Application of fragment-based drug discovery to membrane proteins

Früh, V.

Citation

Früh, V. (2009, October 7). *Application of fragment-based drug discovery to membrane proteins*. Retrieved from https://hdl.handle.net/1887/14047

Version: Not Applicable (or Unknown)

License: Leiden University Non-exclusive license

Downloaded from: https://hdl.handle.net/1887/14047

Note: To cite this publication please use the final published version (if applicable).

Chapter

1

2001
Human genome
sequenced5

>2000

Modern
Drug
discovery

Rational
drug design
protein = target

High
Throuput
Automated
processes;
Biotechnology
advances

Introduction to the application of fragment-based drug discovery to membrane proteins

An overview of drug discovery

The need to relieve pain and suffering by medication has been with us since the dawn of humanity. Spiritual healing, accompanied by herbal medicines was passed on from ancestors who treated the symptoms of the soul. From then, a myriad of information and technological advances have contributed to each step leading to modern day drug discovery (Figure 1), where treatments are developed to treat the body.

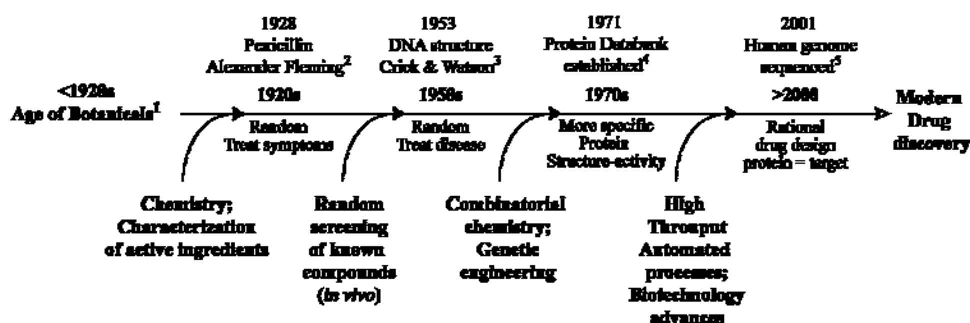

Figure 1. Historical overview of the drug discovery process.

The age of botanicals[2] was defined by the application of ancestral herbal medicines after trial and error determined which treatment was best for a particular ailment. During those days, each plant was believed to be designed to heal a particular symptom. It was not until the 19th century that active ingredients were isolated and characterized with the developing science of chemistry, such as the South American poison arrow curare by Claude Bernard[3]. From then, drug discovery was still very random with accidental discoveries such as Penicillin by Alexander Fleming in 1926[4], among many others. Developments in the 1950s led the way to serendipity in drug discovery, owing its name to the random screening of a variety of known compounds on animal models, with successful emergence of many drugs[5]. With breakthroughs such as the elucidation of the DNA structure by Watson and Crick[6], we started having a better idea of human biology. The low diversity of available molecules and the common usage of animal models with limited knowledge of drug mechanisms however, were still limiting the discovery of drugs with high specificity and low toxicity for the human protein targets. Finally, with emerging biotechnology and high throughput automated processes, the last 50 years have seen a boom in drug discovery strategies owing to our increasing knowledge related to handling and analysing biological

material. The emergence of genetic engineering and the protein data bank[7] provided us with the possibility of zooming into tissues, isolating and studying single protein targets to better understand the relationship between protein and drug structures. Combinatorial chemistry would enable an extremely wide variety of chemicals to be screened on a protein target rather than in an in vivo animal model, with potential specific changes to be made on different parts of the molecule to enhance qualities or decrease toxicity[5]. The sequencing of the human genome by the International Human Genome Sequencing Consortium has estimated that our genome contains 20000 to 25000 genes[8]. 10 – 15 % of these have been estimated to code for drug targets[9], exposing our bodies to modern day rational drug design.

Naturally, increasing knowledge of animal and human biology, developing technologies, and market considerations have altered the way the drug discovery process occurs. Large corporations of pharmaceutical companies for example are incorporating rational drug design strategies, in order to maximise the chances of success. The basic steps of modern rational drug discovery processes, which start by identifying a target and pave the way to identifying drug leads for clinical trials, are highlighted in Figure 2.

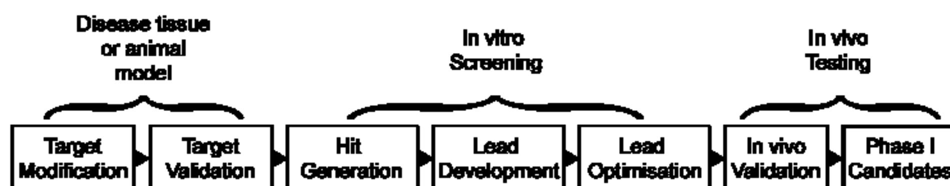

Figure 2. Basic steps in modern drug discovery.

There are increasingly perceptive methods for each of the steps involved, ranging from virtual to experimental ones. The central theme revolves around obtaining valuable information related to structure-activity relationships between the target and hit compounds in order to develop lead compounds which have an advanced trade off between better affinities and specificities, and better drug-like ADME (Absorption, Distribution, Metabolism, and Excretion) properties, as predicted by Lipinski's rule of "5"[10] (Figure 3**a**).

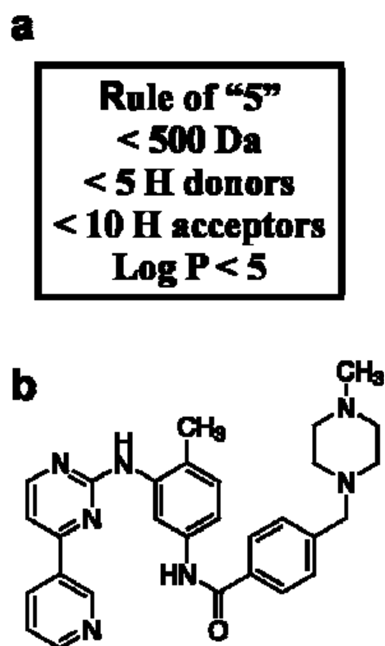

Figure 3. Predicting what makes a compound a good bioavailable drug. Lipinski's rule of "5"[10] defines the prerequisites for a good orally available drug (**a**) such as imatinib mesylate (Gleevec) (**b**), used as a drug to treat leukemia[11].

Before sensitive molecular methods had evolved, the effect of a compound on a target could only be measured by detecting an effect on target function. In order to detect such effects the compounds had to interact strongly with the target, often leading to the necessity of screening fairly large compounds in functionality based assays (Figure 3**b**). These large compounds, however, were often difficult to chemically elaborate without violating the rule of 5 due to increased hydrophobicity upon chemical elaboration. This type of screening, now coined traditional High Throughput Screening (HTS), was widespread in the last decades but was not always successful for some targets[12]. With emerging molecular methods such as Nuclear magnetic resonance (NMR), crystallography, and surface-plasmon resonance (SPR), fragment-based drug discovery (FBDD) was developed to detect hits for those targets for which HTS and other methods failed, and soon, the advantages of this strategy became more apparent.

Smaller is better: Why use a fragment approach to drug discovery?

As can be seen in Figure 4**a** with the example of a complex binding site, the increasing size of a molecule increases the probabilities of detecting it, but the probability of finding a unique match between ligand and target has an optimum at a very low complexity level. There is therefore a trade-off between the probability of detecting binding and the probability of finding a good match between ligand and target. When screening is carried out with smaller molecules, there is a better chance of finding a unique binding mode, but the binding affinity is very low

compared to larger molecules. Molecular methods are therefore required to detect the weak interactions between a target and a fragment.

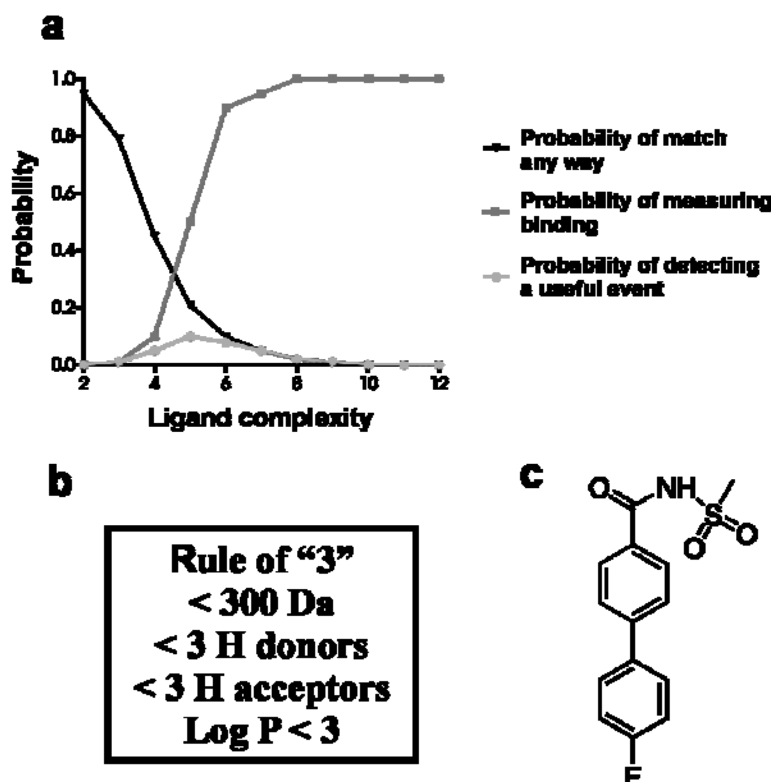

Figure 4. Using fragments in drug discovery. The success landscape for a binding site of high complexity, adapted with permission from reference 13. Copyright 2001 American Chemical Society (**a**). The rule of “3” (**b**) can be applied to predict the bioavailability of a fragment such as the one used as inhibitor of BclxL, a membrane protein involved in the survival of cancer cells[14] (**c**).

Using biophysical methods which detect interactions at the atomic level can now enable screening of fragments which were otherwise undetected by HTS methods. The high probability of finding a good match between a target binding site and a fragment is not the only advantage of FBDD. By definition, fragments are molecules which are smaller, simpler, and more soluble than drug-like compounds. This brings the Lipinski’s rule of 5 down to a rule of 3[15] (Figure 4**b**) where there remains a larger margin to elaborate or link the fragment into a more potent drug which remains bioavailable. An example of a fragment is shown in Figure 4**c**.

Small – With molecular weights ranging from 160 - 300 Da, these fragments are usually easier to synthesize than the larger HTS compounds. The maximum amount of fragments that can possibly be synthesized are estimated at 10^7, as opposed to 10^{60} larger HTS compounds[16]. Furthermore, as can be seen in

Figure 5, these smaller structures are much more efficient at probing key binding areas in a target where larger compounds may sterically hinder access to these areas[17]. Clearly, the advantage of using fragment libraries lies in the fact that fragments can be smaller and yet more efficient at probing the available structure pool! For comparison, a million compounds are often screened for HTS strategies, compared to 1000 fragments in FBDD[18].

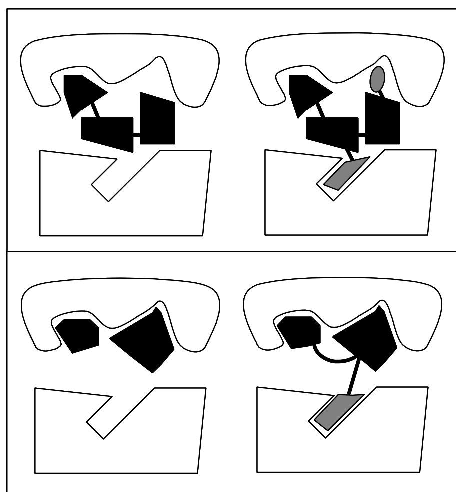

Figure 5: Capacity of an HTS compound (**top, black**) or fragments (**bottom, black**) to probe the key binding areas of a protein binding site (**white**). Fragments are better starting points and can be elaborated into larger compounds (**bottom, grey**) with better specificity and bioavailability than HTS compound (**top, grey**) without easily violating Lipinski's rule of 5.

Simple **–** Fragments are simple structures which tend to be void of reactive groups that can lead to toxicity[19] or metabolic instability[12]. Screening fragments is amenable to finding lead products with better starting bioavailability profiles[20], thereby reducing the often laborious chemistry efforts required to remove initial unwanted functional groups.

Soluble **–** As previously mentioned, previous research has established that elaborating compounds by modifying or adding chemical groups often leads to more hydrophobic compounds which easily violate Lipinski's rule of 5. FBDD is a good alternative approach to HTS because the more soluble fragments are better starting points for optimisation as a variety of chemical elaborations can be carried out with less failure in developing drug-like molecules with good ADME properties[13].

As explained, fragment screening relies on the detection of fragments which bind weakly to target proteins. In most cases, these fragments need to be elaborated into larger compounds with higher affinity. Lead optimization can be carried out in a variety of ways, but often depends on having a high resolution target structure available. This

enables one to identify which chemical groups on the fragment to modify in order to have more potent and specific affinities with the protein target binding site. Computational chemistry and informatics therefore play a key role. The initial idea was to link fragments together, whereby linking small weak fragments would lead to higher potency compounds. Although this was successful in many cases[20,21,22], linking chemistries are not always successful and can be replaced by elaboration of single or overlapping fragments by addition/removal of groups until the desired outcome is produced[14], or by screening focused libraries around a central theme scaffold of previously identified hits[12].

Requirements for fragment-based drug discovery

Interactions between proteins and fragments range from 2 μM to 5000 μM affinity values[23], hence the requirement to apply molecular methods such as NMR, crystallography, or SPR to detect hits in fragment based screens. These sensitive methods which generate invaluable information regarding target-fragment interactions at the molecular level, however, come with some stringent requirements when it comes to sample preparation and assay development.

Clearly, to obtain high resolution information at the molecular level, the target protein sample has to be highly concentrated and purified from its environment in order to increase signal to noise levels. Naturally, individual applications have their additional requirements, such as the need for high resolution crystals and functionally immobilized proteins in crystallography and SPR, respectively. Developing individual protocols for each target protein sample preparation and assay development can be extremely time consuming and often cause large amounts of proteins to be needed in the course of action. Furthermore, large amounts of fragments are often also required.

Crystallography requires 10 – 50 mg of protein, often with purity higher than 95 %, and the resulting crystals do not always diffract appropriately to provide high resolution structural information[24]. Screening involves either co-crystallizing targets with fragment cocktails, or soaking target crystals in fragment cocktails, with individual fragment concentrations ranging between 25 – 100 mM[24]. Nonetheless, there have been numerous success stories with soluble

protein targets at the heart of cancer, including kinases[25,26] and AIDS, such as reverse HIV transcriptase[27].

NMR-based fragment screening is carried out in solution and involves detecting changes in ^{1}H, ^{13}C, or $^{1}H^{15}N$ correlation signals of the protein target or the ligand. Target-based screening such as 'SAR (structure-activity relationships) by NMR' developed by Abbott laboratories enabled identification of inhibitors for cancer target tyrosine phosphatases[28] and tissue target matrix metalloproteases[20]. It is reported that 50 – 200 mg of isotopically labelled protein, soluble within the range of 0.1 - 1 mM, are required for such experimentations. These high quantities are necessary in order to screen one fragment at a time on individual protein samples, due to the inherent system which only identifies structural, albeit valuable, information regarding the molecular effects on the protein and not the ligand. Not surprisingly, the availability of a high resolution structure and assignments for the labelled protein is a prerequisite to establish the structural effects upon fragment binding and to identify the binding site for future chemical elaboration of hits into leads. There are alternative strategies which require smaller amounts of unlabelled protein such as ligand-based screening methods where changes in the magnetisation environment of the ligand are monitored, either by magnetisation saturation transfer (STD)[29] or by proton relaxation differences[30]. These systems enable higher screening throughput due to the smaller amounts of unlabelled protein required (down to μM concentrations) and the possibility of identifying hits within a cocktail of fragments. The changes are often monitored by comparing fragment spectra in the presence and absence of the target and therefore, ligand libraries must be designed to allow appropriate signal deconvolution, or isolation of signals[31]. These NMR techniques can be carried out with off-the-shelf materials and rarely need as much effort in assay development as for crystallography and SPR techniques.

Although SPR techniques were only used for secondary screening due to limitations in throughput capacities, new advances in microarray immobilization and SPR imaging techniques are providing platforms which can now enable SPR-based primary screening of fragment libraries[32]. SPR measures changes in surface electromagnetic waves refraction indices upon adsorption of an analyte to a particular surface, such as with or without an immobilized target protein. Although this was limited to detect only differences caused by adsorption of large

biomolecules, such as target proteins or DNA, advanced techniques now enable fragment binding to an immobilized target. SPR requires high efforts of assay development for each target protein, in order to enable monitoring and to control the stability of the surface, the levels of functional protein immobilized, the level of unbound protein leakage, and levels of non-specific binding of fragments to the surfaces[33]. Nonetheless, once the appropriate assay is developed, very little amounts of protein are required with down-scaling of surface to several squared nanometers in size. Furthermore, labelled analytes or proteins are not required.

Why are membrane proteins such important drug targets?

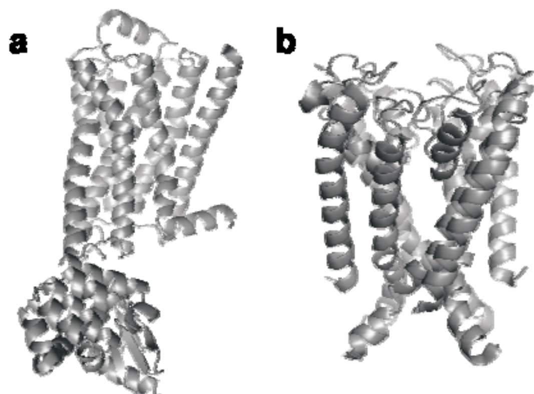

Figure 6. Structures of GPCRs and ion channels. The human GPCR β_2-adrenergic receptor (**a**) and the bacterial potassium channel KcsA which has close homology with human ion channels (**b**) have α-helical transmembrane domains. The structures are modified from pdb IDs 2rh1[34] and 2kb1[35] respectively.

60 % of today's drugs target membrane proteins[36]. As their name indicates, these proteins are located within the native cell membrane and are therefore involved in a multitude of cellular processes related to signalling, transportation, energy production, metabolism, and homeostasis. Clearly, if any of these proteins within these complex networks malfunctions, onsets of a variety of symptoms and diseases can occur.

Two major groups of membrane proteins are currently targeted due to their extremely important roles in maintaining healthy cellular processes. These are G protein-coupled receptors (GPCRs) and ion channels (Figure 6).

GPCRs are proteins which span the membrane with 7 transmembrane α-helices (Figure 6**a**) and are involved in translating an external stimulus (e.g., light, hormones, or neurotransmitters) into an internal signal in the cell. This internal signal is regulated by a signalling cascade, starting with the G protein to which the GPCRs are coupled to on the inside of the cell, down to complex

networks of other effector proteins such as kinases, ion channels, transporters, and enzymes. As such, GPCRs are at the heart of almost all physiological processes and are targeted by 30 - 40 % of drugs on the market today[37]. Faulty GPCR signalling, often due to lack of or surplus of neurotransmitters or hormones, leads to many common human disorders associated with the central nervous system, the cardiovascular system, vision, asthma, allergies, and the immune system[38], to name but a few. Targeting GPCRs in drug discovery is far from simple, for the pharmacology mechanisms are complex and diverse. The active, inactive, or self-activating (constitutive activity) states of GPCRs is determined by the type of ligand and where it binds on the GPCR. Agonists and antagonist ligands will therefore activate or inhibit the receptor, respectively. However, research has revealed constitutive activity of some GPCRs, which translates to a basal activity independent of ligand binding. Inverse agonists are therefore becoming interesting drugs as they block this type of signalling cascade and may at times be more effective than (neutral) antagonists[39]. To make matters more complex, small molecules called allosteric modulators have been noted to modulate the way these ligands bind into stronger stimulation or inhibition of activity[40].

GPCRs which are targeted by drug discovery include the Class A (rhodopsin-like) receptors and bind a range of molecules from small molecules to large peptides[41]. Some of these Class A receptors such as the adenosine receptors, histamine receptors, and the β-adrenergic receptors are good candidates for small molecule drug discovery because of the small sizes of their endogenous ligands[9], which bind to the receptors with much higher affinities than fragment interactions. Other classes of GPCRs targeted by drug discovery, such as the Class C GPCRs, have complex mechanisms such as the involvement of large N-terminal domains (Class C GPCRs) which act as gates to ligand binding only upon dimerization of the GPCR and its N-terminal domain[9]. Although it is beyond the scope of this introduction to discuss all the types of GPCR mechanisms which are targeted by drug discovery, it is clear that GPCRs are a crucial class of proteins to modulate in order to alleviate diseases. Unfortunately, to date, there are only 4 GPCRs for which complete structures are available, including the crystal structures of bovine rhodopsin[42], the β_2-adrenergic receptor[43] (Figure 6**a**), the β_1-adrenergic receptor[44], and the adenosine A_{2A} receptor[45]. Although the drug discovery process has enabled the successful

identification of ligands with mutagenesis, computational methods, and functionality based assays, the structural information of the GPCRs has provided new insights on important key properties of each individual GPCR binding site. Therefore, there is a need to obtain more structural information for other members of this class of proteins to understand where ligands bind, how to evolve them into more specific drugs, and how to modulate GPCR activity in the context of complex signalling interactions.

The second important class of membrane proteins targeted by drugs, the ion channels, are pore-forming proteins which are involved in the transport of sodium, potassium, calcium, and chloride ions through the cellular membrane in order to modulate the ionic potential between the outside and inside of the cell. These ion channels can be 'gated' by different processes such as ligand binding, temperature, mechanical tension in the membrane, and voltage, all of which are often concerted by other proteins in the cell, such as GPCRs[9]. Other molecules which can block ion channels include the deadly snake venom which consists of protein and peptide toxins that physically block the opening of the ion channels, thereby blocking all further transport of ions. Different venoms are being studied now in the hope to modulate this ion channel blockage in diseases such as multiple sclerosis[46]. Needless to state therefore how complex it is to target ion channels without creating undesirable secondary effects. Nonetheless, current drugs on the market are alleviating diseases caused by mutations in these ion channels, such as high blood pressure and muscle dystrophy, faulty acidification levels which may lead to diarrhoea or kidney stones[47], and naturally many diseases linked to the central nervous system such as migraines, seizures, and autism[48]. This was just a short list of examples, which are mostly studied by mutagenesis and 'macroscopic' assays including cell-based, radioligand, or fluorescent-based assays instead of by molecular methods due to the limited amount of complete ion channel structures currently available. Some of the complete structures that exist include the crystal structure of the potassium channel from the gram-positive bacterium *Streptomyces lividans*[49], as shown in figure 6**b**.

Clearly, with GPCRs and ion channels being key proteins of important signalling cascades, current drugs often have high levels of secondary effects. To make matters more complicated, many GPCRs and ion channels work by forming complexes of proteins by homo or hetero-

multimerization[9,50] in order to proceed with certain functions. Research is now unveiling different parcels of complex networks involved in cellular processes regulated by these important membrane proteins, along with important interactions with other components of the cell, such as molecules within the native membrane itself. The dawn of biophysical molecular methods will eventually lead to a better understanding of how to specifically target a membrane protein in a specific state and how to improve drug interactions for a particular system rather than a single protein target. As explained below, there are many reasons why membrane proteins are still difficult to study at the structural level, limiting therefore our possibilities to develop drugs which are more specific in the aim to limit side effects and understand new relationships between disease and target.

Why are membrane proteins generally excluded from fragment based drug discovery?

As previously defined, molecular approaches such as NMR, crystallography, and SPR require high yields of concentrated samples of protein target in order to yield high resolution data regarding molecular interactions. Furthermore, the sensitivity of these methods is such that extensive effort goes into preparing protein samples with the highest level of purity, so as to avoid unwanted interactions with, or signals from, other molecules.

Applying such molecular methods to membrane proteins becomes challenging because of the difficulties in meeting the above requirements, as explained below. It is rarely possible to produce membrane proteins in sufficient quantities, and even when enough is produced, the proteins must be purified, concentrated, and solubilized in a hydrophobic environment which leads to loss of protein stability and an increase in unwanted signals and interactions.

Limitations in membrane protein production

There is no general solution for production of membrane proteins in high quantities and the choice of vector sources and host cells for overexpression and production has to be defined for each new target. Such decisions may be based on the source of the original gene, the protein's requirements for optimal folding, and for posttranslational modifications. Although prokaryotic cells are the most productive hosts for soluble proteins, they often do not possess the machinery required to produce membrane proteins in the correct conformation with suitable posttranslational modifications[51]. Low success rates with prokaryotic cells are also often caused by toxicity induced by overexpressed levels of foreign membrane proteins[52]. Although there have been success stories where inclusion bodies were used to produce high levels of membrane proteins in *E.coli*[53] and cell-free *E.coli* based expression systems[54], there is an extensive amount of work required to find the correct conditions for refolding membrane proteins into stable and functional conformations.

Recombinant expression of human genes in eukaryotic mammalian cells is therefore a more appropriate system to use, but the yields generally obtained remain too low for studies involving molecular methods. GPCRs for example can be obtained in ranges from 10 - 100 fmol/mg of tissue[55], and the lack of purification and solubilization techniques makes it difficult to purify them in abundant amounts from natural sources[56]. Alternative eukaryotic cells which can produce sufficient amounts of membrane proteins for molecular methods include baculovirus infected insect Sf9[57] and yeast[58] cells. The latter was quite a novelty because it proved that, against old beliefs, some GPCRs can be fully functional without glycosylation which is incompatible with crystal formation, and could be produced in high enough levels to be put through crystallization trials.

There is clearly much effort toward overexpression of membrane proteins in high quantities, but the challenge is far from over, for mimicking the native membrane qualities in vitro is the most challenging of all, and has to be faced in each step of the process from membrane extraction to protein solubilization, and, if required by the molecular method, functional and stable immobilization.

Membrane mimics for in vitro handing of membrane proteins

The native membrane which surrounds each living cell is constituted of a lipid bilayer and an array of proteins, either transmembrane or associated with the membrane, which function in concert to create, organize, and complete cellular processes. The membrane does not only enable compartmentalisation of the cell's inner processes and components, but it is actively involved in regulating these processes by interacting with the associated proteins[59]. Although we have begun to understand this important role, elucidating each complex interaction between the different components of the membrane and individual membrane proteins is an extremely difficult task at hand, for the membrane is a complex system. For example, lipid rafts are dynamic arrangements of lipids and cholesterol which have recently been identified as important players in the regulation of membrane receptor activity and localization within the membrane[59]. To study a membrane protein at the molecular level by NMR, crystallography, or SPR, clearly, its isolation from this intricate system is required in order to have a pure sample of the protein of interest.

A membrane protein has a large amount of hydrophobic residues which are in contact with the phospholipids in the native membrane and it is estimated that the free enthalpy cost of solubilizing a membrane protein in water would be in the range of 150 – 200 kcal/mol[60]. Thus, when a membrane protein is removed from its native environment, the hydrophobic domains which were stabilized by phospholipids will be attracted to each other, causing the protein to collapse and precipitate. This is why it is crucial that all steps leading from initial extraction from the membrane to purified protein is meticulously carried out so that the target protein is reconstituted in a synthetic lipid environment which mimics the characteristics of the native one as closely as possible[61]. Which strategies to use depends on the individual protein: some may remain stable only in the native membrane, while others may be reconstituted in the simplest synthetic forms.

Approaches for solubilizing membrane proteins in vitro require the presence of amphipathic molecules which mimic the membrane phospholipid properties by presenting hydrophilic head groups to the aqueous buffer, while maintaining contacts to the proteins' hydrophobic residues with the hydrophobic tail groups. These approaches range from simple addition of high

concentrations of ionic or zwitterionic surfactants such as detergents or lipids in water to create micellar or bicellar vesicles in all handling steps[61] to very complex protocols involving the fusion of various mixtures of detergents/lipids to form a more stable and better mimic of the membrane in lipid bilayers for example. The most popular detergents include alkyl glucosides and maltosides, polyoxyethylenes, alkyldimethylamines, and cholate derivatives such dodecylphosphocholine (DPC)[52] with which micelles are formed when the surfactant is in a higher concentration than its critical micellar concentration (CMC). Below the CMC, the equilibrium shifts from micellar to monomeric forms of detergent, thereby causing loss of protein conformation and functionality as the proteins precipitate in the absence of stabilizing micellar formations. Often, detergents are used to extract membrane proteins from the native membrane, because they are good at dissociating lipids from proteins[60]. While some proteins can continue to be handled in these detergents throughout the purification procedure until the final sample for crystallization, NMR, or immobilization for SPR, others require a more stable lipidic environment in bicelles (micelles formed of lipid bilayers) or bilayers. Molecular methods have therefore been difficult to apply to membrane proteins because the environment which keeps a protein stable may not be compatible with the application we one wishes to use to study the protein.

Detergent micelles often lead to undesirable effects such as low stability or even denaturation, aggregation, and separation of subunits from multimeric formations[52]. This low stability is clearly incompatible with NMR and crystallography as the dynamic processes lead to difficulties in obtaining high resolution NMR data and loss of the initial 'true' protein structure, and prevents crystals from forming. Furthermore, detergent micelles can be too large upon solubilization of membrane proteins to successfully apply NMR[61]. The high levels of detergents necessary to maintain protein conformation can lead to unwanted signals and interactions, such as non-specific binding to the compounds used in drug discovery during SPR or solution NMR screening applications.

Although detergent micelles have been used successfully in applying NMR to the structural determination of the outer membrane protein A in DPC[53] and the Disulphide bond forming protein B[62], the use of more stable lipid formations such as lipid bicelles[63] and lipidic cubic

phases[64,65] are required for crystallization. Strategies are also evolving to include immobilization of detergent or lipid solubilized membrane proteins onto solid supports for SPR[66] or solid state NMR[67] applications. However, although these applications have enabled structural determination of membrane proteins, they are limited to proteins which are obtainable in high amounts, such as bacterial membrane proteins, or require great efforts into finding the appropriate solubilization condition which is both appropriate for protein fold and functionality, but also compatible with the method. When drug discovery is involved however, more difficult targets such as GPCRs and ion channels are involved, which often call for delicate dimerizaton states or the presence of other players in the membrane (G proteins) for a full read out of the activity. Furthermore, the synthetic membrane mimics often lead to surfaces to which fragments bind non-specifically. For all these reasons, to date, there have been no successful applications of fragment based drug discovery to membrane proteins.

New alternative solubilization strategies for membrane proteins in aqueous buffers

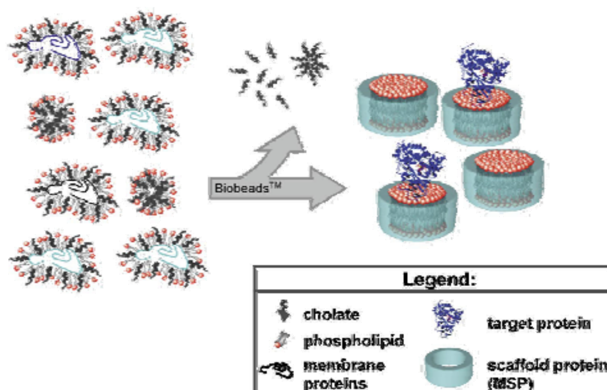

Figure 7. Self-assembly of Nanodisc complexes, reproduced from reference 68 with permission. Copyright 2003 Biotechniques. The detergent micelles of the target protein are replaced by a stable lipid bilayer surrounded by the amphiphilic MSP upon detergent removal. The resulting nanodisc complex is fully soluble in aqueous buffers.

Although there have been important developments contributing to more stable membrane proteins in vitro (better control of the expression systems[69], mutagenesis of key residues for higher protein stability[70] and better solubilization techniques[71]), these all require the presence of detergent. The self-assembly of membrane proteins in new solubilization alternatives such as the amphipols[72] or the Nanodisc[68] complex (Figure 7), however, enables handling of membrane proteins in aqueous buffers without needing to add surfactants to maintain a hydrophobically stable environment.

The Nanodisc procedure consists of mixing lipids which are solubilized in mixed micelles with cholate, and a 23 kDa amphiphilic α-helical membrane scaffold protein (MSP) to the targeted detergent-solubilized membrane protein to be incorporated into the complex. When these entities are mixed together, the lipids, detergents, and cholate form mixed micelles around the MSP and the target protein.

Upon removal of detergents, by dialysis or the addition of detergent adsorbing bio-beads, the target membrane protein, the lipids, and the MSP hydrophobic residues self-assemble into the nanodisc formation. When the appropriate conditions are met, which are different for each target membrane protein to be incorporated, the target protein is solubilized in the center of a lipid bilayer which is consequently stabilized by two monomers of MSP with the hydrophobic residues

making contact with the lipids. Due to the amphiphilic nature of the nanodisc, the final complex can be handled in aqueous buffers without causing conformation or functional loss of the embedded target protein and without requiring any presence of surfactant in the buffer. This enables easy purification and has the potential to be applied to fragment based drug discovery without causing non-specific binding of ligands to the nanodisc complex.

The nanodisc technology has been successful in embedding a variety of membrane proteins such as Bacteriorhodopsin[73], the GPCR β-adrenergic receptor[74], and the metabolically important cytochrome P450 (CYP)[75]. The nanodisc system was also previously used to measure redox potentials on CYP3A4 in the absence and presence of substrates at the active site[76]. Although the procedure of incorporation depends on every target protein and the initial solubilization state of the target protein, the resulting homogenous and standard preparations, void of any surfactant and stable in aqueous buffers, may bridge the gaps which prevent molecular methods from being applied to membrane proteins.

How can Target Immobilized NMR Screening address limitations posed by membrane proteins?

Target Immobilized NMR Screening is, as its name indicates, is a FBDD method which screens an immobilized target for fragment binding using solution NMR. The method involves immobilization of a target protein and a reference protein on an aldehyde resin with a mild Schiff's base chemistry between the primary amines of the protein and the aldehyde groups present on the commercially available resin. A flow-injection, dual sample holder can be placed in the magnet to enable flow-mediated screening of fragments by automated injection of mixtures containing upto 8 fragments over the immobilized proteins. This method enables fragment based drug discovery on membrane proteins for the following reasons.

Immobilization allows efficient use of the target protein

With the target and reference protein immobilized, a single screen of 1000 fragments can be carried out on one single sample of protein. Naturally, because a full screen can last approximately 5 days, the sample may need to be replaced if the protein shows signs of low stability. However, the amounts required are fairly low (50 – 100 nmoles). This procedure is generally applicable to other membrane proteins because of the simple and mild Schiff's base chemistry used to immobilize the protein N-terminus to the aldehyde groups on the resin[23], and because it is compatible with the use of detergents.

NMR detection of fragment signals rather than protein signals

TINS focuses on the ^{1}H 1D spectra of fragments in solution[77]. The difference in intensities of these fragments' peaks between the presence of an immobilized target protein and the presence of an immobilized reference protein allows one to identify if a fragment has bound to the target. The fact that the method directly observes differences in fragment signals in solution means that there is no necessity to produce isotopically labeled proteins and there is no *a priori* requirement for the protein structures to be well resolved, as is necessary in other NMR target-based screening methods. Furthermore, compared to other ligand based screening methods, there is no need to deconvolute the signals, because mixtures are designed to be composed of ligands with a minimum of overlapping peaks.

The reference system accounts for non specific binding of fragments to detergents

When TINS is applied to soluble proteins, the level of non-specific binding to the target protein can be accounted for by simultaneously screening a reference protein of a relatively similar size which has limited small molecule binding properties. With membrane proteins however, the presence of detergent leads to non-specific interactions with the fragments both in solution and upon contact with the immobilized, detergent solubilized protein. It is easy to predict that the

amphipathic nature of the detergents can partially solubilize the fragments (which are relatively soluble yet also composed of hydrophobic moieties) out of solution and temporarily into micelles. To minimize such interactions, the appropriate trade-off can be met, by adding detergent into the wash buffer but not into the independent fragment mixtures, provided the fragment mixture injection does not dilute the effective micelle concentration below local CMC. As stated, this minimizes interactions between fragments and detergents, but can not be eradicated due to the loss of protein functionality which would follow. Therefore, there remains a substantial amount of partial solubilization of fragments into the detergent micelles, with the effect of reducing their effective concentrations and hence, their final intensities in the NMR spectra. The reference membrane protein must therefore not only have minimal small molecule binding properties, but it must be refolded in the same detergent conditions as the target protein in order to obtain a reliable reference system. In our case the outer membrane protein A (OmpA) (Figure 8**a**) which can be folded properly in a variety of detergents[72,78,79], was found to have minimal small molecule binding. This reference system is theoretically applicable to membrane proteins and removes signals from fragments which are only non-specifically binding to the detergent micelles. To minimize signals from detergents with a high CMC, deuterated forms can be used, provided they are available.

Alternative immobilization and solubilization methods can be easily adapted

TINS can be easily adapted to different immobilization chemistries and membrane protein solubilization strategies, provided the appropriate resin is used. In our case, the sepharose aldehyde resin has proven to be compatible with the system by having a minimal line broadening effect on the fragment ^{1}H 1D spectra. A variety of chemical linkers exist which can be used to tailor the immobilization chemistry based on the requirements for functional immobilization of the target protein, such as targeting the C-terminus when the N-terminus is involved in protein functionality. Naturally, the immobilization chemistry has to be compatible with the detergents or alternative solubilization materials used to maintain the correct fold and functionality of the membrane protein. nanodiscs, as an alternative solubilization strategy to DPC micelles, have

potential to be immobilized to the same resin through the N-termini of the scaffold proteins. We envisaged therefore that TINS would be compatible with screening membrane proteins embedded in nanodiscs in the absence of detergents, providing a more stable environment for the target protein without the disadvantages of non-specific binding interactions of fragments with detergent micelles.

The aim of this thesis:
To develop methodologies which enable the application of molecular methods for drug discovery on membrane proteins.

With increasingly perceptive methods for solubilizing and immobilizing membrane proteins in functional ways, the different direct or indirect, random or oriented immobilization strategies are reviewed in **Chapter 2**. Attention is paid to a variety of applications such as chromatography, fluorescent and radioligand based applications, as well as molecular methods such as NMR and SPR.

Prior to starting a TINS screen, we first wanted to test the functionality of an important class of membrane proteins, the GPCRs. Knowing how important the native membrane is for GPCR functionality, we aimed at testing functional immobilization of the human histamine H_1 receptor and the human adenosine A_1 receptor within the native membranes from stable cell lines overexpressing these receptors. Radioligand binding studies of the receptors in solution were compared to the receptors upon immobilization within their native membrane vesicles in the presence of physiologically relevant G proteins, as described in **Chapter 3.**

The H_1 receptor mediates a variety of cellular processes which, upon faulty signalling, are at the heart of many diseases affecting children and adults today. These include immunological responses[80] such as allergies and asthma, as well as eating disorders such as anorexia[81]. Furthermore, antihistamines which are used as potent blockers of histamine induced allergic reactions provoke drowsiness due to the additional role of the H_1 receptor in regulating sleeping behaviour[82]. The A_1 receptor is just as important a pharmaceutical target as the H_1 receptor, because of its role in the central nervous system as a regulator of adenosine and hence, the physiology of many tissues. The adenosine A_1 receptor is ubiquitously found in the whole body and is at the heart of processes leading to asthma, chronic inflammatory problems, and heart failure[83] in addition to, when located in the central nervous system, many types of

neurodegenerative diseases[84]. It has also been reported to form dimers with itself or the dopamine receptor[85]. Both receptors belong to the Class A of GPCRs and generally bind small ligand molecules, thereby making them *a priori* appropriate targets for fragment based drug discovery. Although the structures for these particular proteins do not exist, the homology with other known structures currently enables *in silico* studies of ligand binding interactions[86].

Currently TINS has been explored on soluble proteins. We wished however to prove the concept that detergent solubilized membrane proteins could practically be screened by the TINS methodology as well. In **Chapter 4**, we carried out an initial pre-screen on a bacterial ion channel, the potassium ion channel KcsA, and a bacterial membrane enzyme, the Disulphide Bond Forming protein B (DsbB) with, as a reference, the immobilized bacterial Outer Membrane Protein (OmpA) also solubilized in DPC.

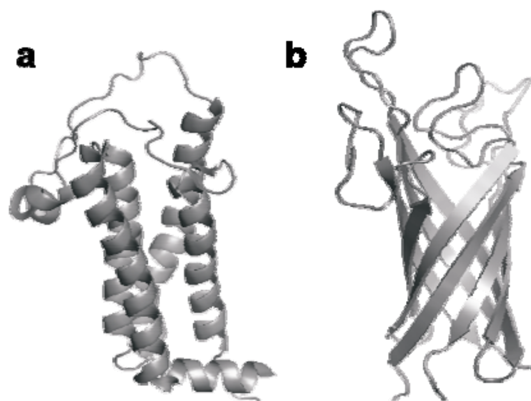

Figure 8. Structures of the target and reference proteins used in the TINS screen. DsbB has α-helical transmembrane domains (**a**) while OmpA has a β-barrel transmembrane domain (**b**). The structures are modified from pdb IDs 2k73[62] and 1g90[53] respectively.

The reference protein, located in the outer membrane of bacteria, is a β-barrel transmembrane protein (Figure 8**b**), which has a variety of roles leading to pore formation, adhesion, invasion, formation of biofilms, and as a receptor for several types of bacteriophages[87]. However, these functions are related to the extracellular domain of the protein. We believed that the transmembrane domain, which has also been studied by NMR[53,72] and crystallography[88,89], however, has minimal small ligand binding and could act as a reliable reference protein to account for non-specific binding of fragments to the detergent micelles due to its similar size and appropriate fold in DPC micelles.

Both target proteins for this TINS pre-screen are from bacterial origin, enabling us to obtain sufficient amounts for initial development of the protocol needed to adapt the TINS procedures to membrane proteins. Furthermore, the crystal and NMR structures of KcsA[90,91] and DsbB[62,92] are available, enabling future studies of protein-ligand interactions on these pharmaceutically important proteins. DsbB is a bacterial membrane protein target (Figure 8**a**) which has pharmaceutical importance due to its role in disulfide bond formation and subsequent regulation of protein toxin folding leading to the virulence of Gram negative bacteria such as *E.coli*[93] and *Bordetella pertussis*[94]. KcsA is a model protein with high homology with human potassium channels which are important pharmaceutical targets due to their role in regulating potassium ion transport through the cell membrane. As such, human potassium channels, are targeted by anaesthetics and drugs which partially alleviate autoimmune and neurodegenerative diseases[95]. On the other hand, ion channels are often non-specifically and unexpectedly targeted or blocked by drugs, causing severe secondary side effects. An example of such a challenging protein is the human ether-a-go-go related gene (hERG) potassium channel which regulates the potassium potential in cardiac myocytes[96]. Drugs not related to the cardiovascular system have been removed from the market because they either blocked the hERG channel activity by non-specifically binding to the large inner binding pocket, or by blocking its traffic to the cell surface, both leading to irregular (and potentially fatal) repolarization of the cardiac muscle[97,98]. It would therefore also be valuable to carry out parallel screens on liable targets such as the hERG channel in order to predict and avoid future drug candidates' interactions.

With the proof of principle validated on membrane proteins, we carried out a full screen on the pharmaceutically important DsbB with OmpA as a reference, in DPC micelles. The identification and validation of the hits on the membrane enzyme DsbB were possible due to an existing robust enzymatic assay which relies on the detection of substrate oxidation or reduction[99]. A soluble partner protein DsbA becomes oxidized in the presence of DsbB and the synthetic DsbB cofactor ubiquinone (UQ1). The enzymatic turn-over rate of DsbB can therefore be compared in the presence of fragment hits in order to validate their inhibitory effect on this target protein. The

results of the enzymatic assays, along with confirmation of the binding modes of fragments by orthogonal biophysical assays (by solution NMR Heteronuclear Single Quantum Correlation) are described in **Chapter 5**.

The possibility of carrying out TINS on DsbB in aqueous buffer completely void of surfactants was tested using the nanodisc system, whereby the membrane proteins can be solubilized in a lipid bilayer stabilized by a belt of two amphiphilic membrane scaffold proteins. The results of a short screen on nanodisc-solubilized DsbB was tested using two references, nanodisc-solubilized OmpA and empty nanodiscs, containing only lipid bilayers and the amphiphilic scaffold proteins. The results are described in **Chapter 6**.

Finally, a conclusion section is presented in **Chapter 7**, including perspectives regarding the future of membrane protein research when applying biomolecular methods in drug discovery.