C h a p t e r 1

INTRODUCTION

The Human Genome Project identified around 20 000 to 30 000 protein-coding genes in the human genome – only one third of what was previously thought (2004; Lander et al., 2001; Venter et al., 2001). This means that the large increase in protein diversity is an output of alternative splicing and post-translational modifications of relatively few genes. Thus gene expression analysis alone is not sufficient for the characterisation of protein diversity in an organism. Functional genomics aims to describe genome functions and thus refers mainly to the proteomic level. This is a very broad field because of the large number of interactions, which proteins can have with different kinds of molecules - from two interaction partners to large complexes. Altogether, functional genomics is huge compared with genomic studies because even in one organism the proteome differs from cell to cell and often changes in even one cell during development, depending on biochemical interactions with genome and environment.

1.1 Microarray technology

Traditional methods in molecular biology based on the principle "one gene in one experiment" are time consuming and cost intensive regarding the amount of knowledge, which is generated. High-throughput technologies like microarrays are indispensable to conduct on genome-wide level. Microarrays are miniaturised sample carriers on special surfaces used to arrange and bind a large number of biological materials. Choudhuri defines them as a *"high-throughput assay system which utilizes spatially ordered discrete, high-density arrangement of biologically important entities immobilized on a solid platform"* (Choudhuri, 2004). "Entities" can be nucleic acid fragments, proteins, whole cells, or tissues. Using microarrays, ten thousands of genes or the whole genome can be analysed simultaneously in a single experiment. This ensures well-founded statistical comparison of the samples and the high yield of information from a particular experiment.

ELISA (enzyme-linked immunosorbent assay) (Engvall et al., 1971; Engvall and Perlmann, 1972) and the dot blotting technique for nucleic acid sequence detection (Kafatos et al., 1979) are technological forerunners of microarrays . In the late 80`s of the last century Ekins and colleagues (Ekins and Chu, 1992; Ekins et al., 1989; Ekins, 1989) designed their "microspots" and produced first microarrays with robots. Only a few years after their invention, microarrays were already well established in genetics and being used in fingerprinting and genome analysis (Hoheisel et al., 1994; Lennon and Lehrach, 1991) and the expression of sequence catalogues (Meier-Ewert et al., 1993). Researchers from Affymax Research Institute in California invented the name "DNA chips" (Fodor et al., 1991; Pease et al., 1994) and commercialised first microarrays as GeneChip® in 1996.

The most common DNA microarrays are mainly used for genome-wide quantification of gene expression or to identify genetic variations through detection of single nucleotide polymorphisms (SNPs) across populations. Also prevalent are protein microarrays, which are used to screen for the ability of the spotted proteins to bind molecules (e.g. receptors, antibodies, enzymes, hormones, or peptides). Thus this method can also be used to complement other protein-protein-interaction methods (see section 1.4). Currently, protein microarrays are often used in diagnosis of diseases by identifying a set of associated proteins.

1.2 Reverse transfection

One of the numerous applications of DNA microarrays is the method of reverse transfection first published by Ziauddin and Sabatini (Ziauddin and Sabatini, 2001). In contrast to the traditional chemical transfection (called "direct transfection" in this paper), where the DNA of interest is in solution and given together with transfection chemicals on top of cells, in reverse transfection a solution containing gelatine and the DNA of interest is spotted and dried on a glass surface. Transfection reagent is already in the samples or alternatively an additional incubation step is interposed before transfection. This microarray of spotted constructs is then covered with a layer of adherent cells, resulting in the transfection of only cells growing on top of the DNA spots and thus expression of specific proteins in spatially distinctive groups of cells. The phenotypic effect of the transfected arrays can be detected by cell-based bioassays like immunofluorescence or using autofluorescent reporter

proteins. The outstanding advantage of reverse transfection over direct transfection is its suitability for high-throughput research. One single slide can contain a set of hundreds of different samples, which are all transfected at the same time and under the same conditions. Thus reverse transfection experiments occur under a more uniform environment than single direct transfection experiments.

Reverse transfection is an appropriate method for many different domains, such as RNA interference (RNAi) research (Erfle et al., 2004; Kumar et al., 2003; Mousses et al., 2003; Vanhecke and Janitz, 2004, 2005; Wheeler et al., 2004; Wheeler et al., 2005) or for cell array-based intracellular localization screenings (Hu et al., 2005; Hu et al., 2006). It can be used to determine members of signalling pathways (Webb et al., 2003), to identify novel therapeutic targets (Mishina et al., 2004) and is also suitable for characterising proteins and their functions (Hodges et al., 2005).

1.3 Protein-protein-interaction research

Most if not all biological processes require cooperation of pairs of proteins or the formation of large functional complexes of proteins. According to estimations of Gavin and Superti-Furga there are hundreds of discrete protein complexes in eukaryotic cells, many of them containing dozens or hundreds of different proteins (Gavin and Superti-Furga, 2003). Presumably human proteins are linked with each other in 150 000 to 200 000 or more interactions (Bork et al., 2004; Figeys, 2003; Peri et al., 2003). Nearly 30 000 of them are already catalogued in the Human Protein References Database (HPRD). It is now clear that almost all proteins in a cell are part of a large protein interactome, the *"complete repertoire of interactions potentially encoded by (…) genome"* of an organism (Sanchez et al., 1999). Thus analysing protein-protein-interactions (PPI) is essential for the elucidation of biological progress, and the determination of the human interactome structure is the next big challenge after the human genome project. *"If we could map the interactions of proteins we would be able to understand protein functions"* (Figeys, 2003).

Protein-protein-interactions can be regulated in several ways. The most important is the control of the protein expression at the genomic and transcriptional level. Post-translational modifications such as phosphorylation or acetylation are also relevant, as well as the location of the protein within the cell. Some transcription factors like NF-κB are activated by decomposition of an inhibitor as a result of phosphorylation with subsequent translocation from the cytoplasm into the nucleus (Ghosh and Karin, 2002). Also important for PPI regulation is the stability of the proteins, the presence of appropriate receptors on the cell membrane, and potential ligands (see section 1.6).

1.4 Methods to detect protein-protein-interactions

It is very difficult to predict interaction partners for particular proteins, even with the knowledge of specific domain properties, like for example in the case of SH3 domain which preferentially binds to sequences containing amino acid proline (Pawson and Nash, 2003). Thus experimental approach is essential to analyse PPI, either *in vitro* or *in vivo*. The most common *in vitro* method is mass spectrometry (MS). Using this technique, not only protein pairs can be analysed but also big complexes. Interactions must be entire to be detected, MS as such is expensive and time consuming. The *in vivo* two-hybrid principle is more suitable for high-throughput PPI research and also allows mapping of interactions within a protein complex, which is very difficult when using MS. A good review of protein-protein research has been published for example by Zhu and colleagues (Zhu et al., 2003).

1.4.1 Yeast two-hybrid system

In 1989 Fields and Song demonstrated *in vivo* detection of PPI in yeast using a method they termed the yeast two-hybrid system (Fields and Song, 1989). They took advantage of the modular nature of transcriptional activators consisting of largely independent DNA-binding and activation domains (Brent and Ptashne, 1985; Keegan et al., 1986; Ptashne, 1986, 1988; Sadowski et al., 1988). Using this and the knowledge of generating hybrid activators (Brent and Ptashne, 1985), Fields and Song separated the two functional domains of the GAL4 protein of the yeast *Saccaromyces cerevisiae* and thus generated a two-part-system. In the first step a gene of interest is cloned into the "bait" vector, so that the gene is placed next to a DNA-binding domain (DBD). The bait (X) has no ability to activate the reporter gene. A second gene (or a library

of cDNAs) encoding a potential interaction partner is cloned downstream of the activation domain (AD) of the GAL4 yeast transcription factor in the "prey" vector. The prey (Y) has no ability to bind DBD-responsive elements. If the two proteins fused to the DBD and AD interact physically, they will bring the domains close together and restore a functional transcription factor that binds to the promoter of a reporter gene. The transcription of this reporter gene is activated and the related reporter protein or its catalytic activity can be detected.

The two-hybrid system, which has frequently been reviewed (Chien et al., 1991; Ito et al., 2001b; Uetz, 2002), can also be used to acquire detailed information about specific interaction domains of proteins or to determine specific amino acid residues through point mutations. The two-hybrid system is applied not only for testing interactions between known proteins but also for screening libraries for determination of new interaction partners. The first array-based two-hybrid screen of a whole proteome (*S. cerevisiae*) was published in 2000 (Ito et al., 2001a; Ito et al., 2000; Uetz et al., 2000). Recently the group of Wankers identified more than 3000 potential human PPI by using this method (Stelzl et al., 2005).

1.4.2 Limits of the yeast two-hybrid system

The yeast two-hybrid system represents one of the most efficient approaches currently available for identifying and characterising protein-protein-interactions. It is highly sensitive and detects interactions not detected by other methods (Li and Fields, 1993). However it has several drawbacks. First, the fusion proteins have to be translocated to the nucleus and must be able to fold and exist stably in yeast cells. In some cases the fusion to a transcription factor domain may occlude the site of interaction. Also interactions that need secondary modifications of the proteins, such as phosphorylation, or third interaction partners cannot be detected, which means that some true interactions stay unrecognised. These false negatives lead to up to 90% (Ito et al., 2001a) or up to 96% (Edwards et al., 2002) undetected interactions when using the yeast two-hybrid system.

On the other hand, some proteins will give false-positive signals by activating transcription without an interacting partner when fused to a DNA-binding domain (Ma and Ptashne, 1987). This may be the case for about 50% of interactions obtained from yeast two-hybrid (von Mering et al., 2002). Even with progress in this field for example from the group of Suzuki (Saito et al., 2002) it stays problematic.

Several variations of the two-hybrid system have been developed to overcome these limitations. For example in the yeast three-hybrid system (Licitra and Liu, 1996; SenGupta et al., 1996) a third partner (Z) is expressed, which is involved in interaction of bait and prey. This protein can be necessary for bridging bait and prey and can thus enable the interaction, or alternatively may prevent the interaction between them (Tirode et al., 1997; Zhang and Lautar, 1996). Other variations of the yeast two-hybrid system are the one-hybrid system (Wang and Reed, 1993), the "reverse" two-hybrid (Vidal et al., 1996) and the split-hybrid (Shih et al., 1996).

To analyse membrane proteins, the yeast two-hybrid is not suitable because of its limitation to protein partners whose interaction is assessed in the nucleus. For this demand other systems like the SRS (SOS Recruitment System) and the RRS (Ras Recruitment System) are more suitable (Aronheim, 1997; Aronheim et al., 1997). Also FRET (fluorescence resonance energy transfer) or its modification BRET (bioluminescence resonance energy transfer) can be used for this purpose (Pollok and Heim, 1999; Truong and Ikura, 2001; Xu et al., 1999). Last but not least PPI can also be detected with systems like USPS (split-ubiquitin system) (Johnsson and Varshavsky, 1994a, 1994b; Stagljar et al., 1998), MAPPIT (Mammalian Protein-Protein Interaction Trap) (Eyckerman et al., 2001), or reverse MAPPIT (Eyckerman et al., 2005).

1.4.3 Mammalian two-hybrid system

Currently, mammalian protein-protein-interactions are mainly investigated using the yeast two-hybrid system. However, in many cases post-translational procedures are essential for correct protein processing, so studying mammalian genes in yeast is problematic. These modifications can differ between organisms and even between cell types. This entails a high rate of false negatives when using a different organism from the one the genes are derived from. For example, TGF-β-induced interaction between Smad3 and c-Jun proteins is detectable only in mammalian cells and not in the yeast two-hybrid system (Feng and Derynck, 2001). Sometimes proteins can bind to an endogenous yeast protein (Luo et al., 1997), so a signal can be detected only in yeast but not in mammalian cells. This means that every potential interaction found in yeast has to be verified. Thus mammalian genes should be studied in mammalian cells, their natural environment.

Dang et al. were the first investigators to use mammalian cells instead of yeast (Dang et al., 1991). The principle of this two-hybrid assay is similar: A gene coding for a protein of interest and another coding for a potential partner are cloned to DNA-binding domain and activation domain from a transcription factor, respectively. After transfection in mammalian cells, interaction of the chimeric proteins brings the domains together and restores expression of the reporter gene.

One of the advantages of the mammalian two-hybrid system is that proteins maintain their native conformation, and additional factors necessary for the interaction of both proteins are available. This is especially relevant for proteins which interacty indirect in multi-protein complexes, for example transcription factors (Feng and Derynck, 2001). Thus, the mammalian two-hybrid-system is often used to further evaluate protein-interaction-partners found in yeast (Leonhardt et al., 1998; Luo et al., 1997) or for small-scale studies (Dixon et al., 1997). But currently mammalian two-hybrid systems involve high reagent consumption and are therefore not practical for screening PPI in mammalian cells. This is true also for variations published by various groups (Fearon et al., 1992; Fotin-Mleczek et al., 2000; Shioda et al., 2000; Vasavada et al., 1991).

The first attempt to use the mammalian two-hybrid system on a larger scale comprised transfection in 96-well or 384-well plate format (Murakami et al., 2002; Suzuki et al., 2001; Zhao et al., 2004). Automated transfection and immunostaining of mammalian cells have been established by Liebel et al. (Liebel et al., 2003). In either case, usage of microwell plate format requires automation of liquid dispensing and is characterised by high consumption of reagents.

1.5 Biology of androgen receptor

Androgens are steroid hormones in vertebrates which are essential for the development and maintenance of primary and secondary masculine characteristics (Gao et al., 2005; Lee and Chang, 2003). The most well known androgen is testosterone and its metabolite 5α-dihydrosterone (DHT). Because of its higher binding affinity (Wilbert et al., 1983) and slower rate of dissociation from the androgen receptor (Zhou et al., 1995) DHT is the more potent androgen in most target tissues.

Most androgens and androgenic components like methyltrienolone (R1881) work through receptor-mediated mechanisms (Fang et al., 2003), only a few target another site than the ligand binding domain (Yamabe et al., 2000). Anti-androgens can lead to incomplete masculinization or reduced fertility (Kelce and Wilson, 1997). They are "pure" with no other endocrine effect (e.g. flutamide) or can have gestagenic effect and in doing so act as anti-androgen (e.g. cyproterone acetate). Recently they have frequently been used to treat prostate cancer (Sharifi et al., 2005), but the precise mechanism of anti-androgenic action is still unclear (Gao et al., 2005).

The androgen receptor (AR) is a member of the nuclear receptor superfamily of ligand-dependent transcription factors (Beato et al., 1995; Evans, 1988; Mangelsdorf et al., 1995; Tsai and O'Malley, 1994), which control essential physiologic and developmental processes in humans and play an important role in prostate cancer (Heinlein and Chang, 2004). It has the common domain structure of nuclear receptors: A N-terminal activation domain (NTD, activation function 1 = AF1), a central DNA-binding domain (DBD), a hinge region and a C-terminal ligand-binding domain (LBD), usually with a second activation function (AF2) (Belandia et al., 2005; Jenster et al., 1991; Simental et al., 1991). Basis for the effects of AR is the binding of a suitable ligand. In unli-

gated status the AR is mainly localised in the cytoplasm (Jenster et al., 1993; Simental et al., 1991), associated with heat-shock proteins (HSPs), which facilitate ligand binding (Fang et al., 1996; Pratt and Toft, 1997). When a ligand binds to the AR, the confirmation changes, and the activated receptor complex enters the nucleus (Georget et al., 1997). A detailed review of this and an alternative mode of action is given by Gao et al. (Gao et al., 2005).

In 1995 Langley et al. reported interaction of NTD and LBD of AR which they inhibited with the anti-androgen hydroxyflutamide (OH-Flu) (Langley et al., 1995). Two years later Doesburg et al. described interaction of the N-terminal transactivation domain (which they named TAD) and the carboxyl-terminal domain of AR as well as its hormone dependence and the blocking by the use of different anti-androgens (Doesburg et al., 1997). They also found a weak LBD-LBD interaction and postulated intramolecular interaction between TAD and LBD of AR, in contrast to the intermolecular interaction proposal of Langley (Langley et al., 1998). However, the interaction of both domains stabilizes bound androgen and slowly its dissociation rate (He et al., 1999; He et al., 2000). Kemppainen et al. classified further AR ligands to be agonists or antagonists of this interaction and found among others OH-Flu and medroxyprogesterone acetate (MPA) (Kemppainen et al., 1999). MPA is a weak androgen *in vivo* (Bardin et al., 1983; Mowszowicz et al., 1974; Raynaud et al., 1980), but failed to induce the N/C-interaction. Its activation mechanism seams to be different from other agonists.

The nuclear receptors alter transcription through interaction with coregulators (Ikonen et al., 1997; McKenna et al., 1999). They bind to receptor complexes in the nucleus and enhance transactivation (coactivators) or reduce it (corepressors). Good overviews of coregulator functions and different modes of action are given elsewhere (Heinlein and Chang, 2002; Kumar et al., 2004; Lee and Chang, 2003; McKenna et al., 1999; Privalsky, 2004). Recently it has been found that beside pure agonists there are also partial ones, which induce a flexible state where the binding of coactivators as well as corepressors is enhanced over the unliganded state (Albers et al., 2006). In 2005 the group of Haendler characterised AR45, a variant of AR, which can either repress or stimulate the AR activity depending on levels of AR and AR45 and of cofactors such as β-catenin (Ahrens-Fath et al., 2005).

1.6 Aims of the study

The rapidly growing collection of gene sequences as a result of genome sequencing projects demands the development of systematic and high-throughput approaches for investigating protein-protein-interactions in mammalian cells. Existing techniques can deal with a limited number of genes and require automated liquid dispensing. DNA microarrays could be used to analyse genes on a genome-wide scale for relatively small costs and reagent consumption. The recently developed transfected cell arrays combine microarray technology with protein expression in mammalian cells, paving the way towards the development of new technologies for functional genomics.

The aim of this study comprised following topics:

1. Development of the high-throughput technique for screening of protein-protein-interactions in mammalian cells through the adaptation of the transfected cell array technology.
2. Establishment of proof-of-principle for this mammalian two-hybrid platform using as an example the set of genes involved in androgen receptor signalling.
3. Extension of the applicability of the established technique towards screening of a prey-library with a bait of interest in different cell lines towards investigation of cell type-specific patterns of protein-protein-interactions.