peptides of SRGAP1 all contain the minimal sequence RPKPAV, but peptide 2 was detected with a higher signal intensity than peptide 1 and 3. Obviously, the flanking amino acids in both directions influence the binding of SH3E1 to the sequence RPKPAV. Peptide 2 of the protein KIAA1295, which is bound strongly by SH3E1, contains the sequence PQVRPKPAP which is very similar to the sequence PALRPKPAV in peptide 1 and 2 of SRGAP1. These two similar sequences can be described by the consensus PxxRPKPA with the first and second x representing a hydrophilic and hydrophobic amino acid, respectively. This motif does not correspond to any endophilin-1 binding motif described so far.

The peptide scan of the hypothetical protein C18ORF11 revealed two regions from amino acid 527-544 (peptide 2, 3) and from 714-731 (peptide 7, 8) bound by SH3E1. The motif ALPPRAP is flanked by two glutamates and is not recognized in peptide 9. The strongest signals in the region 527-544 can be delineated to the sequence NAPPVPPR that contains a class I and a class II motif. Peptides 4-6 which were detected with moderate intensity, all contain the class II motif PSIPPR. Therefore, it is most likely that C18ORF11 is bound *via* the class I motif TSCPALP (719-725) or the sequence NAPPVPPR (531-538) or both.

A peptide scan for the pulldown negative protein NAB2 detected three peptides with an increased signal intensity, whereas for peptide 4 only a weak signal was detected. Peptide 4 of NAB2 contains the class I motif PCVPAKP and the class II motif PCVPAK, suggesting that the class I motif RLSPCVP present in peptide 2 and 3 is bound by SH3E1 in NAB2. Peptide 1, detected more weakly than peptide 2 and 3, contains a truncation of this motif (RLSPC).

A peptide of the protein FLJ10101 (360-373) was recognized by GST-SH3E1, whereas a second overlapping peptide bearing the same potential binding site (356-369) was not. This finding can be deduced to the acidic Glu357 in FLJ10101.

5.4.3 Correlation of results from filter overlay and pulldown assay

Peptide scans of the two pulldown negative proteins neurogenic differentiation factor 6 (NEUROD6) and hypothetical kinesin motor domain containing protein (LOC146909) detected no signals on the peptide array. This finding demonstrates the necessity to verify results from filter overlays by means of an independent method such as a pulldown assay.

In contrast with results obtained in pulldown experiments, a peptide scan for CRP2 could not identify any 14mer which specifically binds to the SH3 domain of endophilin-1. CRP2 contains two LIM domains at either termini, connected by an unstructured region, each consisting of two independent zinc fingers that function as a protein interaction interface capable of recognizing different specific targets (Weiskirchen and Günther 2003). It is possible that the C-terminal LIM domain recognizes the SH3 domain of endophilin-1 rather than vice versa. The N-terminal LIM domain can be excluded since the N-terminus is not present in one of the three identified clones expressing CRP2.

5.4.4 Comparison to a proteome wide peptide scanning approach with the SH3 domain of endophilin-1

Recently, an approach to identify binding motifs for SH3 domains by an whole interactome peptide scanning experiment (WISE) was undertaken (Landgraf *et al*., 2004). SH3 consensus motifs were used to screen the SwissProt/TrEMBL database for proteins bearing peptide sequences potentially interacting with the SH3 domain of endophilin-1. 2032 such peptides were synthesized on an array and examined for their binding to SH3E1. Peptides of the proteins Alix and KIAA1295, also identified in this study, were detected with high signal intensity, while HMG-1 and CIC were detected with moderate and NAB2 and APBB1 with weak signal intensity in the WISE approach. The low signal intensity of the peptide delineated from NAB2 could be explained by the presence of glutamate residues, which probably weakened the binding of SH3E1 in the WISE screen. For NAB2 and HMG-1, this study failed to demonstrate binding in the pulldown assay. Most likely, the interaction of NAB2 and HMG-1 with GST-SH3E1 is weak and does not survive the stringent washing in the pulldown assay.

A binding site in the central region of CIC was identified by Landgraf *et al.* which was not present in the clone detected in this study. A distinct potential binding site in the

C-terminal fragment of CIC was found in this study which supports the observation that CIC is a potential binding partner of endophilin-1. Binding sites of several other binding partners detected in this study had not been included in the screen of Landgraf *et al*. (DAP4, MAP1LC3A, CIC, C18ORF11 and SRGAP1). This may be explained through a more stringent selection of peptides in the WISE approach compared to the more relaxed selection of peptides in this study.

5.5 Physiological relevance of identified interactions with the SH3 domain of endophilin-1

Three endophilins constitute a family of SH3 domain containing proteins. The three endophilins contain a 260 amino acid long conserved N-terminal domain, a C-terminal SH3 domain and a short hinge region (Giachino *et al*., 1997). Although endophilin-1 possesses a nuclear localisation motif, it shows a punctate and diffuse cytosolic staining pattern, enriched in presynaptic neurons (Micheva *et al*., 1997). Endophilin-1, was first described because of its interaction with synaptojanin and dynamin and colocalizes with dynamin, synaptojanin and amphiphysin in the nerve terminal cytoplasm (Ringstad *et al*., 1997).

Endophilin-1 regulates synaptic vesicle formation and participates in invagination, fission and vesicle recycling (Hannah *et al*., 1999). The function of endophilin-1 is realized by its SH3 and N-terminal domains. It has been shown that the SH3 domain is required for interaction with synaptojanin-1, for the assembly with dynamin in rings around necks of clathrin-coated pits and for membrane scission and vesicle release. The N-terminal part of endophilin-1 has lysophosphatidic acid acyl transferase (LPAAT) activity by which arachidonate transfer to lysophosphatidic acid generates phosphatidic acid (Farsad *et al*., 2001). In addition to LPAAT activity, the SH3 domain of endophilin-1 is still critical for synaptic-like microvesicle biogenesis (Schmidt *et al*., 1999).

In addition to its role in synaptic vesicle formation, endophilin-1 interacts with numerous non-endocytic proteins. The SH3 domain of endophilin-1 binds to a 24 aa proline-rich domain (PRD) in the third intracellular loop of the G-protein-coupled ß1-adrenergic receptor (Tang *et al*., 1999). It also binds to a PRD within the cytoplasmic tail of metalloprotease disintegrins, transmembrane glycoproteins acting in cell adhesion and growth factor signaling (Howard *et al*., 1999). Rat germinal centre kinase-like kinase

(rGLK), a serine/ threonine cytosolic kinase, interacts with endophilin-1 (Ramjaun *et al*., 2001). rGLK modulates c-Jun N-terminal kinase (JNK) activity by phosphorylation and binds the SH3 domain of endophilin-1 through a C-terminal PRD. Thus, endophilin-1 may help to recruit and localize intracellular signaling components. The SH3 domain of endophilin-1 also binds Cbl-interacting protein of 85 kDa (CIN85). This complex, when recruited by Cbl, influences internalisation, degradation and intracellular signaling of tyrosine kinase receptors for hepatocyte (Petrelli *et al*., 2002) and epidermal growth factors (Soubeyran *et al*., 2002). In summary, endophilin-1 interacts with endocytic and non-endocytic proteins and regulates synaptic vesicle formation as well as intracellular signaling pathways, perhaps by controlling activation and compartmentalisation. Endophilin-1 is a multifunctional protein that has different functions depending on the interaction partners of its SH3 domain, and its location within the cell.

The SH3-mediated binding of endophilin-1 to the apoptosis-linked gene 2 (ALG-2)-interacting protein X (Alix), a cytoplasmic protein ubiquitously expressed, has already been described (Chatellard-Causse *et al*., 2002). Binding of endophilin-1 to Alix was confirmed in this study. Alix may regulate apoptosis since it binds ALG-2, a Ca^{2+}-binding protein necessary for cell death. Overexpression of Alix leads to cytoplasmic vacuolization into tubulo-vesicular structures. This vacuolization phenomenon was greatly enhanced upon co-expression with endophilin-1 and may be part of the protecting mechanism afforded by Alix.

The screening strategy applied in this work has discovered several interaction partners for endophilin-1 which are already known to be involved in endocytic processes or in signal transduction pathways but were not yet identified as physiological binding partners for endophilin-1. The Slit-Robo Rho GTPase activating protein 1 (SRGAP1) participates in signal transduction from the cell membrane to the cytoskeleton during neuronal migration together with the transmembrane receptor Roundabout (Robo) and the Slit protein (Wong *et al*., 2001). SRGAP1 contains one SH3 domain which mediates binding to Robo. Thus, it is likely that SRGAP1 acts as an adapter protein. The disks large-associated protein 4 (DAP4) was also identified as novel interaction partner for endophilin-1. A potential binding site of endophilin-1 was mapped to the C-terminus of DAP-4. DAP4 belongs to the family of PSD-95/SAP90 associated proteins and is an adapter protein that plays a role in the molecular organization of synapses and in neuronal cell signaling (Takeuchi *et al*., 1997). The protein contains four proline-rich motifs,

indicating that this protein is a potential interaction partner for SH3 domains. The microtubule-associated protein light chain 3A (MAP1LC3A) was able to bind to endophilin-1. Three different light chains with distinctive functions were characterized, which predominantly bind to the heavy chain of MAP1A. It has been shown that MAP1A binds to microtubules and microfilaments. MAP1A is supposed to mediate the interaction of microtubules and actin filaments during neuronal morphogenesis (Toegel *et al*., 1998). The light chains of microtubule-associated proteins are found to be expressed at levels exceeding the level which can be bound by the heavy chains (Mei *et al*., 2000). There is strong evidence that the light chains of MAP1A have additional functions as potential linkers of neuronal microtubules and microfilaments (Noiges *et al*., 2002).

Several other proteins of yet unknown function were found to bind the SH3 domain of endophilin-1. The protein KIAA1295 ranked among the interaction partners with the highest affinity for the SH3 domain of endophilin-1. KIAA1295 contains two SH3 domains at either termini, which leads to the conclusion that KIAA1295 is involved in signaling mediated through SH3 domains. The binding of KIAA1295 to the SH3 domain of endophilin-1 may be a strong evidence for this conclusion.

Binding of endophilin-1 to the cysteine-/serine rich protein 2 (CRP2) seems to be not dependent on a proline-rich motif. No linear peptide epitope in CRP2 was detected in the peptide scan specific for SH3E1. This finding supports the idea that CRP2 recognized a region within the SH3 domain of endophilin rather than vice versa. Since a C-terminal fragment of CRP2 was also able to bind the SH3 domain of endophilin-1 in pulldowns, the binding site is supposed to be at the C-terminus of CRP2. The protein CRP2 contains one LIM domain at either termini, which represent completely independent folded units bridged by a flexible linker region (Konrat *et al*., 1998). Each LIM domain consist of a double zinc finger motif that functions as a protein interaction interface. It is evident, that CRP2 is involved in promoting protein assembly along the actin-based cytoskeleton and is of fundamental importance for distinctive functions in cell differentiation, cytoskeletal remodeling and transcriptional regulation (Weiskirchen and Günther, 2003).

The SH3 domain of endophilin-1 is capable to bind a diverse set of signaling components including other SH3 domain-containing proteins as demonstrated in this study for DAP4, SRGAP1 and KIAA1295. It is becoming increasingly apparent that biological signaling pathways can interact with one another to form complex networks comprising a large number of components. Such complexity arises from the overlapping

functions of components and from the connections among them (Weng *et al*., 1999). Many cellular processes are performed and regulated not by individual proteins, but by proteins acting in large protein assemblies or macromolecular complexes.

5.6 Conclusion

The *in vitro* strategy presented here complements existing techniques to study protein-protein-interactions on a genome-wide scale. The *in vitro* techniques employed in this study provide quick and simple yes or no answers for the examined interactions. Screening high-density protein arrays with a radiolabeled probe and confirming the potential interactions by means of an independent pulldown assay constitutes an efficient strategy in the search for novel protein-protein-interactions. Through combination of pulldown assays with peptide scans for individual binding proteins, this study confirmed the SH3 domain-dependent interaction of Alix to endophilin-1 and was able to identify CIC, DAP4, MAP1ALC3A, SRGAP1, CRP2, GLI3, HNRPA1 and the hypothetical proteins C18ORF11, FLJ10101 and KIAA1295 as novel SH3 domain binders for endophilin-1.

6 Summary

An *in vitro* screening method was developed which allows to identify protein-protein-interactions on protein arrays by overlaying with a radiolabeled protein probe. First, various PCR products were generated which code for proteins that are believed to play a role in neurodegenerative disorders. Among them were several proteins which have been reported previously to interact with each other. In a proof-of-principle approach 96 His-tagged proteins were immobilized on a PVDF membrane and were assayed for their interaction with radiolabeled GST fusion protein of the C-terminus of 70 kDa heat shock cognate protein (Hsc70), the full length heat shock organizing protein (HOP) and the SH3 domain of endophilin-1 (SH3E1). Interaction of Hsc70 with HOP as well as SH3E1 with amphiphysin-1 was demonstrated in this proof-of-principle approach. To discover novel interaction partners for the SH3E1, the developed screening method was extended to a genome-wide scale using high-density protein arrays from a human fetal brain cDNA expression library. The high-density protein array consisted of 36,864 individual clones spotted in duplicates by a robot on two PVDF membranes. Robotic spotting was performed at the German Resource Center RZPD. To distinguish false positive signals, a second high-density protein array was overlayed with 60 kDa heat shock protein (p60) as a negative control. The screening approach identified 30 clones from the cDNA library which expressed a protein in frame with the vector-encoded His-tag and were specific for SH3E1. One clone expressed a C-terminal fragment of (ALG-2)-interacting protein 1 (Alix), a previously reported interaction partner of endophilin-1. His-tagged proteins were purified on NiNTA agarose beads and used in a pulldown assay to confirm the binding of SH3E1 to the discovered interaction partners *in vitro*. Possible binding regions of SH3E1 were mapped, using scans of peptides delineated from the identified interaction partners. Combination of the overlay screen with independent pulldowns and peptide scans confirmed the interaction of SH3E1 with Alix, disks large-associated protein 4 (DAP4), cysteine and glycine-rich protein 2 (CRP2), capicua homolog (CIC), heterogenous nuclear ribonucleoprotein A1 (HNRPA1), microtubule-associated protein 1A light chain 3A (MAP1ALC3), SLIT-ROBO Rho GTPase activating protein 1 (SRGAP1), Gli-Krüppel family member (GLI3) and the hypothetical proteins FLJ10101, C18ORF11 and KIAA1295 and mapped possible binding regions for Alix, DAP4, CIC, MAP1LC3A, SRGAP1, FLJ10101, C18ORF11, KIAA1295 and NGFI-A binding protein 2 (NAB2).

7 References

Aruffo, A. **(1997)** Transient expression of proteins using COS cells. *In* "Current Protocols in Molecular Biology" (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., *et al.,* Eds.), pp. 16.13.11–16.13.17, Wiley, New York.

Auerbach, D., Thaminy, S., Hottiger, M.O. and Stagljar, I. **(2002)** The post-genomic era of interactive proteomics: Facts and perspectives. *Proteomics* 2, 611-623.

Ayer, D.E., Kretzner, L. and Eisenman, R.N. **(1993)** Mad: A heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* 72, 211-222.

Ball, L.J., Kuhne, R., Hoffmann, B., Hafner, H., Schmieder, P. *et al.* **(2000)** Dual epitope recognition by the VASP EVH1 domain modulates polyproline ligand specificity and binding affinity. *EMBO J.* 19, 4903-4914

Bartel, P.L., Roecklein, J.A., SenGupta, D. and Fields, S. **(1996)** A protein linkage map of *Escherichia coli* bacteriophage T7. *Nature Genet.* 12, 72-77.

Bedouelle, H. and DuPlay, P. **(1988)** Production in *Escherichia coli* and one-step purification of bifunctional hybrid proteins which bind maltose. *Eur. J. Biochem.* 171, 541-549.

Blackwood, E.M. and Eisenman, R.N. **(1991)** Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 251, 1211-1217.

Blanar, M.A. and Rutter, W.J. **(1992)** Interaction cloning: Identification of a helix-loop-helix ziper protein that interacts with c-Fos. *Science* 256, 1014-1018.

Boeckmann, B., Bairoch, A., Appweiler, R., Blatter, M.C., Estreicher, A. *et al.* **(2003)** The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 31, 365-370.

Bondos, S.E. and Bicknell, A. **(2003)** Detection and prevention of protein aggregation before, during, and after purification. *Anal. Biochem.* 316, 223-231.

Brinker, A., Scheufler, C., von der Mülbe, F., Fleckenstein, B., Herrmann, C. *et al*. **(2002)** Ligand discrimination by TPR domains: Relevance and selectivity of EEVD-recognition in Hsp70-Hop-Hsp90 complexes. *J. Biol. Chem*. 277, 19265-19275.

Buckholz, R. G. and Gleeson, M. A. **(1991)** Yeast systems for the commercial production of heterologous proteins. *Bio/Technology* 9, 1067–1072.

Büssow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H. and Walter, G. **(1998)** A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* 26, 5007-5008.

Büssow, K., Nordhoff, E., Lübbert, C., Lehrach, H. and Walter, G. **(2000)** A human cDNA library for high-throughput protein expression screening. *Genomics* 65, 1-8.

Cestra, G., Castagnoli, L., Dente, L., Minenkova, O., Petrelli, A. *et al.* **(1999)** The SH3 domains of endophilin and amphiphysin bind to the proline-rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. *J. Biol. Chem.* 274, 32001-32007.

Chatellard-Causse, C., Blot, B., Cristina, N., Torch, S., Missotten, M. *et al.* **(2002)** Alix (ALG-2-interacting Protein X), a protein involved in apoptosis, binds to endophilins and induces cytoplasmic vacuolization. *J. Biol. Chem.* 277, 29108–29115.

Chen, S., Prapapanich, V., Rimerman, R.A., Honore, B. and Smith, D.F. **(1996)** Interactions of p60, a mediator of progesterone receptor assembly, with heat shock proteins hsp90 and hsp70. *Mol. Endocrinol.* 6, 682-693.

Claverie, J.M. **(2001)** Gene number. What if there are only 30,000 human genes? *Science* 291, 1255-1257.

Crowther, J.R. **(2001)** The ELISA guide book. 1st ed., Humana Press.

de Graaf, K., Hekerman, P., Spelten, O., Herrmann, A., Packman, L.C. *et al.* **(2004)** Characterization of cyclin L2, a novel cyclin with an arginine/serine-rich domain. *J. Biol. Chem.* 279, 4612-4624.

De Wildt, R.M., Mundy, C.R., Gorick, B.D. and Tomlinson, I.M. **(2000)** Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nature Biotechnol.* 18, 989-994.

Du, W., Vidal, M., Xie, J.E. and N. Dyson, N. **(1996)** *RBF*, a novel RB-related gene that regulates E2F activity and interacts with *cyclin E* in *Drosophila. Genes Dev.* 10, 1206–1218.

Einarson, M.B. and Chao, M.V. **(1995)** Regulation of Id1 and its association with basic helix-loop-helix proteins during nerve growth factor-induced differentiation of PC12 cells. *Mol. Cell. Biol.* 15, 4175-4183.

Espejo, A., Cote, J., Bednarek, A., Richard, S. and Bedford, M.T. **(2002)** A protein-domain microarray identifies novel protein-protein interactions. *Biochem. J.* 367, 697-702.

Ewing, B. and Green, P. **(2000)** Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genetics* 25, 232-234.

Farsad, K., Ringstad, N., Takei, K., Floyd, S.R., Rose, K. *et al.* **(2001)** Generation of high curvature membranes mediated by direct endophilin bilayer interactions. *J. Cell. Biol.* 155, 193–200.

Feng, S., Chen, J.K., Yu, H., Simon, J.A. and Schreiber, S.L. **(1994)** Two binding orientations for peptides to the src SH3 domain: development of a general model for SH3-ligand interactions. *Science* 266, 1241-1247.

Fields, S. and Song, O.A. **(1989)** A novel genetic system to detect protein-protein-interactions. *Nature* 340, 245-246.

Finley, R.L. and Brent, R. **(1994)** Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci. USA* 91, 12980-12984.

Frank, R. **(1992)** Spot synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* 48, 9217-9232.

Fromont-Racine, M., Rain, J.C. and Legrain, P. **(1997)** Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* 16, 277-282.

Fromont-Racine, M., Mayes, A.E., Brunet-Simon, A., Rain, J.C., Colley, A. *et al*. **(2000)** Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast* 17, 95-110

Flores, A., Briand, J.F., Gadal, O., Andrau, J.C., Rubbi, L. *et al*. **(1999)** A protein-protein-interaction map of yeast RNA polymerase III. *Proc. Natl. Acad. Sci. USA* 96, 7815-7820.

Gavin, A.C., Bösche, M., Krause, R., Grandi, P., Marzioch, M. *et al*. **(2002)** Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-147.

Ge, H. **(2000)** UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res.* 28, e3.

Gerstein, M., Lan, N. and Jansen, R. **(2002)** Integrating interactomes. *Science* 295, 284-287.

Giachino, C., Lantelme, E., Lanzetti, L., Saccone, S., Della Valle, G. *et al.* **(1997)** A novel SH3-containing human gene family preferentially expressed in the central nervous system. *Genomics* 41, 427–434.

Gobom, J., Schürenberg, M., Müller, M., Theiss, D., Lehrach, H. *et al*. **(2001)** α-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. *Anal. Chem*. 73, 434-438.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B. *et al*. (**1996)** Life with 6,000 genes. *Science* 274, 563-567.

Gottesman, S. **(1996)** Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet*. 30, 465-506.

Guan, C., Li, P., Riggs, P.D. and Inouye, H. **(1988)** Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose binding protein. *Gene* 67, 21-30.

Haab, B.B., Dunham, M.J. and Brown, P.O. **(2001)** Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2, 1-13.

Hannah, M.J., Schmidt, A.A. and Huttner, W.B. **(1999)** Synaptic vesicle biogenesis. *Annu. Rev. Cell Dev. Biol*. 15, 733-798.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L. *et al*. **(2002)** Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180-183.

Hochuli, E., Dobeli, A. and Schacher, A. **(1987)** New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* 411, 177-184.

Hockney, R. C. **(1994)** Recent developments in heterologous protein production in *Escherichia coli. Trends Biotechnol.* 12, 456–463.

Holt, L.J., Büssow, K., Walter, G. and Tomlinson, I.M. **(2000)** By-passing selection: direct screening for antibody-antigen interactions using protein arrays. *Nucleic Acids Res.* 28, E72.

Howard, L., Nelson, K.K., Maciewicz, R.A. and Blobel, C.P. **(1999)** Interaction of the metalloprotease disinetgrins MDC9 and MDC15 with two SH3 domain-containing proteins, endophilin-1 and SH3PX1. *J. Biol. Chem.* 274, 31693-31699.

Hu, G., Zhang, S., Vidal, M., Labaer, J., Xu, T. and Fearon, E.R. **(1997)** Mammalian homologs of *seven in absentia* regulate DCC via the ubiquitin-proteasome pathway. *Genes Dev.* 11**,** 2701–2714.

Hudson, J.R., Dawson, E.P., Rushing, K.L., Jackson, C.H., Lockshon, D. *et al*. **(1997)** The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res.* 7, 1169-1173.

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T. *et al*. **(2000)** Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA* 97, 1143-1147.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. **(2001)** A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* 98, 4569-4574.

Jay, D., Garcia, E.J., Lara, J.E., Medina, M.A. and Ibarra, M. **(2000)** Determination of a cAMP-dependent protein kinase phosphorylation site in the C-terminal region of human endothelial actin-binding protein. *Arch. Biochem. Biophys.* 377, 80-84.

Kaelin, W.G., Pallas, D.C., DeCaprio, J.A., Kaye, F.J. and Livingston, D.M. **(1991)** Identification of cellular proteins that can interact specifically with the T/E1A binding region of the retinoblastoma gene product. *Cell* 64, 521-532.

Kingston, R. E., Kaufman, R. J., Bebbington, C. R. and Rolfe, M. R. **(1997)** Amplification using CHO cell expression vectors. *In* "Current Protocols in Molecular Biology" (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., *et al.,* Eds.), p. 16.14.11, Wiley, New York.

Kodadek, T. **(2001)** Protein microarrays: prospects and problems. *Chem. Biol.* 8, 105-115.

Konrat, R., Kräutler, B., Weiskirchen, R. and Bister, K. **(1998)** Structure of cysteine- and glycine-rich protein CRP2. Backbone dynamics reveal motional freedom and independent spatial orientation of the LIM domains. *J. Biol. Chem.* 273, 23233-23240.

Kramer, A., Reineke, U., Dong, L., Hoffmann, B., Hoffmüller, U. *et al.* **(1999)** Spot synthesis: observations and optimizations. *J. Peptide Res.* 54, 319-327.

Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J. *et al*. **(2004)** Protein interaction networks by proteome peptide scanning. *PloS Biology* 2, 94-103

Lee, J. and Bedford, M.T. **(2002)** PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO Reports* 3, 268-273.

Lim, W.A., Richards, F.M. and Fox, R.O. **(1994)** Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* 372, 375-379.

Lockhart, D.J. and Winzeler, E.A. **(2000)** Genomics, gene expression and DNA arrays. *Nature* 405, 827-836.

Ma, J. and Ptashne, M. **(1987)** A new class of yeast transcriptional activators. *Cell* 51**,** 113–119.

MacBeath, G. and Schreiber, S.L. **(2000)** Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760-1763.

MacBeath, G. **(2002)** Protein microarrays and proteomics. *Nature Genetics* 32, 526-532.

Macgregor, P.F., Abate, C. and Curran, T. **(1990)** Direct cloning of leucine zipper proteins: Jun binds cooperatively to the CRE with CRE-BP1. *Oncogene* 5, 451-458.

Mahlknecht, U., Ottmann, O.G. and Hoelzel, D. **(2001)** Far-Western based protein-protein interaction screening of high-density protein filter arrays. *J. Biotechnol.* 88, 89-94.

Maina, C.V., Riggs, P.D., Grandea, A.G. III, Slatko, B.E., Moran, L.S. *et al*. **(1988)** A vector to express and purify foreign proteins in *Escherichia coli* by fusion to, and separation from, maltose binding protein. *Gene* 74, 365-373.

Makrides, S. C. **(1996)** Strategies for achieving high-level expression of genes in *Escherichia coli. Microbiol. Rev.* 60, 512–538.

Mei, X., Sweatt, A.J. and Hammarback, J.A. **(2000)** Regulation of microtubule-associated protein 1B (MAP1B) subunit composition. *J. Neurosci. Res.* 62, 56-64.

Micheva, K.D., Kay, B.K. and McPherson, P.S. **(1997)** Synaptojanin forms two separate complexes in the nerve terminal: interactions with endophilin and amphiphysin. *J. Biol. Chem.* 272, 27239–27245.

Miller, L. K. **(1993)** Baculoviruses: High-level expression in insect cell. *Curr. Opin. Genet. Dev.* 3, 97–101.

Mitchell, P.A. **(2002)** A perspective on protein microarrays. *Nature Biotechnol.* 20, 225-229.

Musacchio, A., Gibbs, T., Lehto, V.P. and Sarasate, M. **(1992)** SH3- - an abundant protein domain in search of a function. *FEBS Lett*. 307, 55-61.

Noiges, R., Eichinger, R., Kutschera, W., Fischer, I., Németh, Z. *et al.* **(2002)** Microtubule-associated protein 1A (MAP1A) and MAP1B: Light chains determine distinct functional properties. *J. Neurosci.* 22, 2106-2114.

Obata, T., Yaffe, M.B., Leparc, G.G., Piro, E.T., Maegawa, H. *et al*. **(2000)** Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J. Biol. Chem*. 275, 36108-36115.

Pawson, T. **(1995)** Protein modules and signalling networks. *Nature* 373, 573-580.

Pawson, T. and Scott, J.D. **(1997)** Signaling through scaffold, anchoring, and adaptor proteins. *Science* 278, 2075–2080.

Pawson,T. Raina, M. and Nash, P. **(2002)** Interaction domains: From simple binding events to complex cellular behavior. *FEBS Lett.* 513, 2–10.

Pawson, T. and Nash, P. **(2003)** Assembly of cell regulatory systems through protein interaction domains. *Science* 300, 445–452.

Petrelli, A., Gilestro, G.F., Lanzardo, S., Comoglio, P.M., Migone, N. *et al*. **(2002)** The endophilin-CIN85-Cbl complex mediates ligand-dependent downregulation of c-Met. *Nature* 416, 187-190.

Phizicky, E.M. and Fields, S. **(1995)** Protein-protein-interactions: Methods for detection and analysis. *Microbiol. Rev.* 59, 94-123.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E. *et al*. **(2001)** The tandem affinity purification (tap) method: a general procedure of protein complex purification. *Methods* 24, 218-229.

Puntervoll, P., Linding, R., Gemünd, C., Chabanis-Davidson, S., Mattingsdal, M. *et al.* **(2003)** ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* 31, 3625-3630.

Ramjaun, A.R., Angers, A., Legendre-Guillemin, V., Tong, X.K. and McPherson, P.S. **(2001)** Endophilin regulates JNK activation through its interaction with the germinal center kinase-like kinase. *J. Biol. Chem.* 276, 28913–28919.

Reineke, U., Sabat, R., Volk, D. and Schneider-Mergener, J. **(1998)** Mapping of the interleukin-10/interleukin-10 receptor combining site. *Protein Sci.* 7, 951-960.

Reineke, U., Volkmer-Engert, R. and Schneider-Mergener, J. **(2001)** Applications of peptide arrays prepared by the SPOT-technology. *Curr. Opin. Biotechnol.* 12, 59-64.

Reutens, A.T. and Begley, C.G. **(2002)** Endophilin-1: a multifunctional protein. *Int. J. Biochem. Cell Biol.* 34, 1173-1177.

Rigaut, G., Shevchenko, A, Rutz, B., Wilm, M., Mann, M. and Séraphin, B. **(1999)** A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* 17, 1030-1032.

Ringstad, N., Nemoto, Y. and De Camilli, P. **(1997)** The SH3p4/SH3p8/SH3p13 protein family: binding partners for synaptojanin and dynamin via a Grb2-like Src homology 3 domain. *Proc. Natl. Acad. Sci. U S A* 94, 8569-8574.

Rüdiger, S., Schneider-Mergener, J. and Bukau, B. **(2001)** Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J.* 20, 1042-1050.

Schenk, P.M., Baumann, S., Mattes, R. and Steinbiß, H.H. **(1995)** Improved high-level expression system for eukaryotic genes in Escherichia coli using T7 RNA polymerase and rare (Arg)tRNAs. *BioTechniques* 19, 196-198.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L. *et al*. **(2000)** Structure of TPR domain-peptide complexes: Critical elements in the assembly of the Hsp70-Hsp90 multichaperone maschine. *Cell* 101, 199-210.

Schmidt, A., Wolde, M., Thiele, C., Fest, W., Kratzin, H. *et al*. **(1999)** Endophilin-1 mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature* 401, 133-141.

Smit, A.F. **(1996)** The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* 6, 743-748.

Smith, D.F., Sullivan, W.P., Marion, T.N., Zaitsu, K., Madden, B. *et al*. **(1993)** Identification of a 60-kilodalton stress-related protein, p60, which interacts with hsp90 and hsp70. *Mol. Cell. Biol.* 13, 869-876.

Smith, D.B. and Johnson, K.S. **(1988)** Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. *Gene* 67, 31-40.

Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W.Y. and Dikic, I. **(2002)** Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors. *Nature* 416, 183-187.

Takeuchi, M., Hata, Y., Hirao, K., Toyoda, A., Irie, M. *et al.* **(1997)** SAPAPs: A family of PSD-95/SAP90-associated proteins localized at postsynaptic density. *J. Biol. Chem.* 272 11943-11951.

Tanaka, A., Gibbs, C.P., Arthur, R.R., Anderson, S.K., Kung, H.J. *et. al.* **(1987)** DNA sequence encoding the amino-terminal region of the human c-src protein: implications of sequence divergence among src-type kinase oncogenes. *Mol. Cell. Biol.* 7, 1978-1983.

Tang, Y., Hu, L.A., Miller, W.E., Ringstad, N., Hall, R.A. *et al.* **(1999)** Identification of the endophilins (SH3p4/p8/p13) as novel binding partners for the ß1-adrenergic receptor. *Proc. Natl. Acad. Sci. U S A* 96, 12559-12564.

Tögel, M., Wiche, G. and Probst, T. **(1998)** Novel features of the light chain of micortubule-associated protein MAP1B: microtubule stabilization, self interaction, actin filament binding, and regulation by the heavy chain. *J. Cell. Biol.* 143, 695-707.

Töpert, F., Pires, R., Landgraf, C., Oschkinat, H. and Schneider-Mergener, J. **(2001)** Synthesis of an array comprising 837 variants of the hYAP WW protein domain. *Angew. Chem.-Int. Edit. Engl.* 40, 897-900.

Tucker, C.L., Gera, J.F. and Uetz, P. **(2001)**. Towards an understanding of complex protein networks. *Trends Cell Biol.* 11, 102-106.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S. *et al*. **(2000)** A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627.

Vidal, M. **(2001)**. A biological atlas of functional maps. *Cell* 104, 333-339.

von Mehring, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G. *et al*. **(2002)** Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399-403.

Wagner, L.A., Weiss, R.B., Driscoll, R., Dunn, D.S. and Gesteland, R.F. **(1990)** Transcriptional slippage occurs during elongation at runs of adenin or thymine in *Escherichia coli*. *Nucleic Acids Res*. 18, 3529-3535.

Washburn, M.P., Wolters, D. and Yates, J.R. III **(2001)** Large scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* 19, 242-247.

Weiergräber, O., Schneider-Mergener, J., Grötzinger, J., Wollmer, A., Küster, A. *et al*. **(1996)** Use of immobilized synthetic peptides for the identification of contact sites between human interleukin-6 and its receptor. *FEBS Lett.* 379, 122-126.

Weiner, H., Faupel, T. and Büssow, K. **(2004)** Protein arrays from cDNA expression libraries. *Methods Mol. Biol.* 278, 1-14.

Weiskirchen, R. and Günther, K. **(2003)** The CRP/MLP/TLP family of LIM domain containing proteins: acting by connecting. *BioEssays* 25, 152-162.
Weng, G., Bhalla, U.S. and Iyenger, R. **(1999)** Complexity in biological signaling systems. *Science* 284, 92-96.

Wong, K., Ren, X.R., Huang, Y.Z., Xie, Y., Liu, G. *et al*. **(2001)** Signal transduction in neuronal migration: Roles of GTPase activating proteins and the small GTPase Cdc42 in the Slit-Robo pathway. *Cell* 107, 209-221.

Woo, M.S., Ohta, Y., Rabinovitz, I., Stossel, T.P. and Blenis, J. **(2004)** Ribosomal S6 kinase (RSK) regulates phosphorylation of filamin A on an important regulatory site. *Mol. Cell. Biol*. 24, 3025-3035.

Wruck, W., Griffiths, H., Steinfath, M., Lehrach, H., Radelof, U. *et al.* **(2002)** Xdigitise: visualization of hybridization experiments. *Bioinformatics* 18, 757-760.

Young, R.A. and Davis, R.W. **(1983)** Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. USA* 80, 1194-1198.

Young, R.A. and Davis, R.W. **(1991)** Gene isolation with lambda gt11 system. *Methods Enzymol.* 194, 230-238.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A. *et al*. **(2001)** Global analysis of protein activities using proteome chips. *Science* 293, 2101- 2105.

8 Abbreviations

aa	amino acid
APS	ammoniumpersulfate
ATP	adenosine triphosphate
dH_2O	deionized H_2O
DTT	dithiothreitol
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
g	gram
g	gravity
GST	glutathione S-transferase
hr(s)	hour(s)
HEPES	N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
IPTG	isopropylthio-ß-D-galactoside
kDa	kilo Dalton
L	liter
m	meter
M	molar (mol x l^{-1})
mA	milli Amper
mg	milligram (10^{-3}g)
µg	microrgram (10^{-6}g)
min	minute
mL	millilitre (10^{-3} l)
µl	microlitre (10^{-6} l)
µm	micrometer (10^{-6} m)
mM	millimolar (10^{-3} M)
µM	micromolar (10^{-6} M)
nm	nanometer (10^{-9}m)
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenyl methyl sulfonyl fluoride
rpm	rotation per minute
SDS	sodium dodecyl sulfate
TBS	tris buffered saline
TCA	trichloroacetic acid
TEMED	1,2-bis-(dimethylamino)-ethane
Tris	tris(hydroxymethyl)aminomethane
v/v	volume/volume
w/v	weight/volume

Publications

Weiner, H., **Faupel, T.** and Büssow, K. (2004) Protein arrays from cDNA expression libraries. Methods in Molecular Biology 264, 1-15

Faupel, T., Weiner, H. and Büssow, K. (2004) Identification of novel endophilin-1 binding proteins using high-density protein arrays. Submitted

Meeting abstracts

Faupel, T., Büssow, K. and Weiner, H. (2002) Comparison of *in vitro* techniques to detect and characterize protein-protein-interactions that play a role in neurodegenerative diseases. Oral presentation on the Symposium of the National Genome Research Network (NGFN) and the German Human Genome Project (DHGP), Berlin, Germany.

Weiner, H., **Faupel, T.** and Büssow K. (2003) A triple technique strategy to screen for protein-protein-interactions *ex vivo*. Oral presentation on the International Congress on Protein Expression and Protein Function. Berlin, Germany.

Faupel, T., Weiner, H. and Büssow, K. (2004) High throughput screening for protein-protein-interactions: Application to SH3-binding proteins. Poster contribution on the 2nd Day of Science of the Max-Planck-Institute for Molecular Genetics, Berlin, Germany.

Acknowledgements

I would like to express my gratitude to Prof. Ulf Stahl for his continous support throughout my diploma and Ph.D. studies and for agreeing to be the referee for the Faculty of Biotechnology at the Technical University Berlin.

I am particularly grateful to Konrad Büssow for providing direction and encouragment throughout my Ph.D. project. I am also grateful to him for the critical reading of this manuscript and for many helpful discussions.

I thank Hendrik Weiner who frequently gave assistance and practical advice in the laboratory and without whose expertise I would not have gained my current methodic knowledge.

I would like to thank Prof. Hans Lehrach and all the other members of the Max Planck Institute for Molecular Genetics, for creating a nice and stimulating working environment.

Many thanks go to Brigitte Hieke and Andrea Krebs for provided technical assistance and to Volker Sievert, who frequently gave assistance on electronic and computing-related matters.

I thank Johan Gobom and Dorothea Theiss, Max Planck Institute for Molecular Genetics and Martina Schnölzer, Barbara Überle and Silke Wandschneider from the DKFZ in Heidelberg for the mass spectrometric analysis of proteins.

I would like to thank Stefanie Wälter and Klaus Genser, Max Delbrück Center for Molecular Medicine, for the help in the initial stage of this Ph.D. project and Prof. Erich Wanker and Uli Stelzl, Max Delbrück Center for Molecular Medicine, for helpful discussions related to this work.

Above all, I would like to thank my family for the constant support and encouragement throughout all stages of my studies.

Appendix A

plate position	protein name	aa sequence	predicted size (kDa)	Genbank accession No.	Gene identifier	DNA sequence	Mass spectrometry
A1	amphiphysin 2	497-593	13	AAC51345	2160795	identified	
B1	c-src	83-144	10	NP_005408	4885609	identified	
C1	CHIP	143-303	22	AAD33400	4928064		identified
D1	CHIP	198-303	15	AAD33400	4928064		
E1	CHIP	2-197	25	AAD33400	4928064		identified
F1	CHIP	2-142	18	AAD33400	4928064		identified
G1	CHIP	2-303	37	AAD33400	4928064	identified	confirmed
H1	cortactin	350-550	25	NP_005222	4885205	identified	confirmed
A2	Hsc70	541-646	14	NP_006588	5729877		
B2	p20-arc	2-168	22	AAB64192	2282040		identified
C2	SH3p7	282-430	19	NP_054782	7661718	identified	
D2	SH3p7	282-430	19	NP_054782	7661718		
E2	SNAP25	2-206	26	XP_045655	14771971	identified	confirmed
F2	synapsin1A	113-420	37	XP_013120	14758353	identified	confirmed
G2	TIM44	39-452	50	XP_049282	14757910		identified
H2	myobrevin	2-75	11	AAC08434	2996192	identified	confirmed
A3	CHIP	143-303	22	AAD33400	4928064		identified
B3	CHIP	198-303	15	AAD33400	4928064		
C3	Hsc70	373-540	21	NP_006588	5729877	identified	confirmed
D3	Hsc70	373-646	32	NP_006588	5729877	identified	
E3	Hsc70	541-646	14	NP_006588	5729877	identified	
F3	Hsc70	541-646	14	NP_006588	5729877	identified	
G3	Hsc70	541-646	14	NP_006588	5729877	identified	
H3	Hsc70	541-646	14	NP_006588	5729877	identified	
A4	Hsc70	541-646	14	NP_006588	5729877	identified	
B4	auxilin	816-913	14	BAA32318	3413908	identified	
C4	endophilin 1	292-352	10	NP_003017	4506931	identified	
D4	amphiphysin 1	560-695	17	NP_001626	4502081	identified	
E4	synaptotagmin	135-422	36	NP_005630	5032139		identified
F4	synaptotagmin	265-422	21	NP_005630	5032139	identified	confirmed
G4	synaptotagmin	265-422	21	NP_005630	5032139	identified	confirmed
H4	SH3p7	282-430	19	NP_054782	7661718		
A5	SH3p7	2-281	34	NP_054782	7661718	identified	confirmed
B5	SH3p7	2-430	51	NP_054782	7661718	identified	confirmed
C5	cortactin	2-330	39	NP_005222	4885205	identified	confirmed
D5	cortactin	2-550	64	NP_005222	4885205	identified	confirmed
E5	cortactin	350-550	25	NP_005222	4885205	identified	confirmed
F5	cortactin	85-330	30	NP_005222	4885205		identified
G5	syntaxin 8	2-236	29	NP_004844	4759188		identified

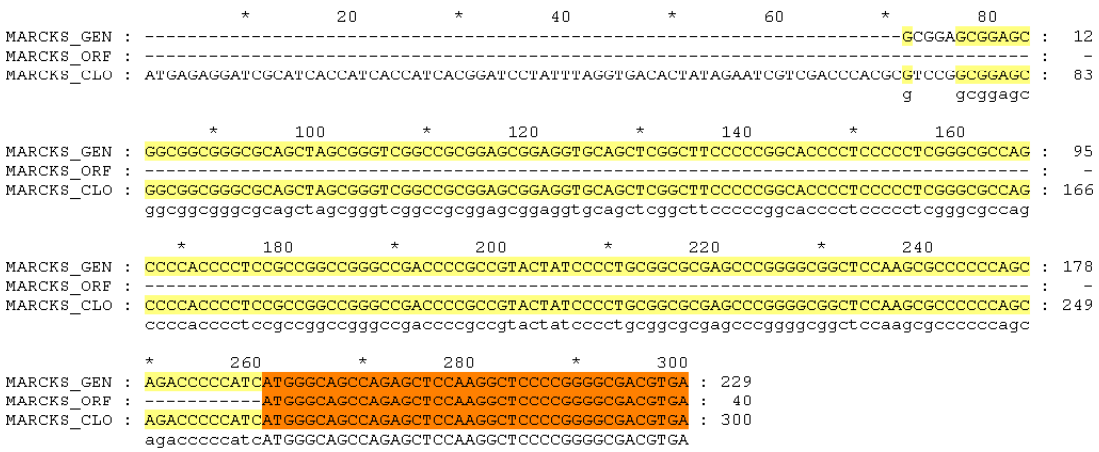
Appendix A (continued)

plate position	protein name	aa sequence	predicted size (kDa)	Genbank accession No.	Gene identifier	DNA sequence	Mass spectrometry
H5	cathepsin	2-339	40	XP_035662	14745062		
A6	centractin	2-376	45	AAH10090	14603262		identified
B6	p60	2-573	64	NP_002147	4504521	identified	confirmed
C6	syntaxin 5	2-301	37	XP_043216	14773476		
D6	syntaxin 5	2-284	35	XP_043216	14773476		identified
E6	syntaxin 12	2-269	33	XP_039018	14725420		identified
F6	dynamitin	2-406	47	AAC50423	1255188		identified
G6	dynamitin	2-87	12	AAC50423	1255188	identified	
H6	p41-arc	2-372	43	AAB64189	2282034		identified
A7	HOP	2-543	65	NP_006810	5803181	identified	confirmed
B7	Hsc70	2-540	62	NP_006588	5729877	identified	confirmed
C7	Hsc70	541-646	14	NP_006588	5729877		identified
D7	auxilin	550-913	42	BAA32318	3413908	identified	confirmed
E7	auxilin	816-913	14	BAA32318	3413908		
F7	endophilin 1	2-352	42	NP_003017	4506931	identified	confirmed
G7	endophilin 1	292-352	10	NP_003017	4506931		
H7	Hsc70	2-540	62	NP_006588	15311554	identified	confirmed
A8	amphiphysin 1	291-559	31	NP_001626	4502081	identified	
B8	filamin A	2015-2198	21	NP_001447	4503745		identified
C8	parkin	2-103	14	BAA25751	3063388	identified	confirmed
D8	parkin	220-465	30	BAA25751	3063388		
E8	parkin	2-318	38	BAA25751	3063388		
F8	parkin	304-404	14	BAA25751	3063388		
G8	parkin	32-465	51	BAA25751	3063388	identified	confirmed
H8	parkin	395-465	11	BAA25751	3063388		
A9	adaptin	319-478	21	NP_570603	19913416		identified
B9	adaptin	2-478	53	NP_570603	19913416		identified
C9	adaptin	319-478	21	NP_570603	19913416		
D9	syntaxin 7	2-261	32	XP_004526	12732634		identified
E9	syntaxin 7	2-236	30	XP_004526	12732634	identified	confirmed
F9	synbindin	2-219	27	NP_057230	7706667		identified
G9	synbindin	2-219	27	NP_057230	7706667		identified
H9	UbcH7	2-154	20	CAA04156	2739215		identified
A10	UbcH8	2-152	20	AAB86433	2623260		identified
B10	syntenin	2-298	35	XP_044627	14749519		
C10	syntenin	114-273	20	XP_044627	14749519		identified
D10	syntenin	2-273	33	XP_044627	14749519		identified
E10	syntenin	114-298	22	XP_044627	14749519	identified	confirmed
F10	Psd95	108-541	50	AAC52113	3318653		identified
G10	Rab4	2-213	26	AAA60244	550068		
H10	syntaxin 1A	2-265	33	NP_004594	4759182	identified	
A11	syntaxin 1A	2-288	36	NP_004594	4759182		identified
B11	syntaxin 1A	2-265	33	NP_004594	4759182		identified

Appendix A (continued)

plate position	protein name	aa sequence	predicted size (kDa)	Genbank accession No.	Gene identifier	DNA sequence	Mass spectrometry
C11	syntaxin 1A	2-288	36	NP_004594	4759182		identified
D11	calmodulin	2-149	18	M19311	179883	identified	confirmed
E11	phocein	4-216	26	XM_002313	13637193		identified
F11	phocein	4-216	26	XM_002313	13637193		
G11	Fyn SH3	2-142	17	NP_002028	4503823	identified	
H11	Fyn SH3	82-142	8	NP_002028	4503823	identified	confirmed
A12	Lyn SH3	2-141	18	AAA59540	307144	identified	confirmed
B12	Lyn SH3	2-121	15	AAA59540	307144	identified	confirmed
C12	Lyn SH3	62-121	8	AAA59540	307144	identified	confirmed
D12	c-src	83-218	17	NP_005408	4885609		
E12	filamin A	2041-2136	12	NP_001447	4503745	identified	
F12	filamin A	2015-2198	21	NP_001447	4503745	identified	
G12	filamin A	1644-2136	54	NP_001447	4503745	identified	confirmed
H12	filamin A	1788-2136	39	NP_001447	4503745	identified	confirmed

Appendix B



Artificial polypeptide (*MARCKS*, AL713653)

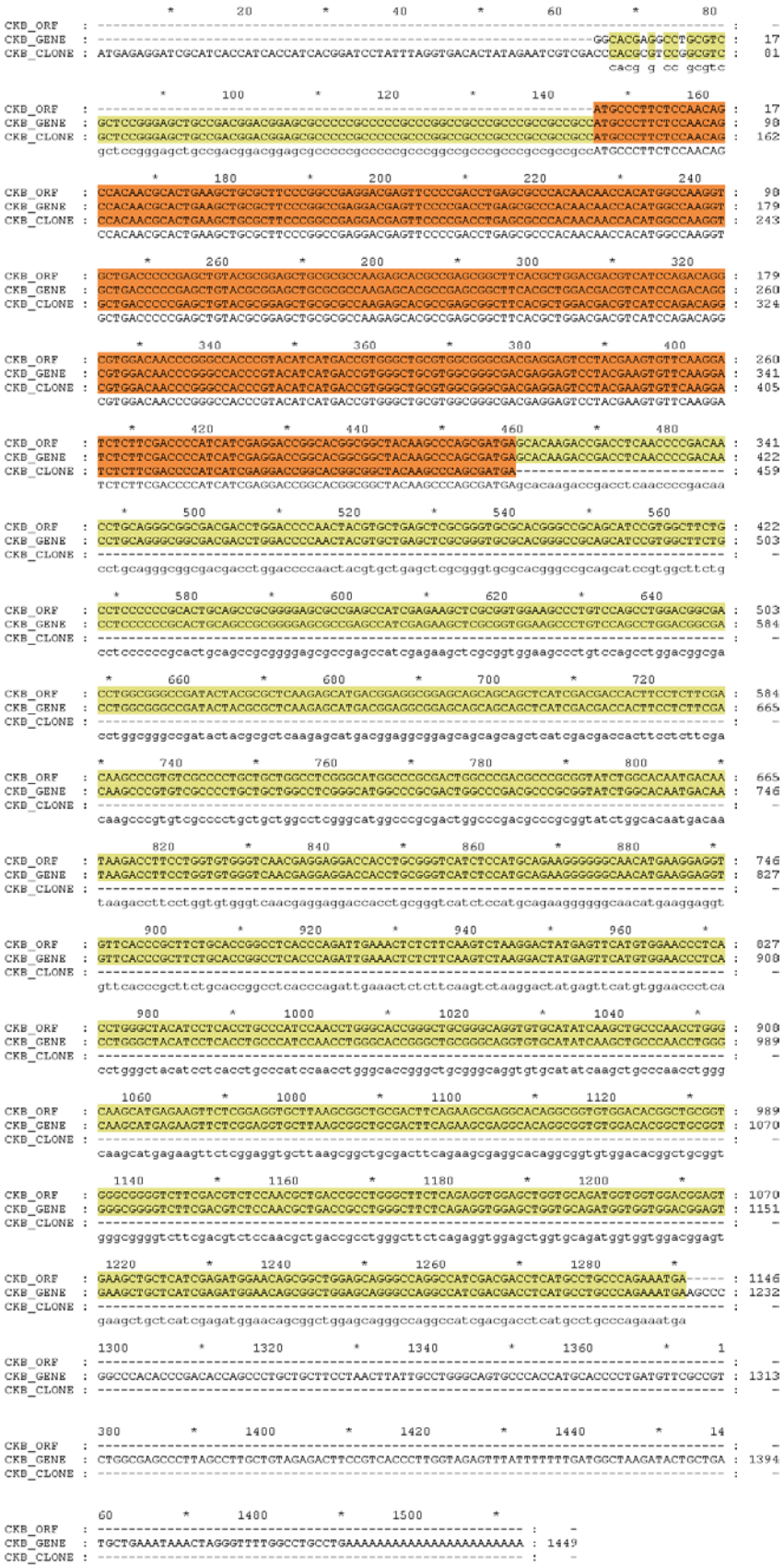
Predicted molecular weight: 10,057 Dalton

MRGSHHHHHHGSYLGDTIESSTHASGGAAAGAASGSAAER

RCSSASPGTPPPRAPAPPLRRPGRPRRTIPCGASPGRLQA

PPSRPPSWAARAPRLPGAT

Appendix C



Appendix C (continued)

Artificial polypeptide (*CKB*, BC010002)

Predicted molecular weight: 16,633 Dalton

MRGSHHHHHHGSYLGDTIESSTHASGVAPGAADGRSAPAPARPPARRRHALLQQPQRTEA

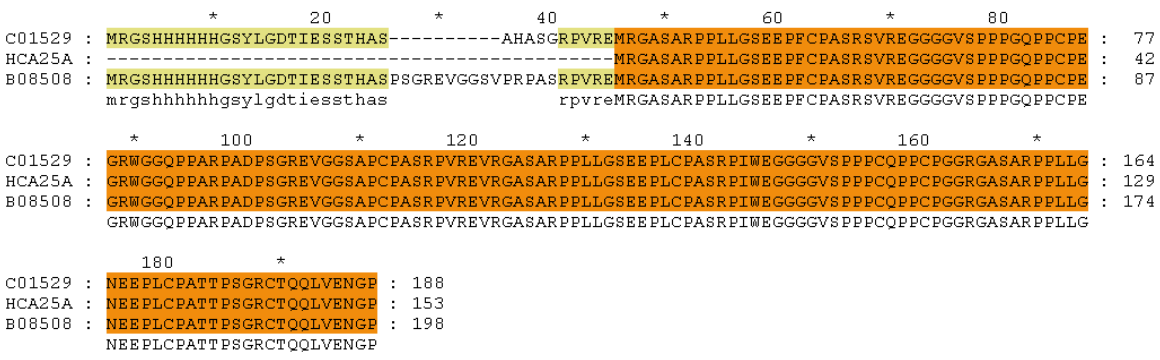
ALPGRGRVPRPERPQQPHGQGADPRAVRGAARQEHAERLHAGRRHPDRRGQPGPPVHHDR

GLRGGRRGVLRSVQGSLRPHHRGPARRLQAQR

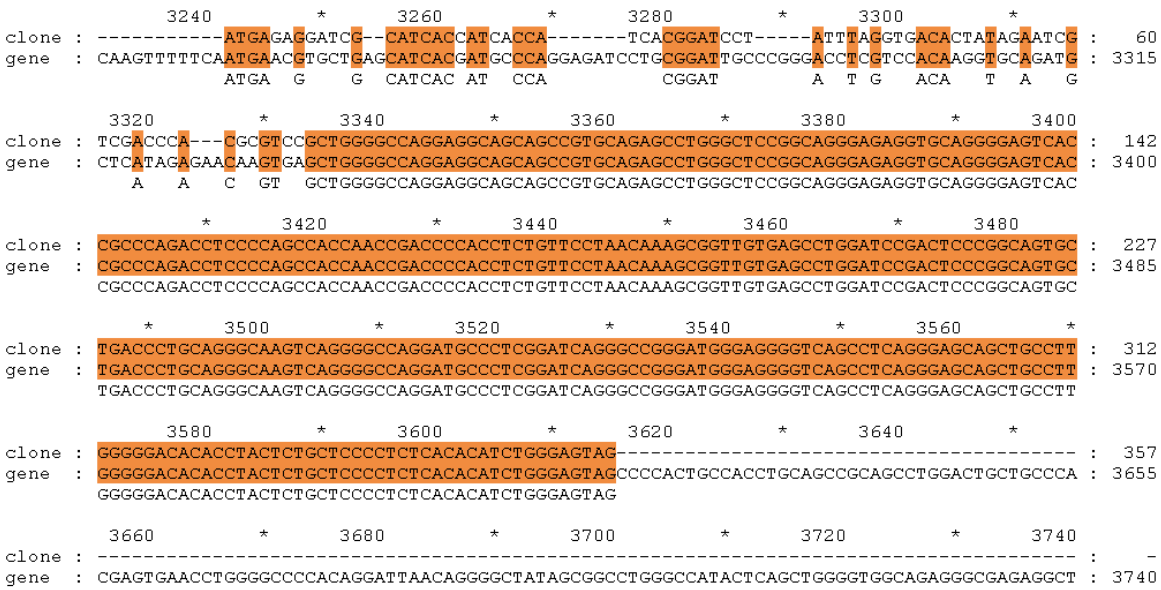
Appendix D

Clone	ENST	Description	Mass spec.	Note
A11548	ENST00000308162	COFILIN	5 peptides	
A15579	ENST00000262868	DISKS LARGE-ASSOCIATED PROTEIN 4	9 peptides	
B08508	ENST00000322981	HEPATOCELLULAR CARCINOMA-ASSOCIATED ANTIGEN		
B13511	ENST00000224864	UNKNOWN	13 peptides	
C01529	ENST00000324441	HEPATOCELLULAR CARCINOMA ASSOCIATED ANTIGEN		
C24599	ENST00000323233	SMOOTH MUSCLE CELL LIM PROTEIN	12 peptides	
D22574	ENST00000311172	AMBIGUOUS		3´UTR
E03569	ENST00000326674	HIGH MOBILITY GROUP PROTEIN 1	7 peptides	
E04569	ENST00000311502	AMBIGUOUS	8 peptides	
E05577	ENST00000160740	UNKNOWN	6 peptides	
E08544	ENST00000215956	NHP2-LIKE PROTEIN 1	3 peptides	
E09559	ENST00000312854	AMBIGUOUS		+ 5´UTR
E10569	ENST00000310837	PROTEIN TRANSLATION FACTOR SUI1 HOMOLOG	4 peptides	+ 5´UTR
E12566	ENST00000216271	HISTONE DEACETYLASE 10		+ 5´UTR
E23538	ENST00000252151	VOLTAGE-DEPENDENT R-TYPE CALCIUM CHANNEL	5 peptides	
E23549	ENST00000299402	AMYLOID BETA A4 PRECURSOR PROTEIN-BINDING	3 peptides	
F05601	ENST00000216259	PHOSPHOMANNOMUTASE 1	15 peptides	
F07530	ENST00000323233	SMOOTH MUSCLE CELL LIM PROTEIN	11 peptides	
F10569	ENST00000322852	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN	14 peptides	
F10589	ENST00000262964	HYPOTHETICAL	4 peptides	3´UTR
G05529	ENST00000275388	UNKNOWN		
G07549	ENST00000335534	HYPOTHETICAL KINESIN MOTOR DOMAIN CONTAINING		
G11565	ENST00000217442	MICROTUBULE-ASSOCIATED PROTEINS 1A	9 peptides	
H01511	ENST00000265388	TRANSPORTIN-SR		
H09541	ENST00000207437	MYOSIN LIGHT CHAIN 1	11 peptides	
H21517	ENST00000264958	DIHYDROPYRIMIDINASE RELATED PROTEIN-1		+ 5´UTR
I01538	ENST00000269209	UNKNOWN	14 peptides	
I07563	ENST00000323233	SMOOTH MUSCLE CELL LIM PROTEIN	8 peptides	
I16511	ENST00000258737	RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1	6 peptides	+ 5´UTR
I17574	ENST00000261232	SLIT-ROBO RHO GTPASE ACTIVATING PROTEIN		
K01518	ENST00000276689	NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT	6 peptides	+ 5´UTR
K21548	ENST00000268989	AMBIGUOUS		3´UTR
L05517	ENST00000313370	ZINC FINGER PROTEIN 106	9 peptides	
L07519	ENST00000262868	DISKS LARGE-ASSOCIATED PROTEIN 4	8 peptides	
M03544	ENST00000301071	ALPHA-TUBULIN 1	2 peptides	+ 5´UTR
M08528	ENST00000265526	ZINC FINGER PROTEIN GLI3	9 peptides	
M10533	ENST00000262395	TNF RECEPTOR-ASSOCIATED FACTOR 4 ISOFORM 1	7 peptides	
M14512	ENST00000326766	BULLOUS PEMPHIGOID ANTIGEN 1		
N06528	ENST00000300131	NGFI-A BINDING PROTEIN 2	7 peptides	
O01523	ENST00000307296	PROGRAMMED CELL DEATH 6 INTERACTING PROTEIN		
O06583	ENST00000297142	NEUROGENIC DIFFERENTIATION FACTOR 6		
O23567	ENST00000311601	UNKNOWN	6 peptides	
P07593	ENST00000290454	UDP-GLUCOSE 4-EPIMERASE		+ 5´UTR

Appendix E


C01529 : MRGSHHHHHHGSYLGDTIESSTHAS----------AHASGRPVREMRGASARPPLLGSEEPFCPASRSVREGGGGVSPPPGQPPCPE : 77
HCA25A : ----------MRGASARPPLLGSEEPFCPASRSVREGGGGVSPPPGQPPCPE : 42
B08508 : MRGSHHHHHHGSYLGDTIESSTHASPSGREVGGSVPRPASRPVREMRGASARPPLLGSEEPFCPASRSVREGGGGVSPPPGQPPCPE : 87
mrgshhhhhhgsylgdtiessthas rpvreMRGASARPPLLGSEEPFCPASRSVREGGGGVSPPPGQPPCPE
C01529 : GRWGGQPPARPADPSGREVGGSAPCPASRPVREVRGASARPPLLGSEEPLCPASRPIWEGGGGVSPPPCQPPCPGGRGASARPPLLG : 164
HCA25A : GRWGGQPPARPADPSGREVGGSAPCPASRPVREVRGASARPPLLGSEEPLCPASRPIWEGGGGVSPPPCQPPCPGGRGASARPPLLG : 129
B08508 : GRWGGQPPARPADPSGREVGGSAPCPASRPVREVRGASARPPLLGSEEPLCPASRPIWEGGGGVSPPPCQPPCPGGRGASARPPLLG : 174
GRWGGQPPARPADPSGREVGGSAPCPASRPVREVRGASARPPLLGSEEPLCPASRPIWEGGGGVSPPPCQPPCPGGRGASARPPLLG
C01529 : NEEPLCPATTPSGRCTQQLVENGP : 188
HCA25A : NEEPLCPATTPSGRCTQQLVENGP : 153
B08508 : NEEPLCPATTPSGRCTQQLVENGP : 198
NEEPLCPATTPSGRCTQQLVENGP

Appendix F



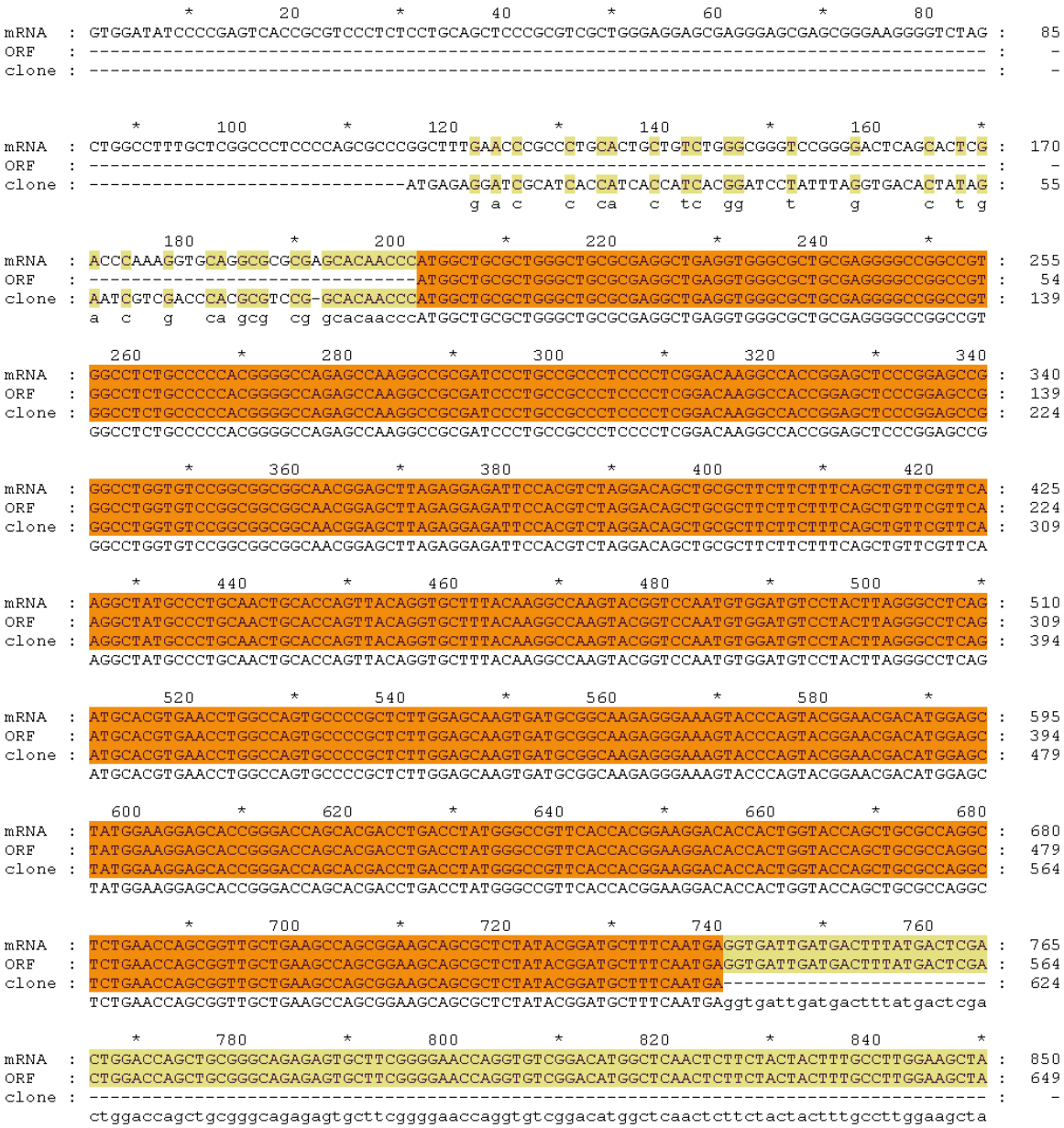
Artificial polypeptide sequence of clone K21548

Matched mRNA: BC039204
Calculated molecular weight: 12.082,08 Dalton

MRGSHHHHHHGSYLGDTIESSTHASAGARRQQPCRAWAPAGRGAGESPPRPPQPPTDPTS

VPNKAVVSLDPTPGSADPAGQVRGQDALGSGPGWEGSASGSSCLGGHTYSAPLSHIWE

Appendix G



Artificial polypeptide sequence of clone H24525

Matched mRNA: NM_000784
Calculated molecular weight: 21.391,75 Dalton

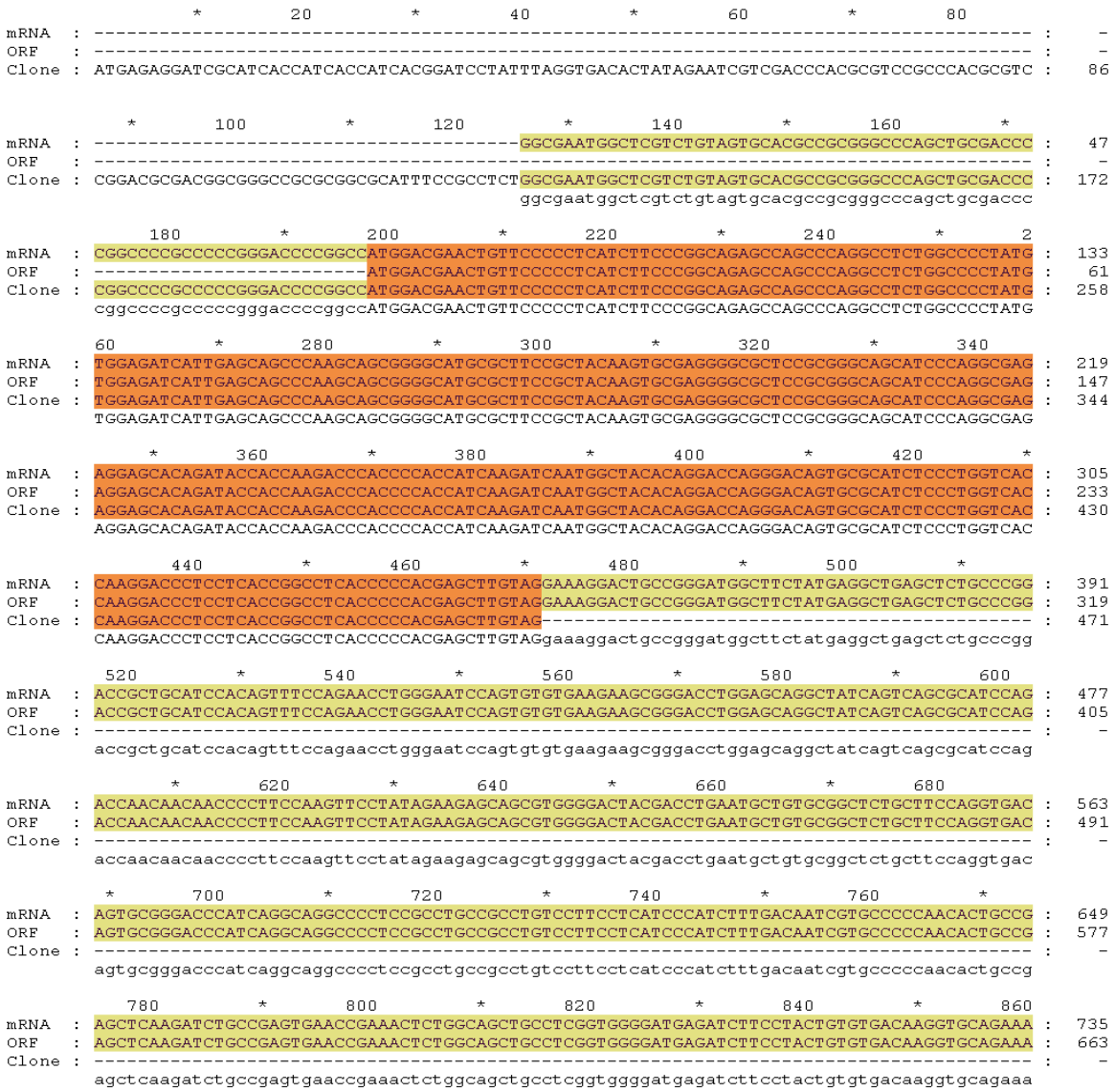
MRGSHHHHHHGSYLGDTIESSTHASGTTHGCAGLREAEVGAARGRPWPLPPRGQSQGRDP

CRPPLGQGHRSSRSRAWCPAAATELRGDSTSRTAALLLSAVRSRLCPATAPVTGALQGQV

RSNVDVLLRASDAREPGQCPALGASDAARGQVPSTERHGAMEGAPGPARPDLWAVHHGRT

PLVPAAPGSEPAVAEASGSSALYGCFQ

Appendix H



Artificial polypeptide sequence of clone M06599

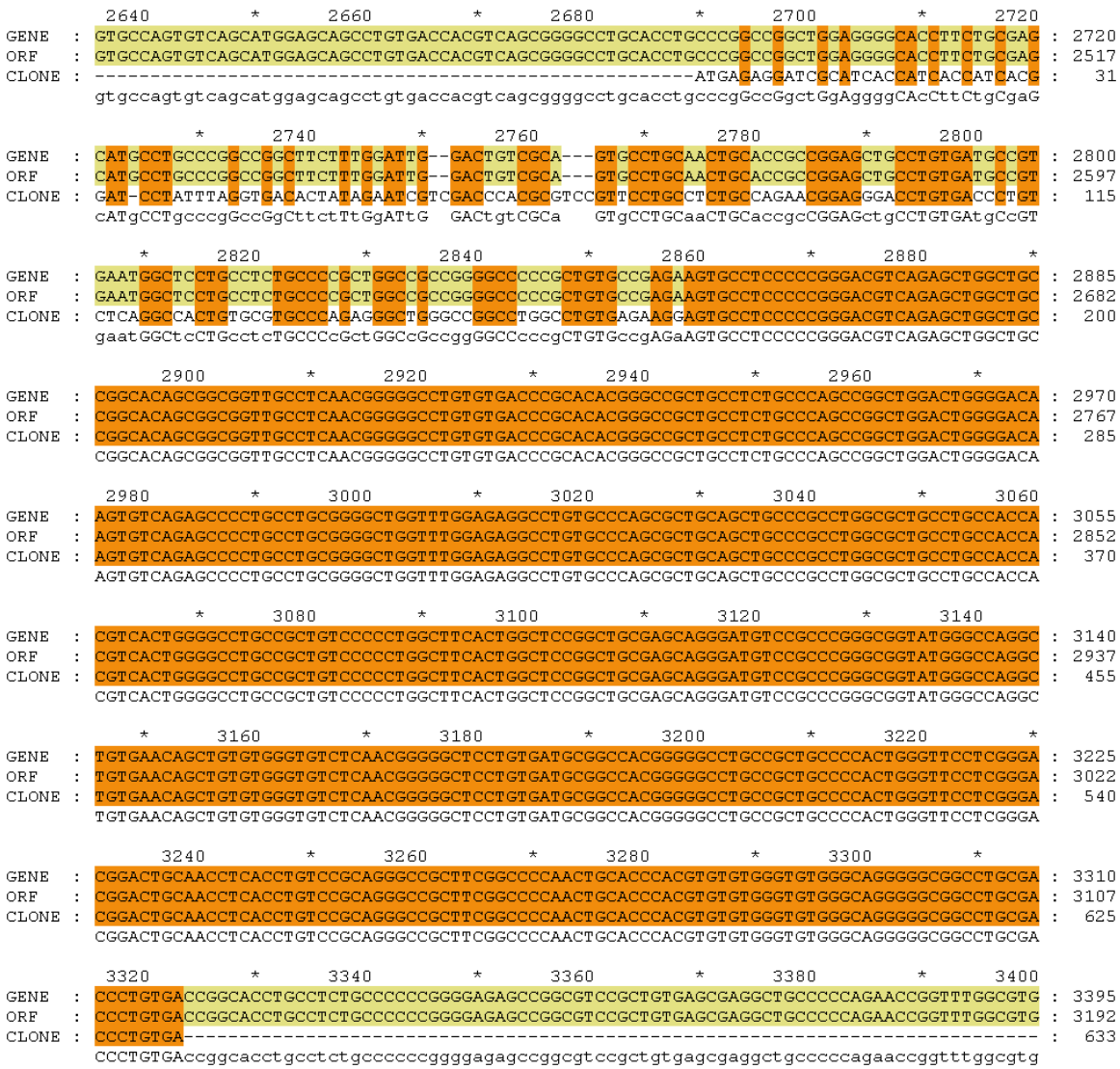
Matched mRNA: NM_021975
Calculated molecular weight: 15.910,57 Dalton

MRGSHHHHHHGSYLGDTIESSTHASAHASGRDGGPRGAFPPLANGSSVVHAAGPAATPAPP

PGPRPWTNCSPSSSRQSQPRPLAPMWRSLSSPSSGACASATSARGAPRAASQARGAQIPPR

PTPPSRSMATQDQGQCASPWSPRTLLTGLTPTSL

Appendix I



Artificial polypeptide sequence of clone M10510

Matched mRNA: XM_031401
Calculated molecular weight: 20.415,02 Dalton

MRGSHHHHHHGSYLGDTIESSTHASVPASARTEGPVTLSQATVRAQRAGPAWPVRRSASP

GTSELAAGTAAVASTGACVTRTRAAASAQPAGLGTSVRAPACGAGLERPVPSAAAARLAL

PATTSLGPAAVPLASLAPAASRDVRPGGMGQAVNSCVGVSTGAPVMRPRGPAAAPLGSSG

RTATSPVRRAASAPTAPTCVGVGRGRPATL