THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN NATURAL SCIENCE

Intrinsically Disordered Domains of the B Cell Receptor

Cell-Free Expression and Characterization by NMR

LINNÉA ISAKSSON

UNIVERSITY OF GOTHENBURG

University of Gothenburg
Department of Chemistry and Molecular Biology
Göteborg, Sweden, 2014

Thesis for the Degree of Doctor of Philosophy in Natural Science

Intrinsically Disordered Domains of the B Cell Receptor
Cell-Free Expression and Characterization by NMR

Linnéa Isaksson

Cover: Schematic interpretation of transcription and translation of an isotopically labeled polypeptide for NMR analysis.

ISBN 978-91-628-9149-7
Available online at http://hdl.handle.net/2077/36703

Department of Chemistry and Molecular Biology
University of Gothenburg
P.O. Box 462
SE-405 30 Göteborg
Sweden

Printed by Ale Tryckteam AB
Göteborg, Sweden 2014

TILL MIRSAD OCH ARVID

Abstract

After the last twenty years of research, the occurrence of flexible proteins without a fixed three-dimensional structure are no longer considered to be rare exceptions from the structure-function paradigm. Instead, intrinsically disordered proteins (IDPs) have become one of the most interesting subjects of modern protein research. NMR is the best and most suitable technique for investigating the details of this protein class, and cell-free protein synthesis (CFPS) offers several advantages compared to conventional *in vivo* synthesis for the production of IDPs.

In this thesis, an integrated approach for efficient characterization of IDPs has been developed, combining CFPS and novel NMR methodology with fast spectroscopy and self-validating automatic assignment procedures. The technique has been demonstrated on disordered cytosolic domains of the B cell- and the T cell receptor. These domains are responsible for signal propagation into the immune cells, initiated by phosphorylation of tyrosines in their *i*mmunoreceptor *t*yrosine-based *a*ctivation *m*otifs (ITAMs). Secondary structure propensities have been observed and followed, going from a non-active form (non-phosphorylated) to an active form (phosphorylated) for the domains of the B cell receptor. A time-resolved technique for studying phosphorylation has also been developed and demonstrated on a B cell receptor domain.

Isotopic enrichment of amino acids is often a prerequisite for studying proteins with NMR, also representing the major cost of the CFPS system. A way to efficiently incorporate these labeled amino acids has therefore been investigated in this work.

CFPS does not only provide a unique technique for producing protease-sensitive IDPs, but also membrane proteins (MPs), inherently difficult to express in functional form. In this work it is demonstrated that CFPS can be successfully applied to express preparative amounts of co-solubilized MPs of varying size and complexities.

List of publications

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I. Isaksson L., Enberg J., Neutze R, Karlsson B.G., Pedersen A. (2012) Expression screening of membrane proteins with cell-free protein synthesis. *Protein Expr Purif*, Mar; 82(1): 218-25

II. Isaksson L., Mayzel M., Saline M., Pedersen A., Rosenlöw J., Brutscher B., Karlsson B.G., Orekhov V.Y., (2013) Highly efficient NMR assignment of intrinsically disordered proteins: application to B- and T cell receptor domains. *PLoS One*, May 7; 8(5): e62947

III. Rosenlöw J., Isaksson L., Mayzel M., Lengqvist J., Orekhov. V.Y. (2014) Tyrosine phosphorylation within the intrinsically disordered cytosolic domains of the B-cell receptor: an NMR-based structural analysis. *PLoS One*, Apr 25; 9(4): e96199

IV. Mayzel M., Rosenlöw J., Isaksson L., Orekhov V.Y. (2014) Time-resolved multidimensional NMR with non-uniform sampling. *J Biomol NMR*, Feb; 58(2): 129-39

V. Isaksson L., Pedersen A., Karlsson B.G. (2014) Improving amino acid incorporation efficiency with cell-free protein synthesis. *Manuscript*

Contribution report

I: I was involved in the entire project. I cloned all constructs, produced them, ran Western blots, purified the protein and ran CD spectroscopy. I also took part in writing the paper.

II: I planned the project, cloned all constructs, produced them and set up purification schemes for all included targets. I took part in NMR measurements and analysis of the data. I wrote a major part of the manuscript and prepared figures.

III: I planned the project, produced and purified the proteins. I took part in NMR measurements and analysis and in writing the paper.

IV: I produced and purified the proteins, analyzed data and proofread the paper.

V: I took part in planning the project. I produced and purified the proteins, ran NMR measurements and made the analysis. I also took part in writing the paper.

Abbreviations

BCR	B Cell Receptor
CD	Circular Dichroism
CECF	Continuous-Exchange Cell-Free
CFCF	Continuous-Flow Cell-Free
CFPS	Cell-free Protein Synthesis
CH	Charge-Hydropathy
CK	Creatine Kinase
CP	Creatine Phosphate
DAM	Dissociation Activation Model
DNA	Deoxyribonucleic acid
FID	Free Induction Decay
HSQC	Heteronuclear Single Quantum Coherence
IDP	Intrinsically Disordered Protein
IMAC	Immobilized Metal Ion Affinity Chromatography
ITAM	Immunoreceptor Tyrosine-based Activation Motif
MDD	Multi-Dimensional Decomposition
MHC	Major Histocompatibility Complex
MoRF	Molecular Recognition Feature
MP	Membrane protein
MTSL	*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NUS	Non-Uniform Sampling
PDB	Protein Data Bank
RNA	Ribonucleic acid
SAIL	Stereo-Array Isotope Labeling
SCS	Secondary Chemical Shift
SCHOOL	Signaling Chain Homooligomerization
TA	Targeted Acquisition
TANSY	Targeted Acquisition NMR Spectroscopy
TCR	T Cell Receptor
TEV	Tobacco Etch Virus
TROSY	Transverse Relaxation-Optimized Spectroscopy

Table of contents

Introduction

Virtually every property that characterizes a living organism is affected by proteins. Proteins are the key actors in the cell, carrying out the duties specified by information encoded in the genome. Proteins store and transport a great variety of substances in the cell, ranging from electrons to large macromolecules. They transmit information between cells and organs, control the passage of molecules across cellular membrane as well as registering what is going on in the surroundings and adjust cellular activities accordingly. Proteins also sustain life by catalyzing chemical reactions, controlling gene expression and they are necessary for our sight, hearing and other senses. Another crucial function of proteins is their involvement in the immune system. Since pathogens can cause fatal infectious diseases, the operation of proteins in the immune system can be a matter of life and death.

Proteins are linear polymers built of various combinations of 20 different amino acids. The numerous arrangements of amino acids, with their diverse chemical characteristics, and the three-dimensional structure that the polypeptide can form, make the vast array of functions protein perform in living organisms possible. Information for the various combinations of amino acids and the structure of proteins are found in genes encoded in DNA. By transcribing DNA to messenger-RNA, that convey the genetic information from DNA to the ribosomes, biosynthesis of proteins can occur by the process called translation. Complex biomolecules, e.g. proteins, built from hundreds of amino acids, cannot be synthesized even by contemporary state-of-the art organic chemistry. Instead the biological mechanisms that generate them in living cells have to be used or recreated, and different protein production systems are used in the field of life science.

In the present work, the immune receptor from B cells has been characterized, with focus on the cytosolic intrinsically disordered domains CD79a and CD79b. A NMR spectroscopy platform for efficient characterization of these disordered proteins and other biomolecules belonging to the same protein class has been established (paper II) and a way for studying real-time events in 3D dimensions has been set up (paper IV). Phosphorylation (paper III) and transient secondary structure (paper II) have been investigated for CD79a and CD79b. The protein production system used in this work has been cell-free expression. The strength of this technique is proven for both membrane proteins (paper I) and for disordered proteins (papers II-IV), and the efficient incorporation of isotopically labeled amino acids with the cell-free expression system has been demonstrated (paper V).

Membrane proteins

Integral membrane proteins (MPs), embedded in the membrane, provide critical roles in cell-to-cell contact, cytoskeleton contact, surface recognition, signaling, enzymatic activity and transportation of substances across the membrane [1]. MPs are also important for drug research, accounting for over 50% of all human drug targets [2]. In particular G protein-coupled receptors have intensively been targeted for therapeutic purposes. Even if nearly 30% of the proteome is comprised of MPs, only 1% of all deposited structures in the protein data bank (PDB) are MPs. This is largely due to the inherent difficulties associated with working with this class of proteins [3].

Cell membranes are permeable barriers that maintain and protect the interior of a cell, and the lipids in the membrane provide the physiological environment for MPs. Most biophysical methods for studying MPs *in vitro* requires a membrane-mimicking system to stabilize the protein, and since the function of MPs is strongly dependent on their environment, the right membrane milieu has to be selected for investigating proteins from this class [4]. Detergents (Figure 1A) are the most common membrane system for structural investigations. However, many MPs require specific types of phospholipids to maintain functional, which cannot be fulfilled by using detergents. Detergents may also lower membrane protein stability. Phospholipid liposomes can overcome these problems since they resemble a native membrane much more than detergent micelles. Unfortunately liposomes are not compatible with crystal formation for X-ray crystallography and they are too big for solution-state NMR, the two main techniques used for structural investigations of MPs [5]. Bicelles (Figure 1B) were introduced in the mid 1990s and has since then been extensively used for both solid and solution state NMR as well as crystallization of MPs [6]. Bicelles consists of a solubilized lipid bilayer formed by the addition of an amphiphile (detergent or short-chain lipid) together with a long-chain lipid. The long-chain lipid form a central planar bilayer surrounded by the amphiphile, protecting the hydrophobic edges of the bilayer. MPs have been shown to be fully functional in bicelles under physiological conditions and bicelle-protein mixtures can be manipulated with almost the same ease as micelle solubilized proteins [6]. The size of a bicelle is determined by the ratio of the long chain lipid to the short chain lipid.

Recently, nanodiscs (Figure 1C) were introduced as membrane-mimicking systems, providing a near native-like environment for MPs. Nanodiscs are non-covalent assemblies of phospholipids and a genetically engineered membrane scaffold protein, based on the sequence of the α-helical human apolipoprotein AI (Apo A-I). Two molecules of the scaffold protein wrap around the bilayer formed by the phospholipid, making a disc-

like particle, called nanodisc. This system closely resembles a native-like lipid environment and at the moment this is the only detergent-free membrane mimicking system for solution NMR spectroscopy [7]. The length of the helix of the scaffold protein determines the size of the nanodisc, and different versions of the Apo A-I have been engineered for biophysical studies of MPs in order to decrease the size of the disc, making them more suitable for NMR studies [5]. The presence of the protein belt of Apo A-I constraints the dimensions of the bilayer and make the particle size more monodispersed compared to micelle- and bicelle systems. The coat of the protein also makes the nanodisc stable over time [8].

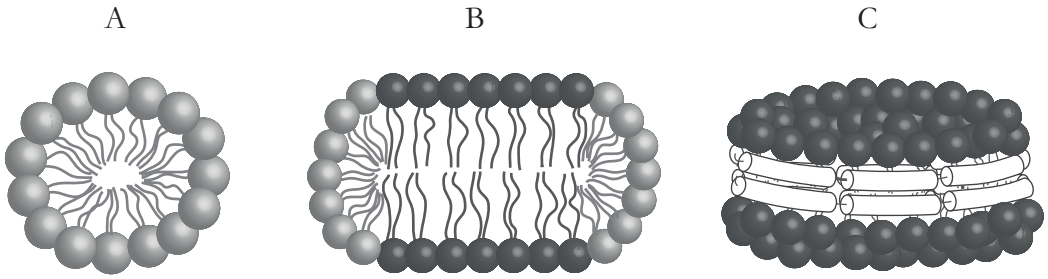


Figure 1. Cartoon of different membrane-mimicking systems. **(A)** Micelle formed in aqueous solution of detergents, orienting their hydrophilic region towards the water and the hydrophobic tails grouped in hydrophobic cores. **(B)** Bicelle consisting of a bilayer of long chain lipids (dark grey) and detergents or short chain lipids (light grey). **(C)** Nanodisc with two human apolipoprotein A1 molecules wrapped around a phospholipid bilayer.

Intrinsically disordered proteins

A long-standing belief has been that the functional properties of proteins depend upon their three-dimensional structure, the so-called structure-function paradigm [9]. The primary origin for this paradigm was the "lock-and-key" model (Emil Fischer 1894), which suggested a strict geometric complementarity of the enzyme and substrate. This theory was further confirmed with the observations that denaturation of enzymes (with e.g. acid treatment, alkali or urea), led to loss of the enzyme activity and that these denatured proteins could not be crystallized [10]. The first reports on X-ray crystallographic structures of myoglobin [11], hemoglobin [12] and the first enzyme, lysozyme [13], reinforced this static view of functional protein structures. Interestingly, exceptions to this view started to appear. Serum albumin, for example, could assume a large number of configurations, binding to different small molecules [14]. Suddenly a protein was not necessarily strictly complementary to its substrate. Casein is another important example because this was the first protein that showed to have an unfolded configuration that was important for the function of the protein [10]. An interesting case is the myelin basic protein (MBP) that failed 4600 crystallization conditions [15]. MBP was suggested to belong to the category "uncrystallizable" proteins. Another early observation of protein disorder is the microtubule-associated protein 2 (MAP2) [16], a homolog of the tau protein involved in Alzheimer's disease. This protein was among the first to be recognized as disordered and functional under native conditions. In 1988, Sigler suggested that several important transcription factors carry out their function without specific structure, instead forming ill defined "acid blobs or negative noodles" [17]. In 1995, a paper written by Gast *et al.* with the title "Prothymosin Alpha: A Biologically Active Protein with Random Coil Conformation" was published [18]. It was not just the title that became an important milestone for the field of disordered proteins, but also the question they raise in the paper: "whether this is a rare or a hitherto-overlooked but widespread phenomenon in the field of macromolecular polypeptides". The observation that a disordered cyclin-dependent-kinase inhibitor, a protein important for the p53-dependent control of cell cycle, adopted a stable structure upon interaction to its partner, was brought up by Wright *et al* 1996 [19]. This became an important element for understanding the function of disordered proteins. In the late 1990s and in the beginning of 2000s, the focus of attention started to shift towards understanding the differences in function between structured and unstructured proteins [20-23]. The observed flexible biomolecules were no longer considered to be rare exceptions from the structure-function paradigm but instead representing a very broad class of proteins. Today, a biological database

collection of intrinsically disordered proteins (IDPs) exists, called DisProt, which currently covers 694 disordered proteins and 1539 disordered regions [24].

General characteristics of disordered proteins

Disorder is common in all species, especially in eukaryotes where as much as 15-45% of eukaryotic proteins contain long disordered regions (>30 consecutive residues) [25]. Defining IDPs is, however, quite difficult. Clearly they cannot be characterized as one specific type of protein. Ordered proteins have a 3D structure that is relatively stable with Ramachandran angles that vary only slightly from their equilibrium positions. IDPs, however, show complete, or almost complete, loss of any ordered structure under physiological conditions, behaving more like random coils (Figure 2). IDPs are often defined as dynamic ensembles with Ramachandran angles that vary significantly over time with no specific equilibrium values [26, 27]. Compared to structured proteins, that have a global minimum in the conformational space, IDPs can be described as having several accessible structural states separated by low energy barriers. Many interesting observations, showing the unusual behavior for IDPs, can also be used for the definition. For instance, IDPs are often resistance to heat, not precipitating after incubation at boiling temperatures. Lowering of pH, causing denaturation of ordered proteins but not for disordered proteins, is another interesting feature of IDPs. The unusual SDS-PAGE mobility due to the atypical amino acid composition, which affects the SDS binding to the protein and thereby the migration in the gel, is another characteristic [10]. IDPs have a high content of uncompensated charged groups and a low content of hydrophobic residues, a combination that has been shown to be an important prerequisite for the absence of compact structure [21]. IDPs are also characterized by low sequence complexity, with amino acid compositional bias and high-predicted flexibility [20, 23]. Bulky hydrophobic residues like Ile, Leu and Val are depleted in IDPs as well as Cys, Asn and the aromatic Trp, Tyr and Phe. The polar residues Ala, Arg, Gly, Gln, Ser, Glu, Lys and the structure breaking Pro are instead enriched in IDPs. So in other words, disorder is encoded in the amino acid sequence.

Several disease conditions have been associated with altered levels of IDPs in cells. Examples are α-synuclein and tau protein that form aggregates upon overexpression, which is linked to Parkinson's and Alzheimer's disease, respectively. IDPs have to be tightly controlled in the cell to avoid overexpression. The levels of mRNA encoding for disordered proteins have been observed to be less abundant compared to transcripts for ordered proteins, due to an increased decay rate of IDP-mRNA. The availability of many IDPs in the cell is also regulated via a reduced translational rate and by increased proteolytic degradation [28]. Fine-tuning of this tightly regulated

system can be achieved through post-translational modifications and interactions of IDPs with other factors [22].

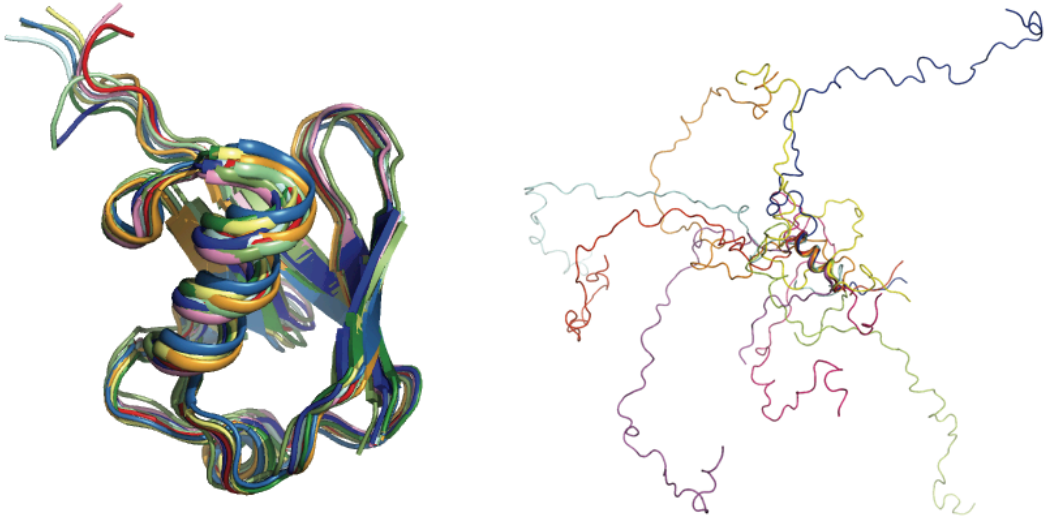


Figure 2. Structures of a well-folded protein and an IDP. To the left, the NMR ensemble structures of Ubiquitin (PDB id: 1D3Z), illustrating that similar structure are maintained over time. To the right, the NMR ensemble structures of Thylakoid soluble phosphoprotein TSP9 (PDB id: 2FFT), an IDP with different conformations that do not overlay over time due to the intrinsic dynamic behavior. The figure was made with MacPyMOL [29].

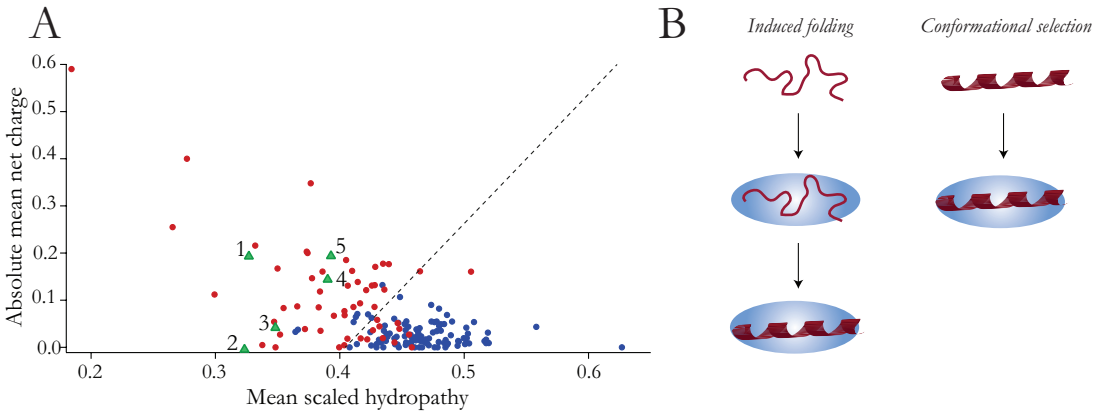


Figure 3. (A) Charge-hydropathy (CH) plot. The mean net charge and the mean scaled Kyte-Doolittle hydropathy are used to distinguish ordered from disordered protein. Clustering of ordered proteins (blue dots) are separated from disordered proteins (red dots) by a linear boundary. The plot clearly shows positions of the targets included in this thesis in the disordered region (green triangles), corresponding to a high net charge and low hydrophobicity, characteristic features of IDPs (1: CD3ε 2: CD3γ, 3: TCRζ, 4: CD79a, 5: CD79b). **(B)** Schematic representation of two folding mechanisms for IDPs. To the left, the IDP associates to its partner and then subsequently folds, so called *induced folding*. To the right, the binding protein selects a member from the ensemble of conformers, so called *conformational selection*.

Prediction of disorder

Based on the specific biochemical properties and biased amino acid composition of IDPs, many different algorithms have been developed in order to predict protein disorder. More than 50 predictors have been established at this stage [30]. *Ab initio* models are one class of predictors that use features extracted from the primary sequence together with statistical models to predict disorder. Examples of *ab initio* predictors are IUPred [31], PONDRs [32], DisEMBL [33] and RONN [34]. Clustering methods, instead, predict tertiary structure models for the target protein, superimpose them and calculate the probability for each residue being disordered. DISOclust [35] is one example of this kind of algorithm. Meta methods, like metaPrDOS [36], combine the output of disorder predictors, which can increase accuracy slightly [37]. The CH-plot (Figure 3A) is also a popular predictor of protein disorder. The mean net charge and the mean hydropathy are calculated for the protein target and the values are plotted in a 2D graph where ordered and disordered proteins are separated into different regions. The theory behind this is that IDPs are enriched with charged amino acids, since these lead to strong electrostatic repulsion, and depleted of hydrophobic amino acids, since this lead to a diminished driving force for compaction [21].

Functions

IDPs have, so far, been found extensively in regulation of key cellular processes such as transcription, translation, signal transduction and cell cycle. Based on their mode of action, IDPs can be classified into five broad functional groups; entropic chains, effectors, scavengers, assemblers and display sites. The *entropic chain* class includes functions when disordered proteins can use their intrinsic flexibility and conformational freedom to search for binding partners (linkers) and separate binding motifs (spacers) [10]. An example is the cytosolic tail of the voltage-gated potassium channel of nerve axons, an IDPs that works as an entropic clock. The cytosolic tail searches in space for its cognate site (body of the channel). When found, the tail sterically occludes ion transport and thereby inactivates the channel [38]. *Effector* IDPs bind to a partner and either inhibit (most common) or activate the partner protein. For example, $p27^{Kip1}$, which is one of the best-characterized IDPs, inhibits Cdk2 by binding to the Cyclin A-Cdk2 complex. The *scavenger* IDP function enables the storage and/or neutralization of small compounds. Casein, for instance, binds calcium phosphate and thereby prevents it from forming precipitate in milk. IDPs can also function by organizing the assembly of multiple proteins into a complex (*assembler*). A very large number of possible connections through binding are allowed via the intrinsic adaptability behavior of IDPs, making them function as so-called hub proteins. Related to the ability of IDPs to

bind multiple partners is the function of molecular chaperones. Chaperones are proteins that bind to a large variety of different folding intermediates to prevent aggregation and facilitate folding of proteins or RNA molecules. Bioinformatics analysis have shown that as much as 54% of the residues of RNA chaperones fall into disordered regions and 40% have long stretches of more than 30 consecutive disordered residues. The corresponding numbers for protein chaperones are 37% and 14% [39]. This means that disorder is even more common in chaperone functions compared to regulatory and signaling processes that normally are considered to be the most disordered function class [40]. Post-translational modifications of proteins are important regulators in the cell. The last class of IDP functions, *display sites*, mediates these types of modifications, since disorder is often required for the modifying enzyme to bind in a specific and transient way.

Disorder and phosphorylation

Protein phosphorylation represents a major regulatory mechanism in eukaryotic cells and as much as one third of all eukaryotic proteins are estimated to be reversibly modified by phosphorylation [10]. Phosphorylation often regulates the activity of proteins involved in signal transduction, key processes in the cell where IDPs have been found to be over-represented. Computational studies using predictions for phosphorylation sites strongly support the fact that disordered regions are enriched in phosphorylation sites [41]. The effect of the modification can, however, be different from system to system. For example, phosphorylation of the pKID domain of the transcription factor cyclin-AMP-response-element-binding (CREB) protein enhances the binding to the KID-binding domain of the CREB-binding protein by providing electrostatic interactions [42]. Another effect of phosphorylation has been observed for the 4E binding protein 1 (4E-BP1) and one of the eukaryotic translation initiation factors (eIF4E). In the non-phosphorylated state, 4E-BP1 binds to the initiation factor by forming an α-helix. Upon phosphorylation of a serine close to the binding site, the helical conformation and thereby the interaction is disrupted and translation can be initiated [43]. This type of regulation is likely relevant for a number of protein-protein interactions modulated by phosphorylation, particularly since helix formation in general is strongly affected by phosphorylation [44].

Disorder and binding

Disordered targets often undergo coupled folding/ordering and binding, also termed disorder-to-order transition [45]. Ordering can occur for the entire protein or small or large segments of the sequence [46]. Short segments involved in binding of IDPs are often referred to Linear Motifs, which in principle means short consensus sequences that are recognized as modification sites for kinases, for instance. The motifs can assume either α-helical structure, β-sheet, irregular structure or even stay in an extended conformation [47]. Binding regions in IDPs that already in the unbound state have a propensity to form structured elements are called preformed elements or molecular recognition features (MoRFs). MoRFs can be categorized in different subgroups depending on the secondary structure they adopt upon binding; α-MoRFs, β-MoRFs and i-MoRFs that adopt α-helical, β-sheet and irregular secondary structures, respectively [48]. Folding of the IDP can occur either before or after binding to the partner molecule [49]. Two extreme mechanisms are proposed for these different folding events; induced folding and conformational selection (Figure 3B). In the first mechanism, the IDP associates to its binding partner in a fully disordered state and then subsequently folds [50]. Examples of induced folding are the largely unstructured pKID that folds after binding to the surface of KIX [51], and binding of the disordered CBD domain of WASP (Wiskott-Aldrich syndrome protein) to Cdc42 [52]. The extent of the structural transition can be somewhat different but nevertheless binding leads to a more ordered state of the IDP [53]. In the conformational selection mechanism, the binding protein instead selects a member from the ensemble of conformers that provides the best complementary to its own structure. One example is the NCBD domain of transcription coactivator CBP and the p160 steroid receptor coactivator ACTR. Interaction of NCBD is mainly initiated by a pre-existing folded conformation [54].

An important notion is that many IDPs can fold into different structures upon binding to different partner molecules [50]. For instance, the regulatory disordered domain of p53 can fold into both helical, β-strand and extended irregular structure upon binding to 4 different partners [55]. The ability of a protein to fulfill more than one function is called moonlighting. It is not just the interesting observation that an IDP can adopt different structures and thereby interact with different partners that is covered in the moonlighting mechanism. For instance, an IDP can form different conformations and bind the same protein at different sites, with the consequence of altered functions. They can also bind a partner in one mode but undergo a conformational change or reorganization in the bound state, that leads to a change in the functional effect [56].

Many observations suggest that the dominant mode for IDP function is by binding and subsequently folding to a target molecule. The term

coupled folding and binding or disorder-to-order transition can be a bit misleading, however. Even in the bound form, IDPs hardly ever become fully ordered. Instead, the concept of fuzziness was coined to describe the continuous spectrum of disorder possible in protein complexes, ranging from static (where the IDP adopt a few well-defined conformations) to highly dynamic states (where either the entire IDP or parts of the protein can remain disordered in the bound state) [57]. The transition of an IDP from the disordered free state to the more ordered bound state is accompanied by a large decrease in conformational entropy. This means that the driving force for the binding interaction is enthalpy, and most often coupled folding and binding give rise to complexes with high specificity and relatively low affinities. This kind of interaction is often desirable in regulation and signaling, where IDPs are abundant. High specificity is needed to initiate signaling and low affinity for the importance of dissociation when signaling is complete [58].

The immune system

In humans the first line of defense against pathogens is the physical and chemical barriers of e.g. the skin, the lung epithelial mucus and the acidic environment of the stomach. If those barriers fail to defend us from invaders, cells of the first category of the immune response, called the nonspecific innate immunity (Figure 4), are activated. These cells respond in the same way to all kinds of foreign material and are therefore considered to be nonspecific [59]. Phagocytic cells like the macrophages, dendritic cells and neutrophils are cells that belong to the innate immunity. These cells recognize and ingest foreign material [60]. The complement system is a biochemical cascade of the innate immunity that helps or "complement" the ability of phagocytic cells and antibodies to lyse the invading pathogens [61]. Natural killer cells (NK cells) are an early component of the nonspecific host response to virus infection [59]. They also have the ability to kill certain lymphoid tumor cell lines [62]. Basophils are cells that release chemicals that mediate inflammation and allergic responses and eosinophils are important in destroying large antibody-coated parasites. These cells also belong to the innate immunity.

The acquired (adaptive) immune system (Figure 4) is composed of highly systematic and specific cells called lymphocytes. Three main types of lymphocytes exist: T lymphocytes, B lymphocytes and NK T cells. These cells recognize and bind full or digested parts of pathogens, called antigens. Once a lymphocyte has encountered a matching antigen, it will divide and create a massive amount of that particular lymphocyte. A response to the antigen is initiated which includes antibody production, cell killing (cytotoxicity), production of cytokines and creation of immunological memory. Besides being important players of the innate immunity, NK cells also play a role in the adaptive response (called NK T cells) by recognizing glycolipid antigens [59]. A minor subset of the T lymphocytes, called γδ T cells, express only a very limited diversity of receptors. Since their receptors are relatively invariant and because they occur only in specific locations within the body, they are known as innate-like lymphocytes. The classification of these cells are at the borderline between the innate and acquired immune system [63]. The function of B- and T lymphocytes will be discussed in the following sections.

Autoimmune diseases arise when the body is attacking its own cells as if they were pathogens invading the body. The opposite situation is also possible, where the immune system's ability to deal with infections can be entirely absent or compromised, which in turn can lead to so-called immunodeficiency.

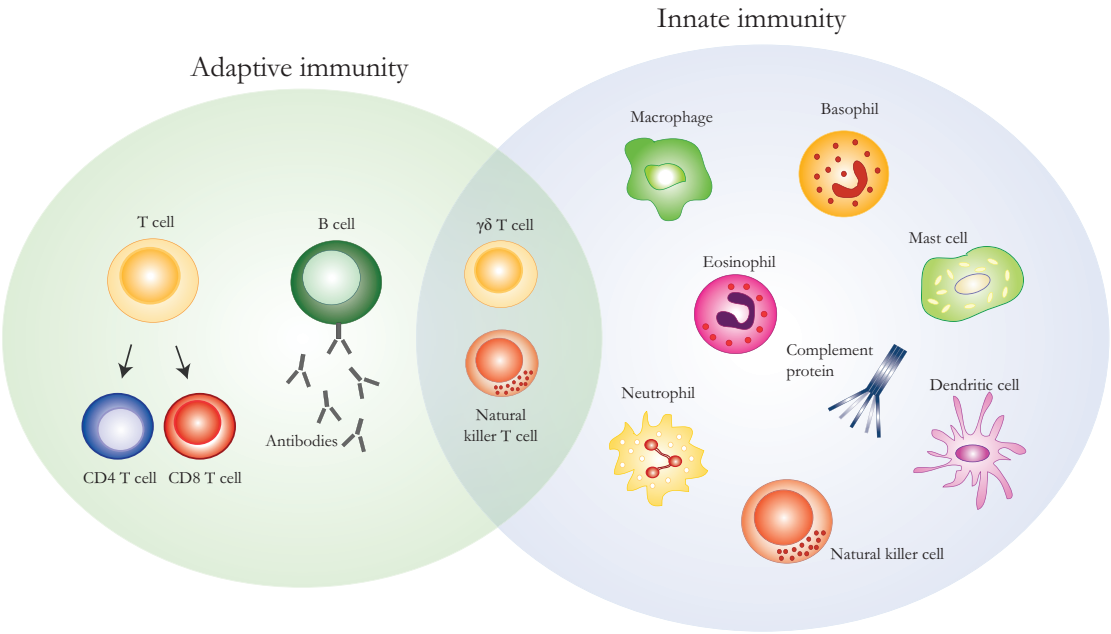


Figure 4. Cartoon of the adaptive and innate immunity. The adaptive immunity, creating a long-lasting memory of a specific pathogen, consists of B- and T cells. T cells fall into two major classes: CD4 helper T cells and CD8 cytotoxic T cells. The non-antigen specific immune cells that respond to foreign material belong to the innate immunity. Cells included in this system are: mast cells, dendritic cells, neutrophils, macrophages, complement proteins, basophils, eosinophils and natural killer cells. On the borderline between innate and adaptive immunity we can find γδ T cells and natural killer T cells.

B cell and B cell receptor

B cells (B lymphocytes) are central components of the adaptive immunity, responding to the myriad pathogens in our environment by producing antibodies, performing the role of antigen-presenting cells, secreting cytokines, and developing into memory B cells after activation [64]. B cells circulate in the blood and lymphatic systems. In the lymphoid organs they encounter its cognate antigen, and together with an additional signal from a T helper cell, it can differentiate into effector plasma cells. These cells secrete specific antibodies that will circulate in the blood to target and eliminate antigens or pathogens [65]. Antibodies have multiple functions, like coating antigens and bacterial toxins so that phagocytic cells can ingest and destroy them. They also activate the complement system and other immune cells [59]. To detect the antigen or pathogen, B cells have up to 120 000 B cell receptors (BCRs) on the cell surface [66]. The BCR is a multicomponent receptor composed of a transmembrane immunoglobulin molecule (mIg) and a disulfide linked heterodimer of CD79a (Igα) and CD79b (Igβ) (Figure 6A). The B cell receptor consists of two heavy (H) and two light (L) chains joined by disulfide bonds. Each chain has a constant (C) and a variable (V) region. How these receptors with a finite number of

genes can have an almost infinite range of specificities can be explained by something called the V-(D)-J rearrangement (also called somatic recombination) [67]. During B cell development in the bone marrow, the gene segments coding for the H and L chains are rearranged forming a complete variable region. The V region of the heavy chain is encoded by three different gene segments (V, D and J) while for the V region of the light chain, two gene segments (V and J) are forming the encoding regions (Figure 5). Since there are multiple copies of all gene segments, random selection from the different sections creates a great diversity of variable regions, producing different immunoglobulins so that each BCR is specific for a single antigen [63].

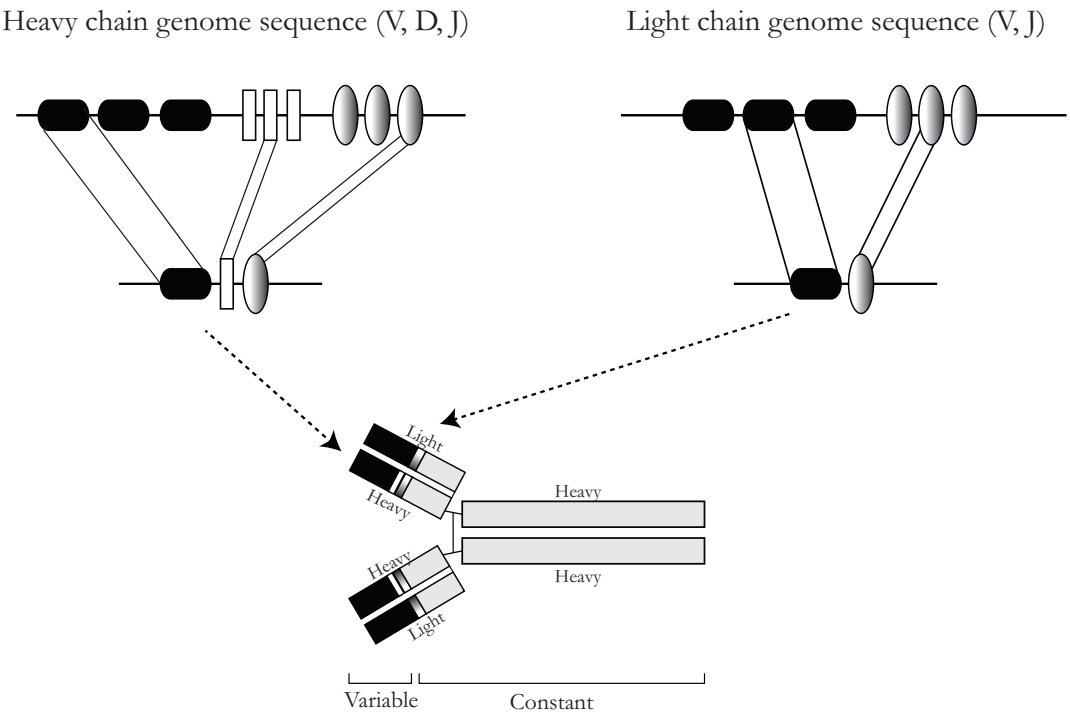


Figure 5. Simplistic overview of V-(D)-J recombination of heavy and light chain. Random selection of the different encoding genes creates a great variety of immunoglobulins.

B cell receptor signaling

The cytosolic intrinsically disordered tails of CD79a and CD79b both contain an immunoreceptor tyrosine-based activation motif (ITAM) with a consensus sequence of $D/EX_7D/EX_2YX_2L/IX_7YX_2L/I$ (Figure 6A). These two tyrosine residues are phosphorylated upon antigen engagement [66]. Multiple intracellular protein tyrosine kinases are responsible for the signaling cascade upon BCR activation (Figure 6C). The first kinase is an Src family kinase, predominantly Lyn, which phosphorylates the tyrosines within the ITAMs for both CD79a and CD79b. Phosphorylation of the ITAMs leads to binding of Syk, an Src-homology 2 (SH2) kinase, which gets activated and phosphorylated upon binding. Once Syk is bound, the BCR signal is propagated via enrollment of a group of proteins associated to the adaptor protein B cell linker (Blnk) that bind to CD79a and becomes phosphorylated by Syk. Blnk serves as a scaffold for the assembly of other components like Bruton's tyrosine kinase (Btk) and phospholipase C-gamma 2 (PLCγ2). PLCγ2 cleaves the phosphoinositide $PI(4,5)P_2$ generating IP_3 and DAG, eventually leading to release of the intracellular storage of Ca^{2+} ions and activation of NFκB, one of the key transcription factors for B cell activation. BCR signaling quickly becomes quite branched and complex [64]. When the stimuli is too weak for the receptor to be activated, a co-receptor called CD19 (in conjunction with the proteins CD21 and CD81) can amplify and activate the Lyn kinase, enabling low-avidity stimuli to start the intracellular signaling cascade.

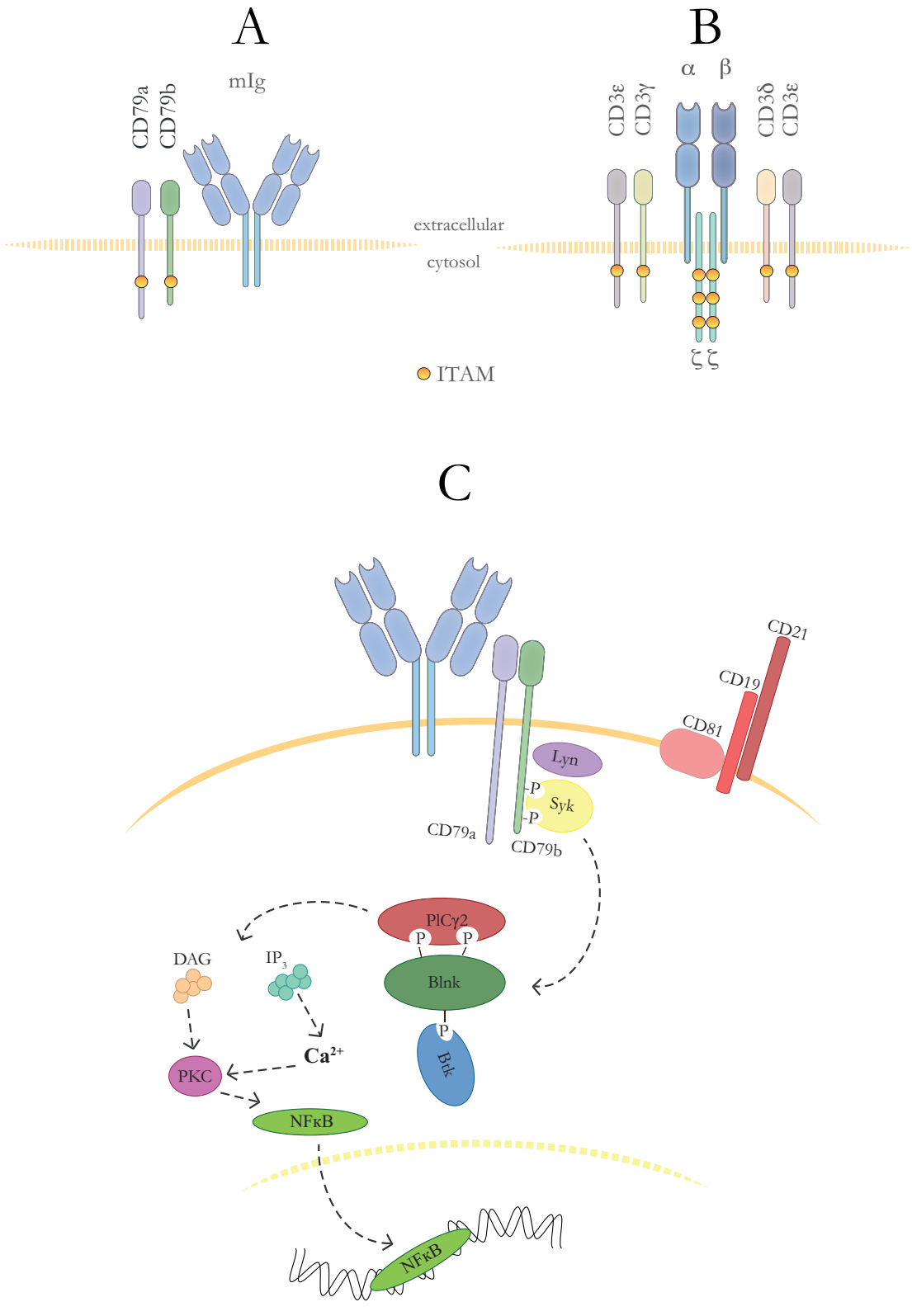


Figure 6. Cartoon of the B cell receptor **(A)** and T cell receptor **(B)** with indicated immunoreceptor tyrosine-based activation motifs (ITAMs). **(C)** Simplistic overview of B cell receptor signaling, showing the main responsible kinases and assembly proteins as well as the co-receptors CD21, CD19 and CD81.

Initiation of B cell receptor signaling

The transmembrane signal transduction mechanism for the B cell receptor is not fully understood, but several different models have been proposed, e.g. the dissociation activation model (DAM) [68], the conformation-induced oligomerization model [69], and the signaling chain homooligomerization (SCHOOL) model (Figure 7) [70]. The latter model proposes that upon ligand-induced clustering and reorientation of BCRs, oligomeric intermediates are formed. In these intermediates, homointeraction between CD79a-CD79a and CD79b-CD79b can take place forming competent signaling oligomers. In these oligomers, tyrosine kinases can phosphorylate the ITAMs, starting the signaling cascade. The model suggests that it is the actual forming of homooligomers that is the key to trigger the downstream signaling instead of the receptor clustering *per se* [70]. The first theory is the dissociation activation model. According to the DAM, the BCRs exist on resting B cells as auto-inhibitory oligomers and in these oligomers the ITAMs are hidden from tyrosine kinases. Antigen binding disrupts these closed oligomers, forming open, signaling active monomers [68].

Soluble monovalent antigens are not able to trigger BCRs [71]. The binding of multivalent ligands have therefore been the key to understand activation and to create models for how this activation is performed, as seen in figure 7A and B for the SCHOOL and DAM models. Pierce and colleagues, however, have shown that monovalent antigens presented on membranes *can* activate BCRs. A model called conformation-induced oligomerization was therefore proposed. In this model (Figure 7C) it is suggested that in the absence of antigen, the receptors are in a conformation that is not receptive to oligomerization, and upon binding of monovalent antigens, presented on membranes, a conformational change of the receptors takes place, bringing the receptors together in an oligomerization-receptive form. BCR monomers are freely diffusing in the membrane, and it is not until the antigen-bound BCR encounter another BCR that oligomerization can take place. In this way, microclusters are formed and initiation of signaling is achieved. This theory explains how monovalent antigens can activate the receptors. For the activation by multivalent antigens this conformational change of the receptors is not taking place, instead it is a physical crosslinking of receptors that will mimic oligomerized BCRs at later points during the process of signal initiation [72].

How oligomerization of the BCRs results in a change of the cytoplasmic CD79a and CD79b making them prone for phosphorylation and how this can recruit the Lyn kinase is still not understood. Studies have shown that the local lipid environment of the antigen-free versus the antigen-bound BCR cluster is different [69]. The receptor clusters seem to be associated with detergent-insoluble lipid-rafts, rich in sphingolipid and cholesterol. Since the Lyn kinase is associated to lipid-rafts, it can come into close proximity with the ITAMs of CD79a and CD79b for phosphorylation.

The change in local lipid environment can also cause alterations of the cytoplasmic tails of CD79a and CD79b, providing a docking site for Lyn kinase. Wuchterpfennig and colleagues have shown that for the ITAM-containing cytosolic tail of CD3ε, belonging to the T cell receptor, membrane interaction hiding the tyrosines in the ITAM takes place. How the tail can be accessible upon antigen binding is still unknown but it is possible that perturbations of the membrane lipid composition can be the trigger [73]. To fully understand the molecular basis of the antigen-induced initiation of BCR signaling, it would be essential to study the signaling responsible domains of CD79a and CD79b at a molecular level. The objective would be to understand how these units contribute to the active versus non-active state of the receptor.

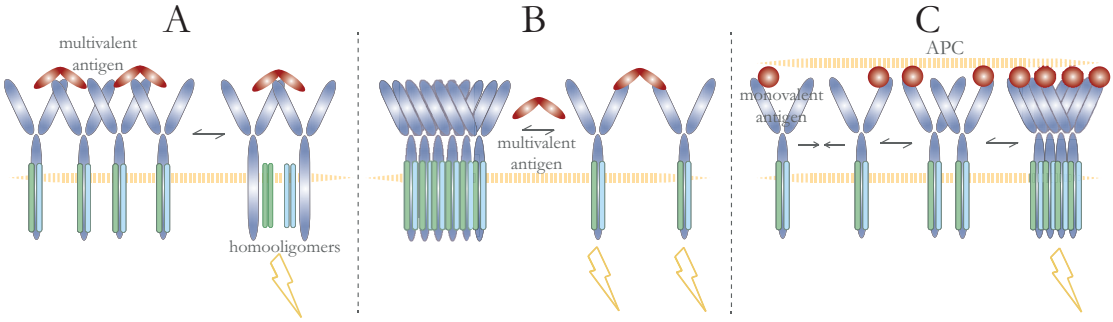


Figure 7. Schematic representation of the three major theories of the signal transduction mechanism for the B cell receptor. **(A)** The SCHOOL model where antigen binding clusters the receptors and homooligomers of CD79a and CD79b are formed. These homooligomers are considered to be the signaling active components. **(B)** Dissociation activation model, where B cell receptors exist as auto-inhibitory oligomers on resting B cells, and upon antigen binding they dissociate from each other forming signaling open monomers. **(E)** Conformation-induced oligomerization model, explaining how monovalent antigens bound to antigen-presenting cells (APC) bind and cause a conformational change of the receptors, bringing them in an oligomerization-receptive form, which initiate signaling.

B cell receptor and diseases

BCR signaling is biologically important for normal B cell development, activation, trafficking and differentiation. Dysregulation of the signaling pathways can cause human diseases. Pathologies coupled to deviating BCR signaling are: hyper-IgM syndrome, common variable immune deficiency (CVID), X-linked agammaglobulemia (XLA), autoimmune diseases such as rheumatoid arthritis (RA), systemic lupus erythamatosis (SLE), idiopathic thrombocytopenia purpura (ITP), asthma, diabetes as well as leukemia and lymphoma [65]. For instance, mutations affecting the tyrosine residues in the ITAMs of CD79a and CD79b are shown to make BCR signaling chronically active, leading to cancer in B cells (DLBCL) [74]. Mutations in CD79a can also result in B cell deficiency with few or no antibodies

produced. These alterations may promote the metastasis of gastric cancer, and therefore CD79a has been considered to be a potential therapeutic target for this type of cancer [75].

T cell and T cell receptor

There are two major groups of T cells (T lymphocytes), helper and cytotoxic T cells. The latter have the most direct effect by destroying virally infected cells and tumor cells. They require that the antigen is presented by a major-histocompatability complex (MHC) molecule (more specifically MHC class I). Since the cytotoxic T cells have the cell-surface molecule CD8 expressed on the membrane they are also called CD8 T cells. Helper T cells are instead called CD4 T cells because they express a CD4 glycoprotein on the cell surface. Helper T cells bind to antigen presenting cells displaying an antigen bound to a MHC class II molecule. These types of cells are very important in fighting bacterial infections and in activation of macrophages and B cells by the secretion of cytokines [63]. Activation of T cells is initiated by binding of the T cell receptor, located on the surface of T cells, to antigens presented by an MHC molecule. The T cell receptor consists of a variable TCRαβ heterodimer that binds to the antigens. TCRαβ forms a complex with non-variable CD3 complex, the CD3ε-CD3γ, CD3ε-Cd3δ and TCRξ-TCRξ dimers (Figure 6B). These cytosolic CD3 subunits are intrinsically disordered and the signal transduction responsible parts of the receptors, also containing ITAM motifs, similar to the CD79a and CD79b subunits of the B cell receptor (Figure 8) [76].

Figure 8. The amino acid sequence of cytosolic constructs used in this work. The consensus sequence of the ITAM motifs is underlined, and tyrosine residues, responsible for signaling upon phosphorylation, are shown in bold.

T cell receptor signaling

The first phosphorylation event of the ITAMs is performed by the Src family tyrosine kinase Lck and/or Fyn. A dually phosphorylated ITAM recruits the ZAP-70 tyrosine kinase, which in turn phosphorylates downstream components of the signaling pathway [73]. Like for the B cell receptor, there is considerable controversy about the mechanism of how the binding event of peptide-MHC to the TCR is transduced across the plasma membrane. Proposed triggering models range from receptor aggregation, receptor segregation and conformational changes [77].

A piston-like displacement of the TCR-CD3 complex, resulting in a change in the conformation of the CD3 cytoplasmic units making them prone for phosphorylation, is one idea behind the conformational change model [77]. Like mentioned earlier, Wucherpfennig and colleagues solved the structure of CD3ε bound to negatively charged bicelles, showing that tyrosine residues in the ITAM are buried in the membrane and that phosphorylation by Lck is therefore prevented. Wucherpfennig proposed different models for how CD3ε is released from membrane upon peptide-MHC binding. For instance, a mechanical force can lead to dissociation of the ITAM from the membrane. It can also be that clustering of TCR-CD3 complexes result in competition among the cytoplasmic units for membrane surface or that clustering of receptors can change the lipid environment in the vicinity of the clustered TCRs, reducing the affinity of CD3ε to the membrane [78].

For the aggregation model it is simply the aggregation of TCR-CD3 complexes, either with or without help of co-receptors, that leads to an increase of proximity of associated Lck kinase and thereby phosphorylation initiation [77].

The segregation model proposes that TCR binding to peptide-MHC complex traps the TCR-CD3 complex into certain zones that are depleted of phosphatases or into lipid-raft regions [79], which are enriched in Lck kinase and deficient in phosphatases [77].

Methodology

Protein production

A prerequisite for all structural and biochemical/biophysical work is to have the protein of interest in amounts necessary for the method of choice. Characterization using NMR spectroscopy usually also requires isotopic labeling of the amino acids. Both bacterial and eukaryotic expression systems are available for the traditional *in vivo* strategy of protein production. They differ primarily in their efficiency of translation and how the protein is being processed after the synthesis [80]. A DNA vector containing the gene of interest is transformed into the cell system of choice. It is common procedure to also include purification tags, protease digestion sites and sometimes signal sequences in the same vector. After transformation, the cells are cultured so that they can transcribe and translate the desired protein. Typically the cells are then lysed to extract the expressed protein for further purification.

Cell-free protein expression

In the early 1950s, it was demonstrated that protein synthesis was continued in cell extracts and disintegrated cells. When it also was demonstrated that disrupted cells were able to produce proteins and ribosomes were identified as the particles responsible for protein synthesis, the first cell-free expression systems were invented [81]. These systems, however, translated endogenous mRNA. A breakthrough in the field occurred in the 1960s when the first *in vitro* protein synthesis from exogenous mRNA template using *E. coli* cell extracts was developed [82, 83]. For this system, the endogenous mRNA was removed without damaging the ribosomes by simple a pre-incubation of the extract at 30-37°C.

In prokaryotes, the ribosomal translation of mRNA is initiated directly after synthesis of the mRNA from the DNA template. The process is therefore called coupled transcription-translation. Adding pre-existing mRNA can, however, hinder the initiation of translation by mRNA folding, especially if structure is formed at the ribosome-binding sites [81]. In 1967, a system based on a crude *E. coli* extract with an efficient bacterial coupled transcription-translation system for exogenous DNA was developed [84, 85].

The next major step in the development of cell-free expression was the use of a specific phage RNA polymerase, orthogonal to the host from which the cell extract was derived. T7 polymerase or SP6 polymerase were

the most successful suggestions, directing synthesis of the specific proteins whose genes are preceded by the corresponding phage promotor. Several advantages resulted from this: higher transcription rate of the phage RNA polymerase compared to the endogenous polymerase, only the gene of interest was expressed and the combination of phage RNA polymerase and DNA template could be combined with both prokaryotic and eukaryotic cell extracts [81].

Until the late 1980s, the cell-free expression reactions were only performed in a fixed volume of a test-tube (batch mode). Synthesis could be maintained for at most an hour, which resulted in low yields of protein, useful mainly for analytical purposes. Other systems were therefore developed that used a continuous supply of the consumable substrates and with removal of reaction products that inhibited the translation at a certain build-up concentration [86, 87]. In the established continuous-flow cell-free (CFCF) system, the feeding solution is continuously pumped into the reaction chamber and products are removed through an ultrafiltration membrane by the outgoing flow. Continuous-exchange cell-free (CECF) systems instead use a dialysis membrane to separate the feeding solution from the reaction solution. The use of these continuous-action systems for cell-free expression of proteins prolonged the translation process up to many hours or even days. The yields of the product therefore increased [81].

Principles of cell-free protein synthesis (CFPS)

CFPS is an *in vitro* method for producing proteins, generally constructed with cell extract prepared from *E. coli*, wheat germ, or rabbit reticulocytes [88]. In order to improve protein expression, much research has been conducted to identify factors affecting *in vitro* transcription and translation. Optimization of cell-free expression conditions by adjusting composition of the lysate, preparing extracts from genetically engineered cells, choosing an efficient energy source and template DNA are among factors that have been intensively studied. In addition, various expression modes have been explored, such as continuous-flow, dialysis, batch and bilayer systems. Consequently, the amount of protein that now can be produced reaches levels of mg/ml in many cell-free systems [89]. In addition to the cell extracts, *in vitro* reconstituted mixtures containing all necessary components for transcription and translation have been made (Protein synthesis Using Recombinant Elements, PURE), which provides higher reaction controllability [90].

The extracts contain all soluble cell components needed for transcription and translation (ribosomes, translation factors, aminoacyl-tRNA synthetases etc.), but are devoid of macromolecules such as cell-surface membranes and genomic DNA. Addition of template DNA or PCR fragments coding for the protein of interest together with RNA polymerase

and components like an energy source, amino acids and tRNAs starts the synthesis of proteins [90].

In vitro protein synthesis is an attractive alternative to the generally laborious *in vivo* protein expression systems (Figure 10B). Not only does it provide simple and quick means, where high yields of protein can be produced within two hours compared to several days using *in vivo* systems [3]. Due to the open nature of the system, it also enables the ability to change and screen different expression conditions (e.g. pH, redox potential, temperature, chaperones) [88]. Since there is no need to maintain cell viability, toxic proteins can be produced. Also, protein with non-natural amino acids and chemical groups integrated at defined positions can be produced at high levels [89]. For NMR spectroscopy applications, there are several advantages using CFPS in order to produced isotopically labeled proteins. Scrambling of the labels due to metabolic pathways in the cell-free system is suppressed compared to *in vivo* expression, and by just simply replace the amino acid(s) of interest with the labeled one(s) the protein can be isotopically labeled and suitable for NMR studies [91]. For large systems, deuteration is necessary to improve NMR signal linewidths. Using conventional *in vivo* expression, the growth must be performed in heavy water, which increases cultivation time and invariably leads to lower yields. Back-exchange of protons at peptide amides might also be a problem. In CFPS the protein is produced in regular water with deuterated amino acids, the result being no change in expression level [88].

The transcription and translation processes require a lot of energy. A problem using bacterial extracts is that they have high endogenous phosphatase and ATPase activities. So in order to enable high yields of protein, an ATP regeneration system is needed. At the moment, energy supply is the main limiting factor for CFPS and many different energy supply systems have been developed. Some examples are the energy substrates of phosphoenol pyruvate, acetyl phosphate and creatine phosphate that are added together with their corresponding kinases (Figure 9). For the batch configuration it is especially important with an effective regeneration system because accumulation of phosphate has been observed to have an inhibitory effect for translation. Modified energy systems for batch setups use oxidation of substrates from glycolytic pathways (e.g. pyruvate, glucose, glucose-6-phosphate) [92]. By combining phosphoenol pyruvate as the conventional energy source together with nicotinamide adenine dinucleotide and coenzyme A, a more effective regeneration system has been developed (PANOx system) [93]. The Cytomim system [94] use potassium and magnesium glutamate together with polyamines spermidine and putrescine in order to mimic the cytoplasmic milieu of the cell and to use oxidative phosphorylation, which is the most effective natural source of ATP, as an energy source. This powerful option is however not suitable for

NMR applications because of the possibility for isotopic scrambling due to the large amounts of glutamate in the buffer [88].

The detailed knowledge of the protein-synthesizing machinery of *E. coli* cells has enabled remarkable improvements of the bacterial cell-free expression systems [81]. In general, *E coli*-based systems provide higher yields and more homogenous samples, which is good for structural analysis. Eukaryotic systems, however, are needed for producing post-translational modified proteins [95]. Synthesis in *E. coli* extracts has also been shown to be tolerant against many different additives. Today there are a multitude of *E. coli* strains that can be used for producing extract. These strains have been extensively modified to improve the performance of the cell-free reaction. For instance there are strains that are depleted in cytoplasmic proteases, strains that carries plasmid to express rare tRNAs, strains that are depleted in RNase activity and strains that enhance the formation of disulfide bonds [81].

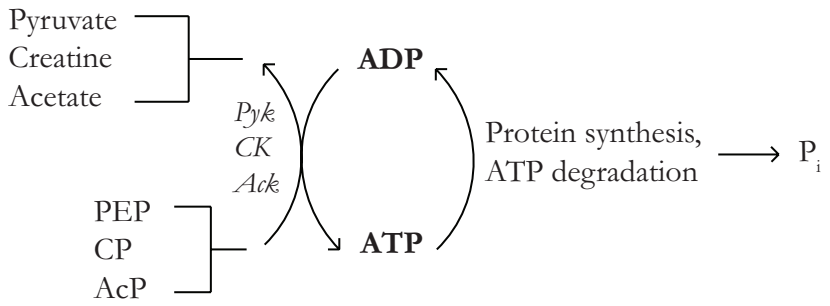


Figure 9. Schematic representation of different methods for energizing cell-free protein synthesis. Phosphoenolpyruvate (PEP) can be converted to pyruvate by the pyruvate kinase (Pyk), converting ADP to ATP, needed for protein synthesis. Creatine phosphate (CP) can also be used in conjunction with creatine kinase (CK) as well as acetyl phosphate (AcP) together with acetate kinase (Ack). Degradation of ATP accumulates inorganic phosphate (P_i).

In-house developed cell-free protein synthesis system

For papers I-V, an in-house developed cell-free protein expression system has been used (Figure 10A) [88]. This system is based on the batch configuration and an *E. coli* extract. Even though batch systems are limited by their short lifetimes and consequently low yields, they represent a more viable option for the parallel and rapid expression of proteins, since miniaturization is possible and setup and handling is uncomplicated [96]. Isotopically labeled amino acids are the most expensive component of the cell-free system, and as much as 72% of the cost of our system are the $^{13}C/^{15}N$-labeled amino acids (1 mg/ml) [97]. The batch configuration is therefore more cost-effective compared to the CFCF and CECF systems with their continuous supply of substrates.

The vector used for the cell-free system in papers I-V contains a T7 promotor, an N-terminal his-tag for detection and purification, Tobacco Etch Virus (TEV) protease recognition site for cleavage of the his-tag, a ribosome binding site (RBS) located at an optimal distance (6-8 bp) from the ATG start codon, a pUC origin for high-copy replication and an ampicillin resistance gene for selection purposes.

As mentioned above, the energy regeneration system is an important factor for maintaining transcription and translation in the cell-free system as long as possible. The system used in papers I-V is based on ATP generation from creatine phosphate (CP) by creatine kinase (CK). Unlike phosphoenolpyruvate (PEP), which can be consumed through a number of metabolic pathways in the cell extract, CP is not a natural substrate of the *E. coli* metabolism. A higher efficiency for ATP regeneration is therefore accomplished using the CP/CK system instead of PEP and the corresponding pyruvate kinase [98].

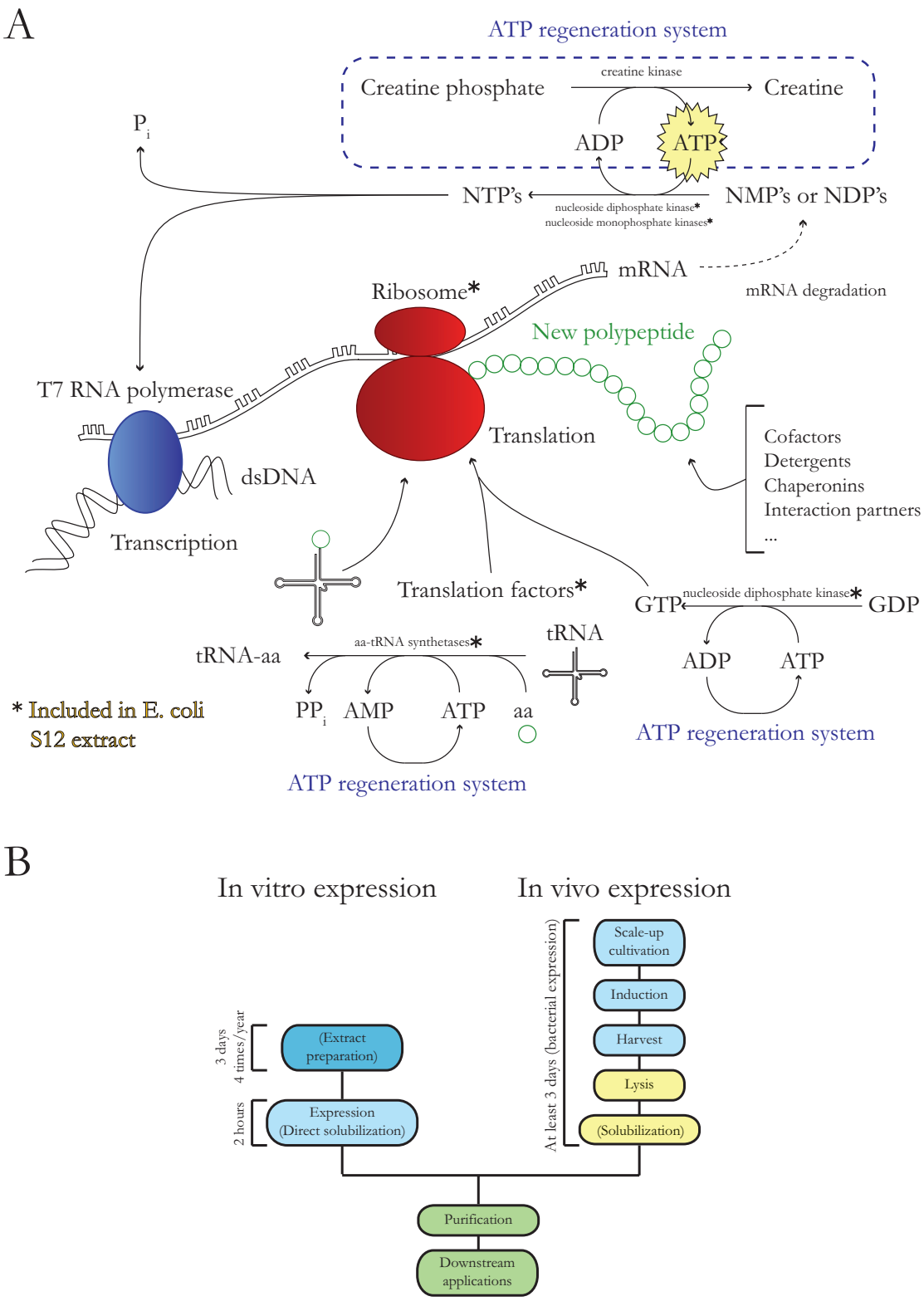


Figure 10. (A) Schematic representation of the CFPS system. **(B)** *In vitro* vs. *in vivo* protein production demonstrating the time efficiency using *in vitro* and the reduced number of steps for setting up protein production. Printed with permission from Dr. Anders Pedersen.

Cell-free protein synthesis and membrane proteins

The inherent possibilities of changing conditions that has a direct impact on the produced polypeptide makes CFPS the most versatile expression system available for membrane protein (MP) synthesis. This group of protein can be extremely difficult to express in a functional state. Often the integrity of cellular membranes, resulting in growth retardation or maybe even lysis of the cell, can be affected upon membrane integration of the produced MPs in conventional *in vivo* systems [99]. Over-expression of MPs *in vivo* frequently also results in cell toxicity mostly due to protein aggregation, misfolding and the inherent hydrophobicity of MPs. In CFPS, co-translational solubilization of the MPs is possible and a wide range of detergents, lipids and nanodiscs are compatible with the synthesis machinery [3]. Membrane-mimicking systems present at the beginning of the reaction gives the nascent polypeptide the opportunity to become solubilized during or shortly after translation, which increases the likelihood for proper folding of MPs [99]. Co-solubilization together with the fact that cell viability does not need to be maintained makes *in vitro* protein expression a unique opportunity for MP production [95]. In paper I, it is demonstrated that CFPS can be successfully applied to express preparative amounts of MPs of varying size and complexities using different solubilization agents (Table 1).

Table 1. A summary of expression yields and solubilization experiments for the different MPs included in this work.

Protein	Abbrev.	UniProt acc. no.	Size (kDa)	TM segments	Expression yield	Tested cotranslational solubilization
Human aquaporin 0	hAQP0	P30301	28.1	6	+++	Brij-58
Human aquaporin 1	hAQP1	P29972	28.5	6	+	Brij-58
Human aquaporin 2	hAQP2	P41181	28.8	6	+	Brij-58
Human aquaporin 3	hAQP3	Q92482	31.5	6	++	Brij-58
Human aquaporin 5	hAQP5	P55064	28.3	6	+++	Brij-58
Human aquaporin 6	hAQP6		29.4	6	++	Brij-58
Human aquaporin 7	hAQP7		32.0	6	+	Brij-58
Human aquaporin 8	hAQP8	O94478	27.4	6	++	Brij-58, DDM
Human aquaporin 9	hAQP9	O43315	31.4	6	++	Brij-58
Human aquaporin 10	hAQP10	Q96PS8	31.8	6	+	Brij-58
Human aquaporin 11	hAQP11	Q8NBQ7	30.2	6	+	Brij-58, CHAPS, DDM, LMPG, LPPG, TX100
Human aquaporin 12, isoform 2	hAQP12	A6NM10	32.5	6	+	Brij-58, CHAPS, DDM, LMPG, LPPG, TX100
P. pastoris aquaporin	*pp*Aqy1	–	29.9	6	+++	Brij-58
S. cerevisiae Fps1p	*sc*Fps1p	P23900	76.3	6	+++	Brij-58, DDM
E. coli EnvZ	*ec*EnvZ	P0AEJ4	50.4	2	++	Brij-58, CHAPS, DDM, LMPG, LPPG, TX100
B. subtilis ResE	*bs*ResE	P35164	66.8	2	++	Brij-58
E. coli PhoR	*ec*PhoR	P08400	49.6	2	++	Brij-58
T. thermophilus PhoR	*tt*PhoR	Q72HX1	40.7	2	++	Brij-58
B. subtilis PhoR	*bs*PhoR	P23545	65.1	2	++	Brij-58
E. coli CpxA	*ec*CpxA	P0AE82	51.6	2	+	Brij-58
E. coli ArcB	*ec*ArcB	P0AEC3	87.8	2	++	Brij-58
B. subtilis YkoH	*bs*YkoH	O34638	50.7	2	++	Brij-58
T. thermophilus 4619	*tt*4619	–	43.8	2	–	Brij-58
Human UCP1	hUCP1	P25874	32.9	6	++	Brij-58
Human UCP2	hUCP2	P55851	33.2	6	++	Brij-58
Human UCP3	hUCP3	P55916	34.2	6	++	Brij-58
Mouse UCP1	mUCP1	P12242	33.4	6	++	Brij-58
Mouse UCP2	mUCP2	P70406	33.2	6	+	Brij-58
Rat UCP2	rUCP2	P56500	33.4	6	++	Brij-58
N. pharaonis SRII	*np*SRII	P42196	25.5	7	+++	Brij-35, Brij-58, Brij-78, Brij-98, DDM, DHPC, DMPC, LPPG, TX100, nanodiscs
N. pharaonis Htr	*np*Htr	P42259	56.6	2	–	Brij-58
N. pharaonis sHtr	*np*sHtr		11.5	2	++	Brij-58
Proterhodopsin, AND4 clade	pR AND4	–	32.4	7	++	Brij-35, Brij-58, nanodiscs
Human C-X-C chemokine receptor 4	hCXCR4	P61073	42.2	7	+++	Brij-58, digitonin, CYMAL-5, TX100, DHPC
Human dopamine receptor 2, long isoform	hD2L	P14416	53.1	7	++	Brij-58, digitonin, TX100, DHPC
S. cerevisiae pheromone alpha factor receptor	*sc*Ste2	D6VTK4	36.0	7	++	Brij-58
Human GLUT1	hGLUT1	P11166	56.6	12	+	Brij-58
Human transhydrogenase	hTH	Q13423	113.9	14	+	Brij-58

Translational efficiency

Translational efficiency of a specific codon has been demonstrated to be affected by the neighboring mRNA sequence [100]. The reason for this is the physical interactions between adjacent tRNA molecules at the ribosomes when translation occurs. The interactions can either be favorable or unfavorable, influencing translational kinetics. Adjacent codons (codon pairs) therefore control the kinetics of translation without changing neither the amino acid sequence nor the protein structure. The codon pair utilization is highly species-specific and it has been observed that genes that are expressed at high levels tend to avoid codon pairs that are overrepresented (codon pairs used in protein coding sequences that occur more frequently than expected from the usage of the individual codons of these pairs), suggesting that these codon pairs lead to incompatible tRNAs. This gives a slower translation process, which can function as a kind of translational pause for the ribosomes, potentially important for folding [100]. In paper I, an apparent correlation between cell-free expression yields and numbers of potential ribosomal pause sites could be observed (Figure 11B-D), judging from so-called SpeedPlots. These plots display the relative deviation from random utilization of each successive codon pair (Z-score) for each of the 3721 nonterminating codon pairs that can occur in a gene sequence. Also, shorter fragments appeared after protein synthesis that correlated well with predicted ribosomal pause sites in respective genes (Figure 11A). Other factors, observed in literature that affect protein expression yields, are levels of mRNA secondary structure and the codon usage. The presence of infrequently used codons in a gene can significantly depress protein expression levels [101].

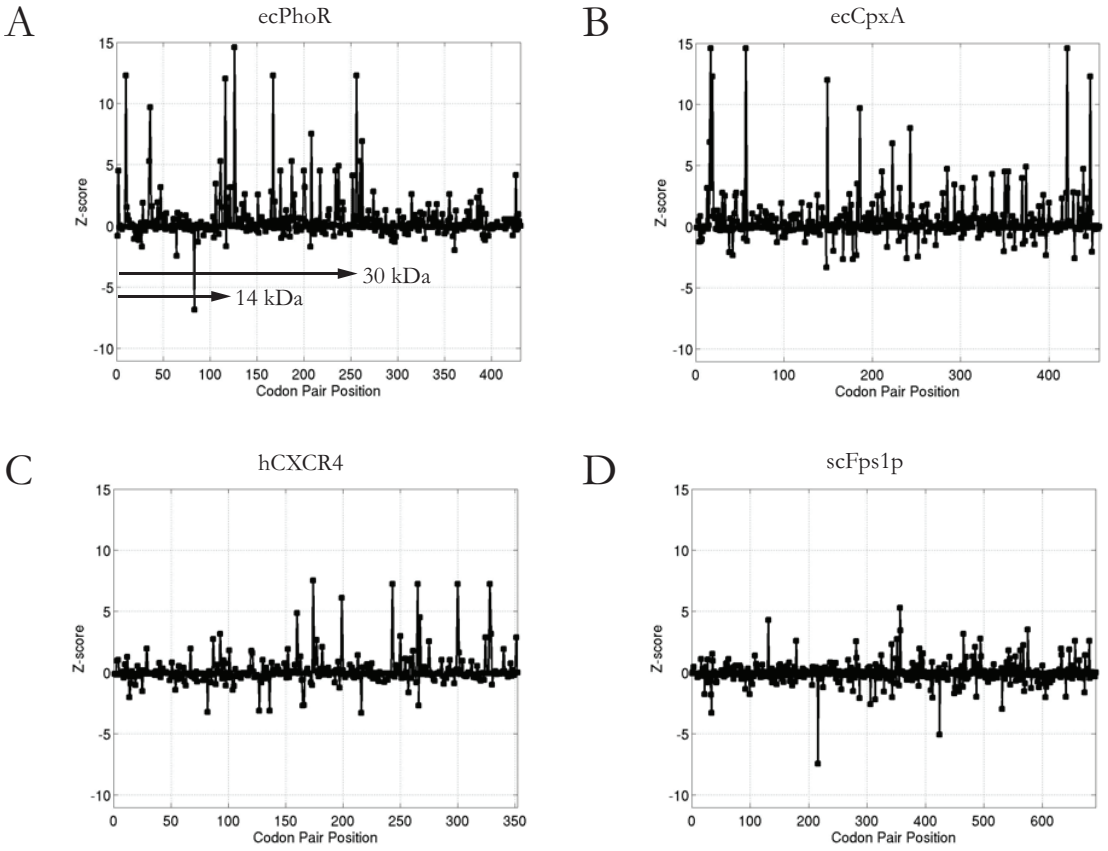


Figure 11. SpeedPlots of ecPhoR **(A)**, ecCpxA **(B)**, hCXCR4 **(C)** and scFps1 **(D)** (Figure S1 Paper I). The plots show the likelihood of a potential ribosomal pause site at a specific codon pair, distributed along the sequences. High Z-score (>10) is indicative of a ribosomal pause site. Two sites of high Z-score in ecPhoR are marked with arrows. The corresponding protein sizes, if the ribosome were to spontaneously terminate at these sites, could be observed on a Western blot.

Cell-free protein synthesis and intrinsically disordered proteins

In vivo, regions of low complexity and disorder of overexpressed heterologous proteins are often prone to proteolytic degradation, which may decrease the yields or even prohibit protein expression. By having a short expression time and by adding protease inhibitors directly in the cell-free reaction, the proteolysis can be circumvented. The lack of need for complex folding and the fact that IDPs generally do not aggregate, also work in favor for CFPS (papers II-IV) [97]. In a comprehensive bioinformatics study of 3066 human proteins expressed in an *E. coli* based cell-free expression system, several physiochemical and structural properties were shown to correlate with the success of protein expression [102]. The proportion of high yields and solubility were found to be directly correlated with the content of intrinsic disorder (where disorder was defined using the RONN

predictor). It was also observed that the higher content of charged residues, which is characteristic for IDPs, was associated with increased soluble protein expression. Glutamic acid appeared to be one of the most enriched amino acid in the subset of well-expressed proteins, which may not be so surprising since glutamic acid also has been observed to be the most abundant amino acid in IDPs [23].

Cell-free protein synthesis and amino acid labeling

Cell-free expression offers several advantages for the preparation of isotope labeled protein samples for NMR spectroscopy analysis. Since it is only the target protein that is produced and labeled, the amino acids are used very efficiently and isotopic scrambling are kept at a minimum due to lower activity of metabolic processes in the extract compared to live cells (paper V). Selective labeling can also be achieved by simply replacing the amino acid(s) of interest with the labeled one(s). In contrast, amino acid type selective labeling in bacteria is restricted to certain amino acid types or requires auxotrophic strains [103]. Selective labeling allows assignment for large systems with broad, overlapping signals or for functional studies where a specified residue in the binding or active site can be labeled and followed during a biological process. Different combinatorial strategies for selective labeling can be used to minimize the number of samples for assignment [103, 104]. Site-specific incorporation of unnatural amino acids during protein synthesis is also possible using cell-free expression due to the open and accessible nature of the reaction environment. The codon for the residue of interest is mutated to a nonsense codon that is recognized by a suppressor tRNA. This suppressor tRNA is chemically aminoacylated *in vitro* with the desired unnatural amino acid, and by adding this to the cell-free system together with the mutated gene, the unnatural amino acids is incorporated at the target site [105]. This instantly provides an "assignment" for the NMR signal of the unnatural amino acid, which can be used for site-directed screening for binders [106]. Cell-free expression also has a major contribution for the incorporation of more sophisticated stable isotope NMR approaches, such as the Stereo-Array Isotope Labeling (SAIL) method [107]. The SAIL amino acids are prepared by stereo-selective replacement of ^{1}H by ^{2}H and ^{13}C to ^{12}C at specific locations. This drastically reduces the number of observable protons and carbons relative to uniform sampling, with a corresponding improvement of magnetic relaxation rates generating sharper lines in acquired spectra. Taken together, the use of SAIL amino acids extends the molecular weight limit for routine NMR structure determination from 20 kDa to 40 kDa and beyond.

Protein purification

To properly characterize a protein, it needs to be relatively free from contaminants, and therefore, purification of protein mixtures is most often performed. Having the protein of interest in a pure form without contaminants also often increase stability of the sample, especially for the proteolytically sensitive IDPs. This group of protein is also often resistant against heat treatment. By increasing the temperature of the cell-free expression reaction after protein synthesis, most of the contaminating proteins can be denatured and precipitated, leaving the heat stable IDPs in solution. For CD79a and CD79b we have managed to go from gene to a $^{1}H^{15}N$-HSQC spectrum in less than 2 hours by heat-treating the cell-free reaction.

A powerful separation method often used today is chromatography. The protein mixture is dissolved in a liquid or a gas, known as the mobile phase, which carries it through a column consisting of a porous solid matrix, known as the stationary phase. The different solutes interact with the stationary phase in different manners depending on their individual properties, which affect their retention time from the column, i.e. the solutes are separated. In papers I-V, affinity chromatography and reverse-phase chromatography have been used extensively.

Affinity chromatography is one of the most diverse and powerful chromatographic methods for purification of specific molecules. It is based on the ability of proteins to bind specific molecules tightly but noncovalently. The interacting molecule (ligand) is placed onto the stationary phase while the target protein is in the mobile phase. Proteins not interacting with the ligands are washed away with the buffer. By changing the elution conditions, the target protein can be released from the matrix and thus be recovered in a highly purified form [108]. A widely employed affinity method is immobilized metal-affinity chromatography (IMAC). IMAC is based on the interaction between a metal ion (e.g. Co^{2+}, Ni^{2+}) immobilized on a matrix and specific amino acids in the proteins. Since histidine is the amino acid that interacts strongest with metal ions, sequences of consecutive histidines (called poly-histidine tags) are often used for affinity purification of recombinant proteins [109]. In papers I-IV, the constructs of the different protein genes adds an N-terminal poly-histidine tag, and thus IMAC purification has been performed.

Reversed-phase chromatography separates molecules according to hydrophobicity. Different solutes in the mobile phase binds the hydrophobic ligands attached to the stationary phase with different affinities. The solutes are initially applied to the column in aqueous buffer, and the solutes are eluted by the addition of organic solvents to the mobile

phase [110]. In papers II-IV, a gradient elution using acetonitrile and trifluoroacetic acid (TFA) as counter-ion has been used, where proteins were eluted in order of increasing molecular hydrophobicity.

Protein NMR spectroscopy

In 1946, the groups of Bloch and Purcell first described the phenomenon of NMR. The first NMR spectrum of a protein, a ribonuclease, was recorded in 1957, and Wüthrich and coworkers reported the first three-dimensional structure solved by NMR in 1985 [111]. Since then, NMR has been a powerful technique for obtaining structural information of macromolecules in solution at atomic resolution.

In NMR spectroscopy, radiofrequency (RF) waves are applied to a sample immersed in a strong and homogenous magnetic field (B_0). The atomic nuclei in the sample with odd mass number will have the property of spin, i.e. the nuclei behave as if they were spinning around a given axis. Since it is well known that a charged spinning particle will produce a magnetic field, these atoms will behave as small bar magnets in the large outer magnetic field. The magnetic field, induced by the rotating nucleus, can be described by a nuclear magnetic moment vector (μ). In the presence of B_0, the nuclear magnetic moment will precess around the axis of B_0, called Larmor precession. The frequency of this precession depends on intrinsic properties of the nucleus and the strength of the outer magnetic field. In a sample immersed in B_0, the nuclear magnetic moment vector will have the same direction as the outer field. In order to observe the nuclear magnetization, RF pulses are applied with a frequency that corresponds to the specific Larmor frequency of the nuclei to be observed. The RF pulse will put magnetizing vector perpendicular to B_0, and the nuclear magnetization will start to rotate in the plane perpendicular to the outer magnetic field with the Larmor frequency. This will be registered as a tiny current in a receiver coil. This signal is called free induction decay (FID). Since the nuclear magnetization strives to reach equilibrium, with the direction as the outer magnetic field, the FID will decrease with time. Fourier transform is applied to the signal in order to go from the time domain to the frequency domain, and an NMR spectrum will be obtained. The flow of electrons around a magnetic nucleus generates small local magnetic fields that oppose the outer magnetic field. The effective field of a certain nucleus is known as the chemical shift. The degree of shielding depends on the strength of B_0, the chemical structure and structural geometry of the molecule. The same nucleus in two different positions in the molecule can therefore possess different chemical shifts.

Using various pulse sequences with different RF pulses and delay times, the nuclear spins can be manipulated in numerous ways to extract different kinds of information and to obtain different dimensionalities of the NMR spectrum. When studying proteins, 1D spectra are far too complex for interpretation as most of the signals overlap heavily. By transferring magnetization from the excited nucleus to nearby nuclei in space or through-bond, NMR signals can be resolved in multiple dimensions. A prerequisite is that the transferring nuclei are NMR active, which often requires isotope enrichment of the molecule. Protein NMR utilizes mainly ^{1}H, ^{13}C, ^{15}N and ^{31}P as NMR active nuclei. Since it is only ^{1}H that normally are present in proteins, ^{13}C and ^{15}N enrichment of the protein during its biosynthesis is often performed for resolving signal overlap. Two examples of 2D spectra are shown in Figure 12, $^{1}H^{15}N$-HSQC of CD79a and CD3ε. Each peak corresponds to a correlation between the nuclei of a nitrogen-proton pair. For a deeper insight into NMR theory see Levitt [112] and Keeler [113].

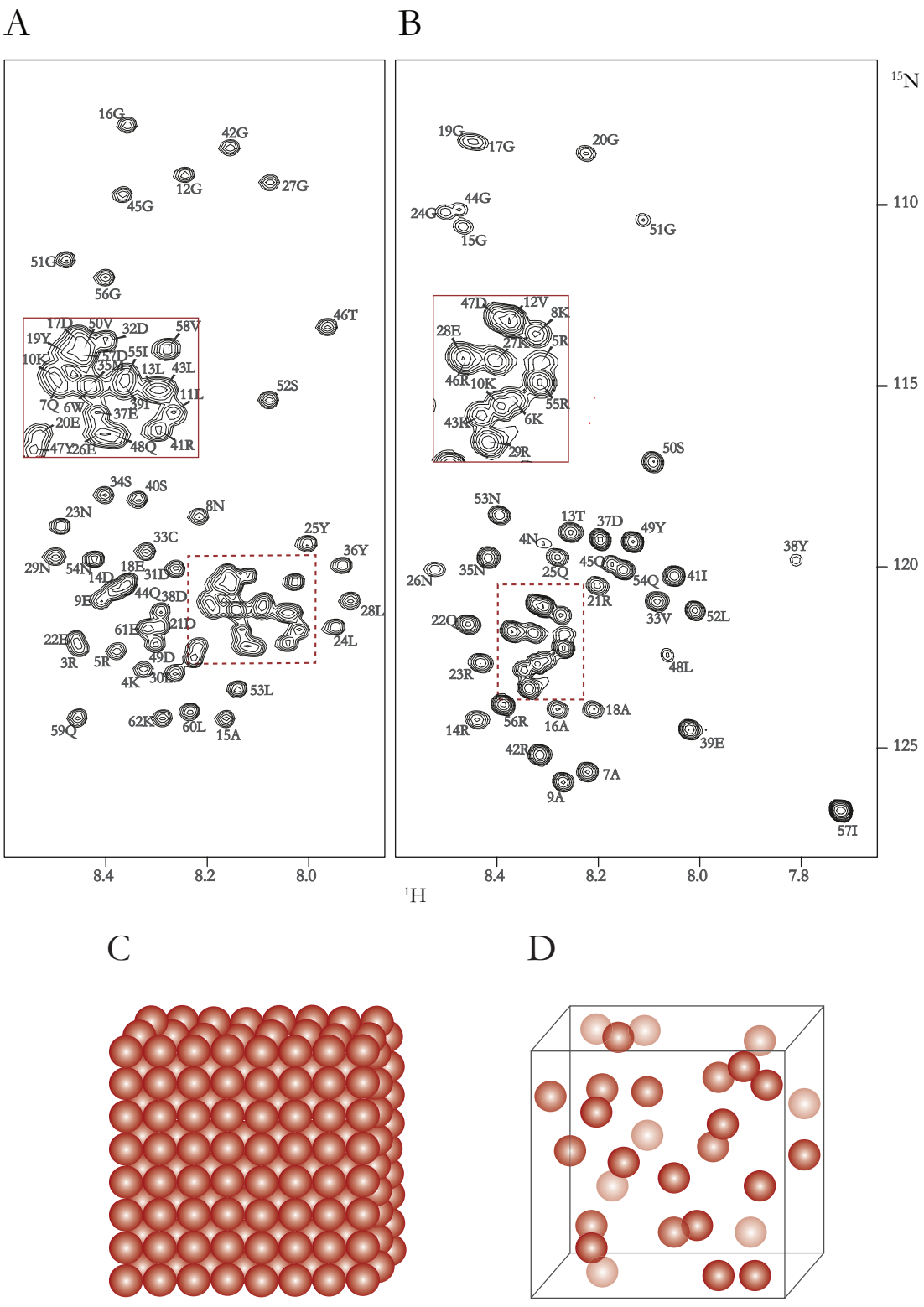


Figure 12. Annotated $^{1}H^{15}N$-HSQC of CD79a **(A)** and CD3ε **(B)** with crowded regions shown as inserts (Figure S1A and B Paper II). **(C)** and **(D)** Cartoon demonstrating the difference between uniform- **(C)** and non-uniform sampling **(D)**. For non-uniform sampling the dominating part of the sampling points are omitted.

Non-uniform sampling and Multi-Dimensional Decomposition

The oscillating NMR signal (FID) is a function of time that is converted to digital form by the analogue-to-digital converter (ADC). The ADC measures the voltage levels of the signal and store the information as a set of values in the computer. This process is called sampling [112]. Digitization involves sampling the signal at regular intervals chosen to give a good approximation to the smoothly varying signal. The digital representation of the FID consists of several data points evenly spaced in time. In order to convert the NMR signal into a NMR spectrum, Fourier transform (FT) is subjected to the digital signal. Fourier transform is a mathematical technique, which converts a function of time into a function of frequency. For a detailed description of FT see Levitt [112] and Keeler [113]. Various mathematical operations, including FT, are subjected to the digital signal in order to give a NMR spectrum. This procedure is called processing. A prerequisite for applying Fourier transform to the NMR signal in the traditional way is that the input data has to be sampled in a uniform way. This means that the time between each sampled point, called the dwell time, has to be constant. Since the time required to perform an experiment is proportional to the number of data points, this uniform sampling approach leads to very lengthy experiments for high-resolution multidimensional spectra. Instead of systematic sampling, non-uniform sampling (NUS) can be used that does not need a constant dwell time between each sampling point (Figure 12 C and D). Instead, randomly selected data points with varying time separations is used according to a pre-defined sampling schedule. In this way, a complete spectrum can be recovered from only a small number of data points. These data points are selected optimally by taking into account prior information about signal properties, e.g. the higher signal to noise in the beginning of the FID [114]. The non-uniform sampling approach requires nontraditional signal processing schemes such as non-linear Fourier transform, Maximum Entropy (ME) [115], Compressed Sensing (CS) [116], Maximum likelihood method (MLM) [117] or Multi-Dimensional Decomposition (MDD) [118-122].

In papers II-IV, non-uniform sampled data have been processed using MDD. Very briefly, this processing technique replenish missing data points (from the non-uniform sampled data) into a full data matrix using an interpolation model. The assumption, required for reconstructing the full spectrum in MDD, is that a NMR signal is completely defined by its line shapes in all spectral dimensions, i.e. one peak in a 3D spectrum can be described as three one-dimensional vectors. In the MDD model, an M-dimensional matrix corresponds to an input spectrum, whereas the one-dimensional vectors represent the output line shapes. By incorporating also constraints on the appearance of the line shapes (since in most cases the NMR line shapes are known), a larger fraction of the missing data can be predicted. This variant of MDD processing is called recursive-MDD

(r-MDD). The complete spectrum built from MDD is further processed as usual with Fourier transform, linear prediction etc.

Assignment

In order to get information associated with a particular atom in the protein, resonance assignment of the protein has to be accomplished. Assignment is typically recognized as an essential intermediate in NMR analysis of proteins [123], and in principle it means that each detected resonance frequency in the spectra will be assigned to its corresponding nucleus in the protein. Resonance signals that belong to the same amino acid are called to be part of a spin system. A number of different NMR experiments are recorded to obtain all necessary resonances for assigning the protein of interest. Assignment can be obtained using homonuclear methods (that observe only protons) or heteronuclear methods (involving ^{15}N, ^{13}C and ^{1}H) [124]. Heteronuclear assignment using multi-dimensional triple-resonance spectra has dramatically improved the speed and reliability of the protein assignment process, but it requires uniformly $^{15}N/^{13}C$ labeled proteins [125]. Sequential linking between neighboring residues are identified via spin-spin couplings. An example of a triple resonance experiment used for assignment is HNCA (Figure 13A). Magnetization is transferred from the amide proton to the nitrogen, which is recorded as the first dimension. Magnetization is then further transferred to the Cα and Cα-1, which is recorded as the second dimension. Finally the magnetization is transferred back to the amide proton, which corresponds to the third dimension. In principle, assignment could be based on pairs of experiments, for instance HNCO and HN(CA)CO. However, chemical shift overlap of the CO shift in the spectrum can lead to ambiguities. Instead, triple resonance experiments (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB and CBCA(CO)NH) are often used which relies on two shared shifts. This is more likely to give long segments of unambiguously assigned residues [124].

Instead of picking peaks manually and linking residues sequentially, automatic assignment procedures can be used. Conceptually the assignment process is very simple and several automated assignment programs have been developed, e.g. MARS [126], GARANT [127], and AutoAssign [125]. Commonly, a five-step analysis is performed for resonance assignment of $^{13}C/^{15}N$-labeled proteins: (1) peaks are picked and filtered out from spectral noise, and resonances are referenced across different spectra; (2) resonances are grouped into spin systems; (3) amino acid types are linked to the spin systems; (4) spin systems are sequentially linked; (5) spin systems are mapped onto the primary sequence. Even if all steps can be performed automatically, it is good to inspect manually the quality of picked peaks and spin systems, since a reliable distinction between peaks and noise is crucial for the assignment process [126].

Targeted Acquisition NMR Spectroscopy

In papers II-IV, a procedure called TANSY (Targeted Acquisition NMR Spectroscopy) has been used for rapid automatic backbone assignment (Figure 13B). The procedure combines several techniques to achieve high efficiency of data collection and analysis. The triple-resonance spectra needed for assignment are recorded using NUS with a sampling schedule matched to the signal intensities. BEST-TROSY [128] experiments are used in order to reduce measurement time. Additional time saving, besides using NUS and BEST-TROSY experiments, is obtained in the course of Targeted Acquisition (TA). Instead of completely recording all triple resonance spectra one after another, which often results in acquisition of either insufficient or highly redundant data, TA record and collect a small fraction of data in NUS mode for all experiments. This is performed in steps, where the data from each step is combined with previously acquired data (if present). Using TA, the time allocation for each triple resonance spectrum can be adjusted and optimized depending on the sensitivity of each experiment. For instance, the time allocation for the HNCA experiments can be much less than for the insensitive hnCOcaNH experiment. The data from TA is processed using r-MDD and signals, in forms of shapes obtained from MDD, are detected and used for automatic assignment with the AutoAssign program. Further time savings is obtained by picking signals automatically in a hyperdimensional way [129], which, in short, means that shared shapes in the recorded experiments (for instance the H and NH shifts in HNCO and HNCA) are fixed and only the unknown shape (CO and CA) are the ones that needs to be picked. Scores on completeness of the peak lists and achieved assignments are then evaluated and, if necessary, the TA procedure is repeated. When the number of peaks and assigned residues level off, the TA procedure is finished. By using this incremental acquisition, substantial timesavings can be made since the acquired information is monitored concurrently with data collection.

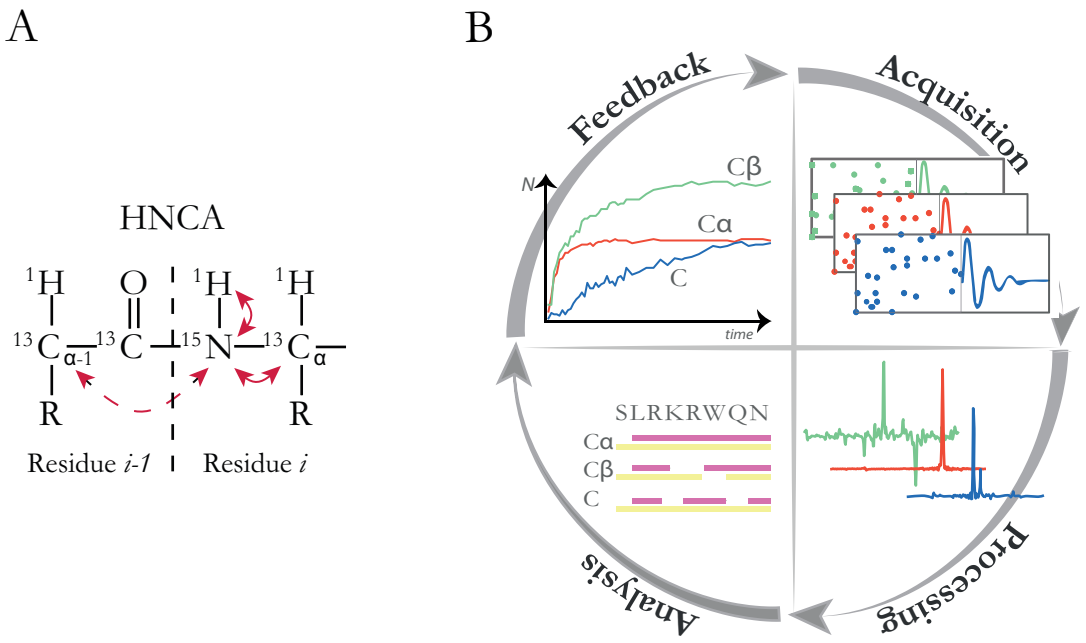


Figure 13. (A) Magnetization transfers in a HNCA experiment. All the Cα and Cα-1 shifts of the studied peptide are measured and recorded. **(B)** Schematic representation of the TA procedure. A small fraction of data is acquired using non-uniform sampling for all 3D experiments. The data is combined with earlier collected measurements, if present, and co-processed using co-MDD. Signals are detected and resulted one-dimensional shapes are used as input for AutoAssign. Feedback is received as scores on completeness of the peak lists and the achieved backbone assignment are evaluated and, if needed, the procedure starts from the acquire step again until number of peaks and assigned residues level off.

NMR and intrinsically disordered proteins

NMR is the most powerful and suitable technique for investigating the details of disordered proteins. Initially, NMR was used for demonstrating disorder but later emphasis was shifted to characterization of residual structure and to correlate this with protein functions. Today, IDPs are also investigated with NMR in living cells, so-called in-cell NMR. Using this technique, proteins can be studied at atomic level in intact cellular systems. The crowding effect from non-interacting molecules as well as post-translational modifications and interactions with other proteins and cellular components are taken into consideration [130].

IDPs are very dynamic and NMR is therefore especially well suited for studying dynamics and structure propensities for this protein class. Although the proton shift dispersion of IDPs are very poor, the slower relaxation rates often allows acquisition of spectra with narrow lines even for large proteins. Advances in sample preparation techniques, labeling schemes (like selective labeling) and pulse sequences tailored for assigning resonances for disordered proteins have facilitated the analysis of IDPs [131]. The primary observables in NMR studies of disordered proteins are

the chemical shifts. Variations in chemical shifts from random coil values (so-called secondary chemical shifts, SCS) can give important insights into secondary structure propensities [132]. Generally for IDPs, secondary structure is transient and limited to short segments of the protein. Structured elements can be indicative of binding sites for partners, and in order to detect the presence of residual structure, chemical shift deviations from random coil values are calculated. Residues located in β-sheet structure have negative $^{13}C\alpha$ and positive $^{13}C\beta$, the opposite being true for residues located in α-helices (Figure 14A) [131]. However, several random coil value sets are published in the literature and these differ because of the experimental conditions. Temperature, pH and neighboring residues are factors that affect random coil values. For instance, the most popular random coil data sets have been recorded at low pH, so corrections have to be made when NMR spectra are recorded at another pH. The effect of neighboring residues is the main challenge in defining random coil chemical shifts, since neighbors can directly and indirectly affect the chemical shifts for a certain residue. Introducing a systematic error because of incorrect referencing can therefore be a potential problem for secondary structure propensity calculations [133].

An alternative to the traditional random coil data is to use intrinsic referencing (used in papers II-IV). Intrinsic random coil shifts are obtained by assigning the protein of interest at denaturing conditions. These shifts are then subtracted from the shifts obtained at native conditions (Figure 14A). Using this type of referencing, the secondary chemical shifts fluctuation between neighboring residues is decreased compared to the traditional approach, providing a more relevant reference for the specific protein under study.

Regions of proteins that are distant in the primary sequence but close in space in a folded structure can be detected via NOEs. The NOE depends on the internuclear distance with an upper limit of about 5 Å. In the case of folded proteins, NOEs are measured and used to determine the global fold. In unfolded proteins, detection and assignment of NOEs is difficult, either because the population of tertiary structures is very rare or the ensemble that contains folded structures is too heterogeneous [132]. Another way for getting long-range information for IDPs is to covalently attach nitroxide spin labels to the protein of interest [134, 135]. These spin labels are paramagnetic, causing a broadening of nuclear spins. The spin label is introduced via site-directed mutagenesis. One example of a spin label is (S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl) methyl methanesulfonothioate), MTSL (paper II). MTSL is attached to the protein via a disulfide bond to a cysteine residue. By introducing cysteines at various positions in the amino acid sequence, several different versions of spin labeled protein can be obtained. The site for the spin label must be chosen with care in order to not disrupt or influence potential structure formation

sites. The distance information is obtained by assessing the broadening effect in a spectrum, often a $^{1}H^{15}N$-HSQC. The spectrum is recorded in both the oxidized (paramagnetic) and the reduced (diamagnetic) states. Differences in intensities, relaxation rates or line widths can give information about potential medium-range or long-range contacts within the IDP.

Other important NMR techniques for characterizing IDPs are the use of residual dipolar couplings (RDCs) to probe for conformational behavior, long-range order and structure propensities [131] as well as amide proton deuterium-hydrogen exchange. After placing the protein in D_2O, successive HSQCs are recorded and since the exchange is too fast for the HSQCs, the inability to resolve rapid exchange works as an indication of structural disorder [10]. To probe for dynamic information of IDPs, mainly R1 and R2 relaxation measurements as well as steady state $^{1}H^{15}N$ heteronuclear NOEs are used [132]. A complementary tool for the routinely used ^{1}H-detected NMR experiments for the study of disordered proteins is ^{13}C-detection approaches [136, 137]. The high solvent-exposure of the peptide backbone in disordered proteins leads to broadening of signals due to an increase in the amide proton exchange. This fact together with the low dispersion in the ^{1}H dimension and the fast relaxation of protons makes ^{13}C detected experiments a good complementary technique for the study of IDPs [131].

NMR is also a versatile tool for studying kinetic processes such as post-translational modifications, common events in IDP regulation. Series of two-dimensional ^{1}H-^{15}N correlation spectra, like SOFAST-HMQCs [138], have been used for studying these events in real-time [139]. For more complex systems, where signal overlap in 2D spectra is too severe and when tyrosine phosphorylation is the main objective, the resolution obtained in 2D mode is not sufficient. Unlike serine and threonine residues, for which amide signals acquire distinct chemical shifts upon phosphorylation, tyrosine amide peaks move only moderately in $^{1}H^{15}N$-HSQCs from their non-phosphorylated positions (Figure 14C and D). This was also observed in papers III and IV. In paper IV, a new alternative for following *in vitro* phosphorylation of CD79b using the 3D BEST-TROSY type of experiments in conjunction with non uniform sampling and multidimensional decomposition is presented. This new technique allows quantitative time-resolved kinetic studies.

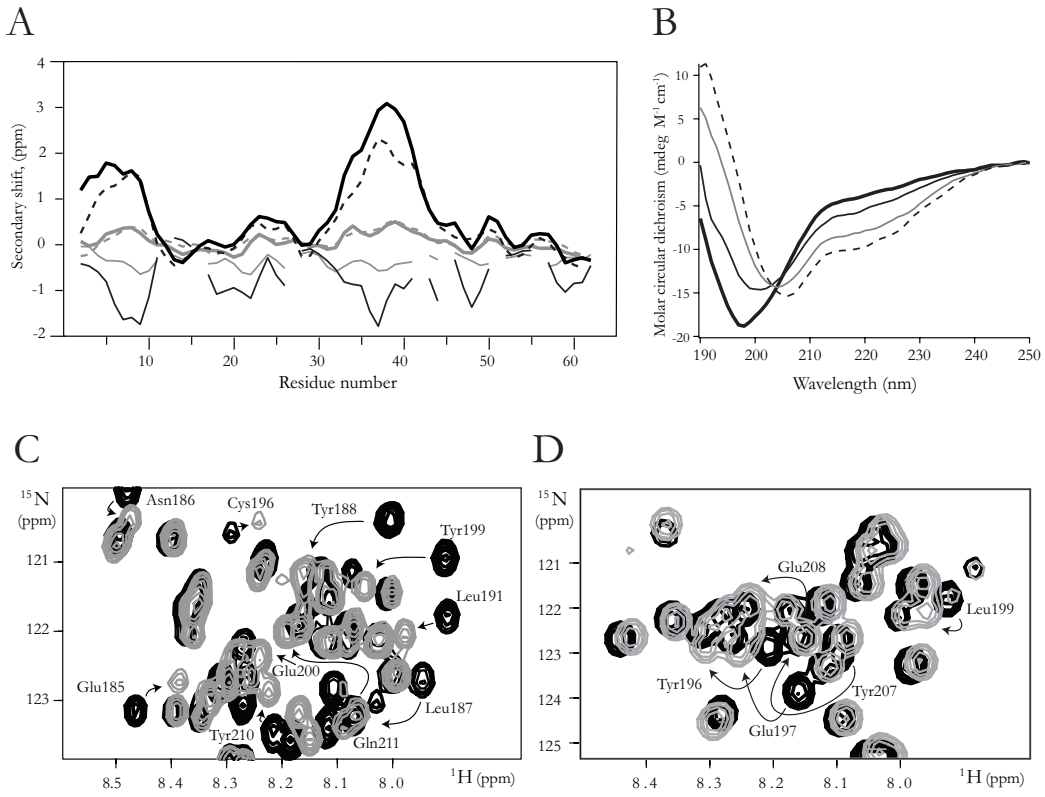


Figure 14. (A) ^{13}C secondary chemical shifts for CD79a in native (grey) and 20% α-helix promoting TFE (black) states. Secondary chemical shifts for CO (dashed lines), CA (bold lines) and CB shifts (thin lines) are obtained by subtracting the shifts with intrinsic random coil shifts, obtained in 6M urea. α-helix propensities are indicated by positive values for CO and CA, and negative for CB shifts (Figure 5 Paper II). **(B)** Far-UV circular dichroism spectra of CD79a in buffer (thick line) with increasing amounts of α-helix promoting TFE, (thin, grey and dashed line). α-helical features are visible as minima at 222 nmm, 208 nm and a positive value at 190 nm (Figure S3 Paper II). **(C)** and **(D)** Overlays of $^{1}H^{15}N$-HSQCs of CD79a **(C)** and CD79b **(D)** in their native (black) and phosphorylated state (grey) demonstrating the moderate shift of peaks upon phosphorylation and the need for higher dimensionality for real-time kinetic studies of these systems (Figure 3C Paper III).

CD spectroscopy

Plane polarized light can be viewed as being made up of two circularly polarized components of equal magnitude, one that rotates in a counter-clockwise manner, called left handed (L), and one that rotates in a clockwise manner, called right handed (R). Circular dichroism (CD) spectroscopy measures the difference in absorption of these two components. Only chiral molecules (optically active) result in a CD signal due to different absorption of the L and R handed circularly polarized light. In proteins, it is the peptide bonds, aromatic amino acid side chains and disulphide bonds that are observables for CD measurements. The spectra in the near-UV region (250-

350 nm) arise from the aromatic amino acids. The shape and magnitude of the CD spectrum in this region depend on the environment around these aromatic residues, giving a fingerprint of the tertiary structure of proteins (however not sufficient to yield significant structural insight) [140]. In the far-UV region (190-230 nm) it is mainly the peptide bonds that are observed. The different types of secondary structure found in proteins (α-helix, β-sheet, turn, PPII helix and random coil conformations) give rise to characteristic CD spectra in the far-UV region (Figure 14B). A number of different algorithms exist that estimate the secondary structure composition of proteins by approximating a spectrum as a linear combination of the contributions of different elements [10].

Results and discussion

Paper I

Can cell-free expression be used for synthesizing membrane proteins in sufficient amounts needed for structural biology studies?

Membrane proteins are critical for many essential cellular functions in the cell, e.g. accounting for over 50% of all human drug targets. However, less than 1% of the atomic structures deposited in PDB represent MPs, and since more and more structures of globular proteins are continuously added, this number is decreasing every day. Why is it that such an important protein class is falling behind the rest of the protein field? The major bottleneck is the inherent difficulties in producing proteins at milligram quantities necessary for most structure determination techniques. One of the major concerns for MP production *in vivo* is the potential aggregation and cytotoxicity of these proteins to the host, limiting the ability to express proteins at high levels. CFPS offers an alternative methodology for the overproduction of MPs. The proteins are produced in a test tube by reconstituting transcription and translation of a DNA template, added to the reaction. Since there is no need to maintain cell viability, cytotoxicity issues are avoided.

In paper I, we demonstrate the applicability of cell-free expression for a limited screen of 38 integral membrane proteins. The constructs were of both prokaryotic and eukaryotic origin with sizes ranging from 12 to 114 kDa. All included targets were α-helical MPs with 1-14 predicted transmembrane helices. 37 of these 38 constructs expressed at sub-milligram quantities, suitable for structural biology studies.

Savage *et al.* have reported a correlation between expression yields and target protein size [141]. Targets most likely to express well were of lower molecular weight with fewer transmembrane helices. In paper I, we show that no general bias towards high expression levels of smaller proteins was observed. For instance, *sc*Fps1p and *ec*EnvZ, with sizes of 76.3 and 50.4 kDa, respectively, were produced at high levels (>0.5 mg/ml). We also observed that constructs with more than 10 helices (hTH and hGLUT1) expressed weakly probably because of the increased complexity of the membrane domains. Only 1 of the 38 included MPs (tt4619) did not express with our system. The size of tt4619 is 43.8 kDa and the number of transmembrane helices for this protein is two. Why this protein did not express is not known, and no further attempts to express it were made. A correlation between large numbers of predicted ribosomal translational pause sites and low expression yields of the included targets could, however,

be observed, as well as shorter fragments for some proteins, with sizes correlated to the presence of potential pause sites.

Cell-free expression offers the MPs to be co-translationally solubilized due to the open nature of the system, and detergent screens can easily be performed. The choice of membrane-mimicking system is crucial for both folding and functionality of MPs [142]. In paper I, we tested detergents of different charges, tail length and head group functionality (Figure 15A). Brij58 appeared to be the best choice from a solubilization point of view. Digitonin also proved to be successful in all tested cases. Out of the NMR suitable detergents: DHPC, LMPG and LPPG, only DHPC was a good candidate for co-translational solubilization. LPPG and LMPG have been used succesfully by others to solubilize MPs produced with CFPS as precipitates after expression is completed [92].

Several of the MPs included in this paper have previously been described to be hard to overexpress *in vivo*. Yields of *sc*Fps1p have been reported to be below 1 mg/L fermentor culture, more likely in the μg scale [143-145]. In this paper, we show expression levels well above 0.5 mg/ml. In this work, we also observed high-level expression of hAQP6, 9 and 11, previously reported to be hard to overproduce [146]. The most exciting result in this work was the well expressing hCXCR4, *sc*Ste2p and hD2L, proteins belonging to the G-protein coupled receptors (GPCRs), a class notoriously difficult to produce and the targets for approximately 50% of all modern medical drugs [147].

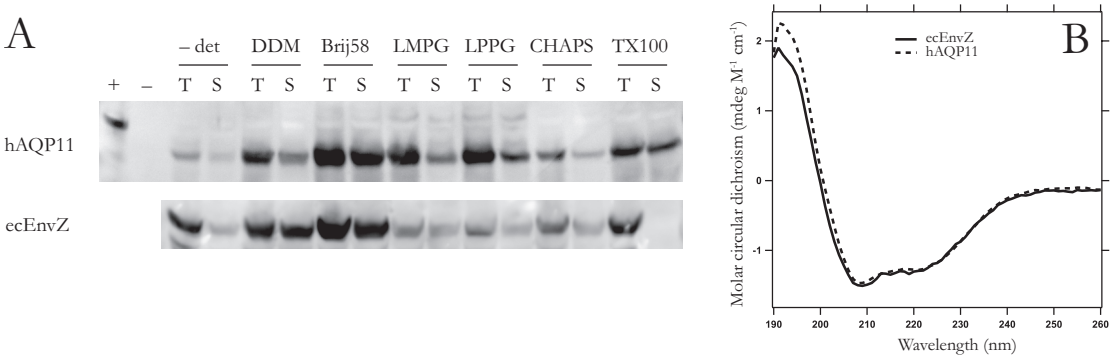


Figure 15. (A) Western blots of detergent screens for hAQP11 and *ec*EnvZ. T and S denote total and soluble fractions, respectively. **(B)** Far-UV circular dichroism spectra of hAPQ11 and *ec*EnvZ, indicating folded proteins.

A solubilized MP need not be equal to a functional protein. In spite of many attempts we were not able to generate functional *np*SRII or distinguish specific binding from unspecific binding for hD2L using a standard radioligand-binding assay. However, we could generate functional pR incorporated into nanodiscs and we observed folded hAQP11 and *ec*EnvZ using CD spectroscopy (Figure 15B). It was beyond the scope of this paper to test functionality of all included MP targets.

In conclusion, we have proved that CFPS can be successfully applied to express preparative amounts of MPs, suitable for structural biology studies. We demonstrate successful folding and functionality for selected targets and present effective co-translational solubilization agents for the investigated proteins.

Papers II & III

Can cell-free expression and NMR be combined for efficient characterization of intrinsically disordered proteins? Do the domains of the B cell receptor contain any secondary structure propensities, and how do their phosphorylation patterns look like?

The characterization of the conformational properties and interaction mode of intrinsically disordered proteins has recently become a major research topic in order to understand biological function at a molecular level. NMR spectroscopy has become the technique of choice for extracting information at atomic resolution for this protein class. In particular, NMR is very useful for revealing regions of increased propensity to form α-helical or extended (β-strand) structure in the otherwise random coil-like structured states. Characterization of IDPs typically requires preparation and analysis of many protein samples. In paper I, we present new methodology for efficient characterization of IDPs. CFPS is used as a rapid and low-cost protein production system in conjunction with fast pulsing NMR experiments, combined with non-uniform sampling, targeted acquisition, automatic signal detection and backbone assignment. The technique is demonstrated for five cytosolic domains of the B- and T cell receptors. The combined NMR methodology is introduced with the name Targeted Acquisition NMR Spectroscopy (TANSY).

In this work, we demonstrate for the first time the systematic use of CFPS for producing IDPs for NMR analysis. CFPS is an excellent tool for producing the proteolytically prone disordered proteins because of the presence of protease inhibitors in the cell-free reactions. Since most of the protein is synthesized within one hour, screening of different conditions, temperatures, and additives can be performed very rapidly. Because of the high amino acid incorporation level we have in our system, the amount of protein produced in just a reaction of one ml (corresponding to a cost of $20) was sufficient for backbone assignment. We also present assignment for a sample with a concentration of only 30 μM. This is among the lowest protein concentrations for which protein backbone sequential assignment has ever been reported. In this work, the time needed to assign a 330 μM sample was only 48 minutes, which is about 100 times faster in comparison to the conventional methods of data collection and analysis.

We also introduce a new method for validating protein sequential assignment using jackknife-type resampling. This basically means that using the raw-data, we randomly extract about 70% of the points, process it, detect signals and perform automated assignment. This resampling is repeated a number of times and assignment is cross-validated for every residue. Demonstration of this new resampling-based validation for two

protein samples showed that our validation technique better revealed problematic assignments and more efficiently found the correct solution, compared to the validation suite from e.g. AutoAssign.

The NMR platform for characterization of disordered proteins, described above, has been implemented to study CD79a and CD79b, the disordered domains of the B cell receptor. Analysis of the secondary chemical shifts was made using intrinsic referencing, where both proteins were assigned under denaturing conditions (6M urea) in order to overcome some of the problems traditionally associated with reference random coils shifts found in literature. For CD79a, the CO, CA and CB secondary chemical shifts (Figure 16A) revealed propensity for α-helical structure around the signaling-responsible ITAM motif. In the N-terminus, we observed another residual helix. Both CD79a and CD79b have ITAM motifs, containing tyrosines that upon phosphorylation start the intracellular signaling cascade. Regions containing secondary structure propensity are often biologically relevant and involved in, for instance, interactions. Phosphorylation of residues located in or close to these sites can modulate secondary structure, which consequently can result in altered function or mode of interaction. In paper III, we examine the secondary structure propensities of CD79a and CD79b before and after *in vitro* phosphorylation by Fyn kinase, and how phosphorylation affects these propensities. Similar to CD79a, CD79b displayed propensity for α-helix formation in the ITAM region (Figure 16B) in the non-phosphorylated state. An additional helix could also be observed in the N-terminus. Subtracting the chemical shifts of the phosphorylated state from the non-phosphorylated state for CD79a revealed three phosphorylation sites (the two ITAM tyrosines plus one tyrosine preceding the ITAM), which was also confirmed with mass spectrometry. For CD79b, only the two ITAM tyrosines were phosphorylated, which was also confirmed with mass spectrometry. Since possible long-range interactions can perturb chemical shifts, we subtracted the chemical shifts from the phosphorylated state in 6M urea with the non-phosphorylated state in 6M urea. Under these conditions, long-range interactions are disrupted and the change is largely contributed by the changes in random coil chemical shifts induced by the covalent attachment of a phosphate group. The result from these measurements showed that the shift perturbation was induced by phosphorylation and not long-range interactions. We also compared the secondary chemical shifts of CD79a and CD79b in the phosphorylated and non-phosphorylated forms in order to see whether the secondary structure propensities were affected (Figure 16). For CD79b, the region of the transient α-helix located in the ITAM motif was enhanced upon phosphorylation. This stabilizing effect we believe come from the favorable electrostatic interactions between the phosphoryl groups, positioned in the N-terminus of the helix, and the helix dipole. The other tyrosine that we observed to be phosphorylated did not alter the α-

helical content around its site. In contrast, CD79a phosphorylation resulted in an overall reduction of helical propensity for the α-helix region observed around the second ITAM tyrosine. This tyrosine is located close to the center of the helix region, and since it previously has been shown that phosphorylation of a serine in the middle or in the C-terminal part of a helix has an overall destabilizing effect on that helix [44], this result was not unexpected.

In conclusion, we present an efficient methodology for extensive and rapid characterization of highly disordered proteins in solution. This technique was implemented for the detailed studies of CD79a and CD79b, revealing α-helical structure propensities for both proteins in the C-terminal part of the ITAM region. The phosphorylation patterns for CD79a and CD79b were analyzed and for CD79b, an increase of helical structure was observed upon phosphorylation, in contrast to CD79a, where a decrease in helix content was detected.

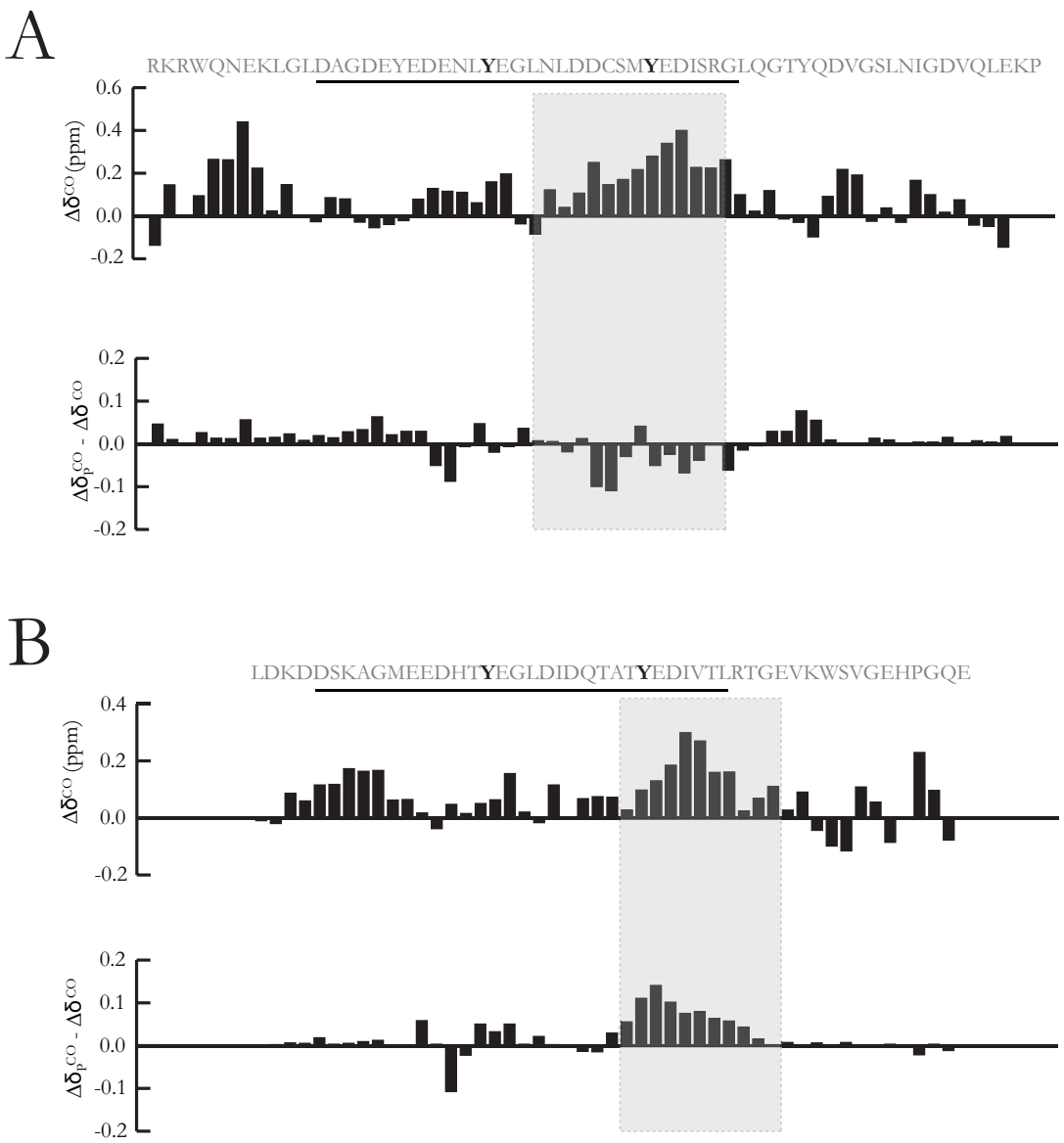


Figure 16. CO secondary chemical shifts for CD79a **(A)** and CD79b **(B)** in the upper diagrams respectively. The ITAM motifs are underlined and the two tyrosines that get phosphorylated in the ITAM are shown in bold and black. **(A)** CD79a shows α-helical propensity at two major sites: in the N-terminal region and in the ITAM motif (grey box). **(B)** CD79b also contains two major α-helical tendencies: in the N-terminal part of the protein and in the C-terminal part of the ITAM (grey box). The propensities for α-helix after phosphorylation of CD79a and CD79b are shown in the lower diagrams, respectively, where secondary chemical shifts for the non-phosphorylated states are subtracted from the phosphorylated states. For CD79a **(A)** this propensity is decreased for the ITAM α-helix, in contrast to CD79b **(B)**, where the ITAM α-helix tendency is increased.

Paper IV

Can 3D NMR experiments be used for time-resolved kinetic studies with timescales of minutes?

A detailed understanding of cellular mechanisms requires knowledge of structure and dynamics of the involved biomolecules at atomic resolution. NMR combines high resolution with dynamic analyses on timescales ranging from picoseconds to days [148]. Since 1D NMR spectra are not sufficient for studying large molecules such as proteins, 2D spectroscopy have been widely used both in single scan 2D methods, for highly concentrated samples, and by recording series of ^{1}H-^{15}N correlation spectra, e.g. the SOFAST-HMQC experiment. As the complexity of the biological system increases, 2D experiments can be insufficient for resolving the critical resonances in the spectra. Conventional 3D experiments, however, requires a data collection scheme which is incompatible with kinetic processes if the temporal resolution is in the order of minutes. In paper IV, we demonstrate a unique method for studying dynamic events using non-uniformly sampled 3D BEST-TROSY HNCO experiments together with co-processing and multidimensional decomposition (Figure 17). We use non-uniform sampling to continuously acquire a 3D experiment with the duration covering the period of the studied reaction. Instead of processing the entire spectrum, the experiment is sliced in a number of time windows, where each window gives an individual spectrum. Since the size of the time window is specified first in the processing stage, the duration of each window can be optimized depending on the reaction rate studied. This stands in contrast to the method of consecutively recording a series of 2D spectra, where the duration for each spectrum needs to be decided prior to the experiment. In this paper, we both demonstrate a linear time window selection, where number of points are equal for each window, and an exponential time window where every window contains more points than the previous one. Using co-processing of series of NMR experiments, only >1% of the measurements required for the conventionally sampled 3D experiments are acquired. The reaction kinetics is obtained from the analysis of peak intensities in the different spectra corresponding to individual time points of the reaction.

In this work, we demonstrate that our technique can be used to measure amide hydrogen-deuterium exchange rates with half-times of a couple of minutes and larger. Also, the phosphorylation of the two ITAM tyrosines of CD79b has been followed in a time-resolved way using this technique. In contrast to serine and threonine, whose shifts moves to distant often empty regions for non-phosphorylated proteins, tyrosine shifts are not so dramatic, often overlapping with other amide proton peaks. For CD79b, both of the tyrosines are in the crowded region in the $^{1}H^{15}N$-HSQC, both

for the phosphorylated and non-phosphorylated states (Figure 14D). However, both signals could be well resolved in the 3D HNCO spectrum. Time evolution of the signal intensities for both tyrosines, and directly adjacent residues, were plotted and first order kinetic was observed for both tyrosines with phosphorylation rates around 0.25 and 0.19 h^{-1} respectively. This observation indicates that phosphorylation of the two CD79b ITAM tyrosines by the Fyn kinase, used in this study, occurs independently.

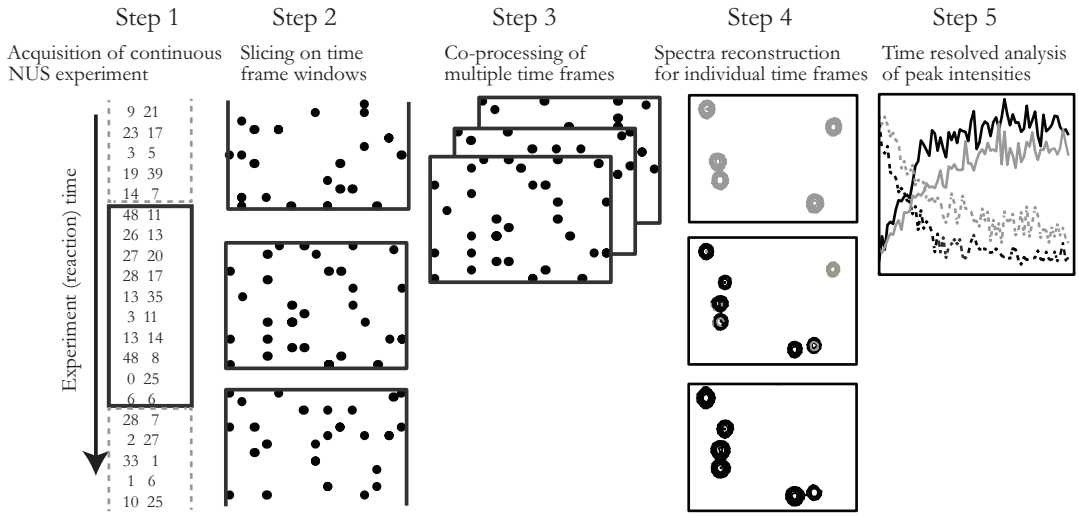


Figure 17. Schematic overview of the time-resolved analysis using non-uniform sampling and co-MDD.

In conclusion, we present a unique method for studying time-resolved kinetic processes in multidimensional experiments with time resolutions of a few minutes. A continuous 3D experiment is acquired using non-uniform sampling, with a duration covering the reaction time. Co-processing together with multidimensional decomposition can slice up the total experiment in time windows and hence quantitative site-specific analysis can be made throughout the entire kinetic process.

Paper V

How can the incorporation of amino acids be improved in the cell-free system?

CFPS has emerged as a powerful technology for protein production, especially for expression of isotopically labeled amino acids for NMR spectroscopic analysis. *In vitro* protein production offers the ability to both generate uniform, selective and site-directed labeling of amino acids in a protein without severe scrambling, due to the diminished metabolic activity of the necessary cellular extract. The small reaction volumes also make *in vitro* protein expression a more convenient option for labeling, and heteronuclear single quantum coherence spectra can be recorded directly on the reaction mixture without purification. Isotopically labeled amino acids represent the most expensive reagents for our batch mode *E. coli* based CFPS system (accounting for about 72% of the total cost). In this work, we therefore wanted to investigate how the efficiency of incorporation can be improved and why serine and glutamine are crucial for optimal expression yields.

Using stoichiometric amino acid mixtures, based on the relative occurrence of the 20 amino acids for GB1, we obtained the same yields compared to the reference sample (using 1mM of all amino acids). Since serine and glutamine are important for high expression yields, these amino acids were kept at a concentration of 2 mM and 4 mM, respectively, for both samples. Total amount of amino acids added to the optimized and reference sample of GB1 was 0.12 and 0.41 mg, respectively, which means that only 31% of the standard amounts of amino acids could be used without affecting overall yields of GB1. Excluding serine and glutamine, incorporation efficiency was increased from 4.8% to 33%.

Metabolization of glutamine could be attributed to the two glutaminases present in *E. coli*. This activity could be inhibited by adding 6-diazo-8-oxo-L-norleucine (DON) or partly inhibited by adding D-glutamine. Serine was rapidly deaminated to pyruvate by serine deaminases present in the extract. Pyruvate dehydrogenase complex (PDC) together with phosphate acetyltransferase and acetate kinase further processed the pyruvate to acetate. NAD^+ and CoA are important cofactors for PDC, and addition of these molecules confirmed the involvement of PDC as the deamination rate of serine increased severalfold. For an efficient incorporation of serine, deletion of the responsible deaminase genes (sdaA and sdaB) of the extract strain has to be made.

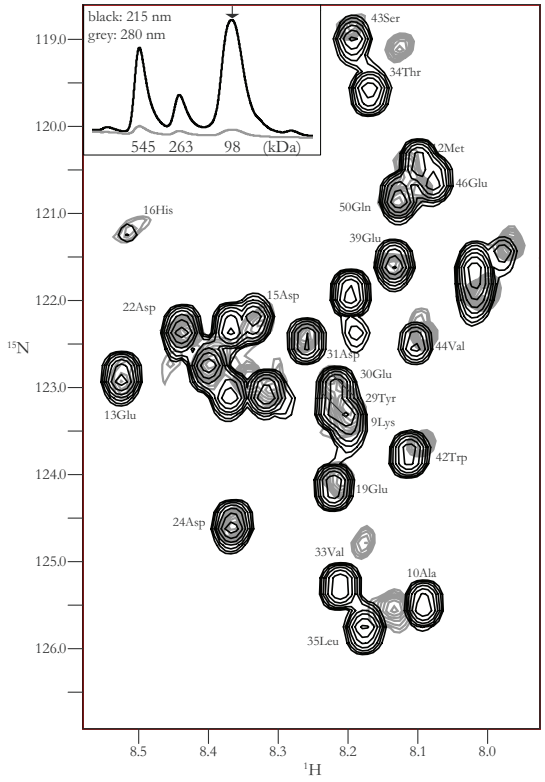
Future perspective

At the turn of the century it was systematically shown that many partially or completely disordered proteins are functional, expanding the protein universe to also include intrinsically disordered proteins (IDPs). The interest for this "new protein class" has increased significantly during my PhD studies. In 2010, for instance, I took part in the first Gordon research conference exclusively about IDPs. Since then, many meetings, summer schools and conferences have been held with IDPs as the major topic. The field has changed from just reporting that disorder exists and trying to define what it means to show actual systems and how they function and interact with other proteins. In paper I, we demonstrate a unique methodology for characterizing these systems. As far as I know, we are the first group that systematically uses CFPS for producing IDPs for NMR analysis, despite of all its advantages.

The transmembrane signal transduction mechanism for the B cell receptor is not fully understood. The importance of proper signaling pathways in the immune system makes the study of B cell activity an essential research topic, since dysregulation of activity are related to a number of diseases. The SCHOOL model, presented by Sigalov *et al.* [70], where homooligomers of CD79a and CD79b are responsible for signaling, is highly doubtful. During my PhD period I have measured intermolecular PREs between CD79a without seeing any oligomer formation. A recent publication by Mittag *et al.* [149], also reports the absence of dimerization for the T cell receptor zeta unit and the CD3ε domain. These results together with my observations stand in contrast to a paper by Sigalov *et al.* [150], reporting dimer formation of the zeta unit without disorder-to-order transition.

How oligomerization of the BCRs results in a change of the cytoplasmic CD79a and CD79b making them more prone for phosphorylation is also still not addressed. Reports on α-helix formation for CD3ε, hiding the ITAM tyrosines in the membrane, can be a potential theory also applicable for the B cell receptor. According to this theory, the binding of antigen perturbs the membrane lipid composition making the tails detach from the membrane and thereby be accessible for tyrosine kinases. However, Tolar *et al.* have observed that the cytoplasmic domains of CD79a and CD79b do not interact with the cell membrane [151]. I have studied potential membrane interaction for both CD79a and CD79b. Membrane lipid strips, spotted with 15 different biological important lipids found in cell membranes, have shown that CD79a interacts with negatively charged lipids like phosphatidylserine (PS) and phosphatidylinositol (PI), present in the membrane of B cells. It also appears that CD79a interacts

with a component present in the cell-free reaction, since the $^{1}H^{15}N$-HSQC, recorded for the reaction mixture, contains very few signals compared to CD79b, where all signals are visible. Our suspicion is that it is the lipids present in the extract that are the interacting components for CD79a. I have also performed titration of negatively charged bicelles to the cytoplasmic domain of CD79a, observing interaction of N-terminal residues as well as residues in the ITAM domain. I have also examined potential interaction of both constructs in liposomes and *E. coli*- and mammalian cell lysates, with a trend showing that CD79a interacts in contrast to CD79b. Finally, I have cloned both domains to also include the transmembrane part to see whether the membrane interaction is altered compared to the cytoplasmic constructs. In Figure 18, CD79b is incorporated into DMPC nanodiscs. It seems like CD79b weakly interacts with the membrane with a region close to the second tyrosine in the ITAM motif. We were not able to get any nice spectrum of CD79a, probably because of strong interaction to the nanodiscs. All these results points to that membrane interaction can be an important regulation event for the B cell receptor, in contrast to Tolar *et al.* In the future it would therefore be interesting to continue to work with this theory, confirming the above mentioned results as well as testing new nanodisc lipid compositions and other membrane-mimicking systems.



Another important question is what role the transient α-helices, observed for CD79a and CD79b, have for B cell receptor signaling. Most often when a propensity for structure is observed in IDPs, it is believed that they are sites of interactions. However, transient structure formation might also work as protection from proteases or to increase stability of the proteins. Since a previous publication demonstrates formation of helix-like character for an ITAM-derived peptide upon interaction to the Lyn kinase, it is tempting to believe that these observed transient helices work as α-molecular recognition

Figure 18. $^{1}H^{15}N$-HSQC spectrum of CD79b cytosolic (black) and CD79b TM+cyt incorporated into dDMPC nanodisc (grey). In the upper left corner the gelfiltration profile for the CD79b TM+cyt in nanodisc, where the marked peak is CD79b incorporated into nanodisc.

features (α-MoRFs). Through a conformational selection mechanism, these α-MoRFs interact with Src family kinases, and the transient helices are stabilized.

It is still unclear what role the phosphorylation of CD79a and CD79b has from a structure point of view for B cell receptor signaling. Why is the α-helical propensity increased for CD79b upon phosphorylation and decreased for CD79a? Interestingly, tyrosine phosphorylation has previously been reported to correlate with helix-to-coil transitions in structured systems, affecting protein function [152, 153]. Also regulation of binding through phosphorylation-mediated modulation of secondary structure propensity has been reported for instance for the disordered protein 4E-binding protein I (4E-BP1) [43]. The tendency for α-helical structure for CD79a and CD79b, indicated by secondary chemical shifts, does not exclude the presence of other secondary structure species. Since the presence of helical and β/extended structure have opposite effects on secondary chemical shifts, the only conclusion we can draw is that residual helix has the largest contribution. In addition, the possibility of onset of non-helical structure for CD79a upon phosphorylation cannot be excluded. It might be that instead of destabilizing the helix, β/extended structure is formed. Futterer *et al.* [154] have solved the crystal structure of the tandem SH2 domain of Syk complexed with a dually phosphorylated ITAM peptide. The peptide is in an extended conformation in the two pYxxL/I motifs while residues in the inter-motif region formed nearly a complete α-helical turn. This can be an example of the promiscuity of IDPs, that they can adjust their structure to specific binding partners via conformational selection. However, this does not explain why CD79b adopts more helical structure upon phosphorylation, in contrast to CD79a. Clark *et al.* [155] suggest that the two signaling domains serve non-redundant function in B cells, based on findings that they associate with distinct sets of effector molecules, and thereby activate distinct second messenger pathways.

Acknowledgements

This thesis summarizes my years as a PhD student. It has been an exciting time with both joy, enthusiasm, despair and coffee breaks (or rather tea breaks). I have experienced a lot and gotten to meet a lot of incredible people I hereby would like to thank:

Vladislav, for trusting me and enrolling me as a PhD student. I have enjoyed working in your research group and I have learned a lot from you during the years, ranging from how to run NMR to how eggs are cooked. :) Thank you for always keeping the door open and for patiently explaining everything to me! Thank you for your inspiration and for your guidance!

Anders, vart ska jag börja? Du har varit klippan från dag ett när jag började exjobba och fram till nu. Tack för att du alltid tagit dig tid för att handleda, diskutera och hjälpa mig med allt från himmel till jord. Du är en otroligt omtänksam person med en enorm kunskapsbas. Har lärt mig så ofattbart mycket från dig! Vi har hängt mycket genom åren, både på labb, konferenser och lunchraster, något som jag kommer att sakna väldigt mycket!

Martin Billeter, Jan Rydström, för att ni är och har varit mina examinatorer och för trevliga kakstunder!

Peter Tompa, for accepting to be my opponent. It is a true honour!

Göran, för din ständiga omtanke och för att du alltid har tagit dig tid när jag kommit med mina funderingar... Tack för alla trevliga (och ibland udda) lunchdiskussioner med allt från global uppvärmning, brasilian wax och huruvida Gotland är värt att resa till. :)

Maria S, min siamesiska lab-tvilling! Är ju inte konstigt att jag säger fel och säger mamma till dig ibland... Du har varit som en storasyster (ibland även lillasyster) och vi har haft riktigt kul genom åren! Inte bara på labb utan även privat. Gillar att hänga med dig och din ständigt gravida mage! :)

Joakim, inte har jag ångrat en sekund att jag och Maria tog dig som examensarbetare. Med exakt samma godissmak kan det ju inte gå fel! Du är otroligt duktig och ambitiös och vi har haft det riktigt kul ihop. Våra fenomenala tidsplaner har kanske inte alltid fungerat så bra, men viljan har i alla fall funnits där! Saknade dig när du slutade (och när du dessutom flydde staden). Vi måste dock hålla kontakten och ses ändå!

Maxim, for always being so friendly and helpful! I have enjoyed working with you during these years. You are so talented and smart. Thank you for always taking time for my many questions, even if you have million other things to do! :)

Cissi, för alla delade duschtillfällen. :) Du är en genuint genomtrevlig och ärlig människa som jag kommer att sakna! Tack för att du alltid tagit dig tid för att hjälpa mig med NMR experiment, CCPN, tebeställningar och annat livsnödvändigt. Någon gång ska vi allt få till den där tennisträningen, om jag bara vågar...

Johan, en sådan smarting är det inte ofta man stöter på! Saknade dig när du slutade. Vi hade en riktigt kul tid när vi delade kontor och på vår Sicilienresa.

Diana, för all hjälp jag fått genom åren, och alla fina barnsaker! De har använts flitigt. Du är amibitiös och duktig och kommer att gå långt, det vet jag!

Madelene P, vad vore livet utan trånga passager? Du har varit en frisk fläkt och härligheten personifierad. Har haft himla trevligt på labb tack vare dig!

Karin R, Karin E, kul att dela labb med er! Synd (för mig) att ni flyttade till Lund. Kommer att gå toppen för er båda, ni är duktiga!

Olga B, for a nice and productive collaboration! Hope to see you soon in Gothenburg again!

Pernilla, Emma, Nina, Rowena, Julia, Billy, **Millie** för sällskap på labbet/NMR centrum och trevliga diskussioner.

Urszula, Jenny, Mike, David, Linda, Mikael, Anna F, Erik M, Gerhard, Fredrik Ö, Weixiao, Daniel F, Jonas F, för sällskap i labbet och massa roligheter!

Kristina Hedfalk, Jonas Carlsson, Martina K, Lennart Sjölin och Christer Gruvberg för ett riktigt bra samarbete under kemins år! Vi fick till det bra! :)

Lars-Gunnar, Patrik, Anders, NMR-maskinexperterna som håller dem i skick!

All the PIs at Lundberg and the rest of the lab people

Alex, Åsa, Helén, Magnus, Mate, Ulrika för intressanta lunch-diskussioner och sällskap på NMR centrum.

Maj-Lis Sjöbäck, för ditt stöd, engagemang och mycket trevliga sällskap under mitt år (och även efterkommande år) i mentorprogrammet.

Nina, min stadiga sjömatros till bästa kompis. Tack för att du alltid finns där!

Madde, min lilla norska. Tack för att du alltid tar dig tid att ringa och hälsa på. Gillar dig!

Jane, för alla lekparksträffar, kalas, nyårsmiddagar, after work m.m. Du är en fin vän som alltid tar dig tid att lyssna!

Elin och Lasse, för alla roliga resor, luncher och middagar! Vi är sjukt lika Elin, det bara är så...

Kerstin, tack för att du va den bästa kompisen i Linköping. Jag saknar att hänga med dig! Du är och förblir min smartaste kompis! :) Tack även till, den inte lika fullt smarta, **Carl Persson**, för att du kom och förgyllde Linköpingstiden med din sjuka humor! Hade inte haft så roligt om inte du flyttade dit... :)

Frida, Jacqueline, David, Ingrid, Jenny och Sanna, för 5 roliga år i Linköping med allt vad det innebar!

Suhra, för all din hjälp med barnpass, renovering, matlagning, koloniodling m.m. Du har varit ovärderlig! **Azra**, för att du alltid är så glad och snäll! Arvid har tur som fått en sådan fin faster.

Eric, Anders och John, för att ni är de bästa bröder som man kan ha! Och tack **Malin** och **Marie** för att ni står ut med dem! :)

Stina Gylling, för din otroliga generositet och hjälpsamhet! Vi har haft det mycket trevligt på alla våra resor, något som jag aldrig kommer att glömma!

Mamma och pappa, för att Fjällstigen fortfarande är den bästa platsen på jorden och för allt stöd och hjälp som ni alltid ger!

Mirsad och Arvid, för att ni är den finaste lilla familj man kan ha! Ni betyder allt!

References

1. Katzen, F., *Cell-free protein expression of membrane proteins using nanolipoprotein particles.* BioTechniques, 2008. **45**(2): p. 190.
2. Katzen, F., et al., *Insertion of membrane proteins into discoidal membranes using a cell-free protein expression approach.* Journal of proteome research, 2008. **7**(8): p. 3535-42.
3. Isaksson, L., et al., *Expression screening of membrane proteins with cell-free protein synthesis.* Protein expression and purification, 2012. **82**(1): p. 218-25.
4. Raschle, T., et al., *Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs.* Journal of the American Chemical Society, 2009. **131**(49): p. 17777-9.
5. Hagn, F., et al., *Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins.* Journal of the American Chemical Society, 2013. **135**(5): p. 1919-25.
6. Ujwal, R. and J.U. Bowie, *Crystallizing membrane proteins using lipidic bicelles.* Methods, 2011. **55**(4): p. 337-41.
7. Denisov, I.G., et al., *Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size.* Journal of the American Chemical Society, 2004. **126**(11): p. 3477-87.
8. Nath, A., W.M. Atkins, and S.G. Sligar, *Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins.* Biochemistry, 2007. **46**(8): p. 2059-69.
9. Campbell M, F., S, *Biochemistry*. 4th ed. 2003 Thomson learning
10. Tompa, P., *Structure and Function of Intrinsically Disordered Proteins.* 2010.
11. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature, 1958. **181**(4610): p. 662-6.
12. Perutz, M.F., *Structure of hemoglobin.* Brookhaven symposia in biology, 1960. **13**: p. 165-83.
13. Blake, C.C., et al., *Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution.* Nature, 1965. **206**(4986): p. 757-61.
14. Karush, F., *Heterogeneity of the Binding Sites of Bovine Serum Albumin.* Journal of the American Chemical Society, 1950. **72**(6): p. 2705-2713.
15. Thomas, P.D. and K.A. Dill, *An iterative method for extracting energy-like quantities from protein structures.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(21): p. 11628-11633.
16. Hernandez, M.A., J. Avila, and J.M. Andreu, *Physicochemical Characterization of the Heat-Stable Microtubule-Associated Protein Map2.* European Journal of Biochemistry, 1986. **154**(1): p. 41-48.
17. Sigler, P.B., *Transcriptional Activation - Acid Blobs and Negative Noodles.* Nature, 1988. **333**(6170): p. 210-212.
18. Gast, K., et al., *Prothymosin-Alpha - a Biologically-Active Protein with Random Coil Conformation.* Biochemistry, 1995. **34**(40): p. 13211-13218.
19. Kriwacki, R.W., et al., *Structural studies of p21(Waf1/Cip1/Sdi1) in the free and Cdk2-bound state: Conformational disorder mediates binding diversity.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(21): p. 11504-11509.
20. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* Journal of molecular biology, 1999. **293**(2): p. 321-31.
21. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-27

22. Dunker, A.K., et al., *Intrinsically disordered protein.* Journal of molecular graphics & modelling, 2001. **19**(1): p. 26-59.
23. Tompa, P., *Intrinsically unstructured proteins.* Trends in biochemical sciences, 2002. **27**(10): p. 527-33.
24. Sickmeier, M., et al., *DisProt: the Database of Disordered Proteins.* Nucleic acids research, 2007. **35**(Database issue): p. D786-93.
25. Tompa, P., *Intrinsically disordered proteins: a 10-year recap.* Trends in biochemical sciences, 2012. **37**(12): p. 509-516.
26. Uversky, V.N. and A.K. Dunker, *Understanding protein non-folding.* Biochimica et biophysica acta, 2010. **1804**(6): p. 1231-64.
27. Mittag, T. and J.D. Forman-Kay, *Atomic-level characterization of disordered protein ensembles.* Current Opinion in Structural Biology, 2007. **17**(1): p. 3-14.
28. Gsponer, J., et al., *Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation.* Science, 2008. **322**(5906): p. 1365-1368.
29. *DeLano Scientific LLC, macPyMOL.* 2006.
30. He, B., et al., *Predicting intrinsic disorder in proteins: an overview.* Cell research, 2009. **19**(8): p. 929-49.
31. Dosztanyi, Z., et al., *The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins.* Journal of molecular biology, 2005. **347**(4): p. 827-39.
32. Peng, K., et al., *Length-dependent prediction of protein intrinsic disorder.* BMC bioinformatics, 2006. **7**: p. 208.
33. Linding, R., et al., *Protein disorder prediction: implications for structural proteomics.* Structure, 2003. **11**(11): p. 1453-9.
34. Yang, Z.R., et al., *RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.* Bioinformatics, 2005. **21**(16): p. 3369-76.
35. McGuffin, *Intrinsic disorder prediction from the analysis of multiple protein fold recognition models.* Bioinformatics, 2008. **24**(16): p. 1798-1804.
36. Ishida, T. and K. Kinoshita, *Prediction of disordered regions in proteins based on the meta approach.* Bioinformatics, 2008. **24**(11): p. 1344-8.
37. Deng, X., J. Eickholt, and J. Cheng, *A comprehensive overview of computational protein disorder prediction methods.* Molecular bioSystems, 2012. **8**(1): p. 114-21.
38. Magidovich, E., S.J. Fleishman, and O. Yifrach, *Intrinsically disordered C-terminal segments of voltage-activated potassium channels: a possible fishing rod-like mechanism for channel binding to scaffold proteins.* Bioinformatics, 2006. **22**(13): p. 1546-1550.
39. Tompa, P. and P. Csermely, *The role of structural disorder in the function of RNA and protein chaperones.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2004. **18**(11): p. 1169-75.
40. Iakoucheva, L.M., et al., *Intrinsic disorder in cell-signaling and cancer-associated proteins.* Journal of molecular biology, 2002. **323**(3): p. 573-584.
41. Iakoucheva, L.M., et al., *The importance of intrinsic disorder for protein phosphorylation.* Nucleic acids research, 2004. **32**(3): p. 1037-49.
42. Zor, T., et al., *Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators.* The Journal of biological chemistry, 2002. **277**(44): p. 42241-8.
43. Tait, S., et al., *Local control of a disorder-order transition in 4E-BP1 underpins regulation of translation via eIF4E.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(41): p. 17627-32.
44. Andrew, C.D., et al., *Effect of phosphorylation on alpha-helix stability as a function of position.* Biochemistry, 2002. **41**(6): p. 1897-905.

45. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Current Opinion in Structural Biology, 2002. **12**(1): p. 54-60.
46. Mittag, T., L.E. Kay, and J.D. Forman-Kay, *Protein dynamics and conformational disorder in molecular recognition.* Journal of molecular recognition : JMR, 2010. **23**(2): p. 105-16.
47. Neduva, V. and R.B. Russell, *DILIMOT: discovery of linear motifs in proteins.* Nucleic acids research, 2006. **34**(Web Server issue): p. W350-5.
48. Mohan, A., et al., *Analysis of molecular recognition features (MoRFs).* Journal of molecular biology, 2006. **362**(5): p. 1043-59.
49. Onitsuka, M., et al., *Mechanism of induced folding: Both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants.* Proteins, 2008. **72**(3): p. 837-47.
50. Wright, P.E. and H.J. Dyson, *Linking folding and binding.* Current Opinion in Structural Biology, 2009. **19**(1): p. 31-8.
51. Sugase, K., H.J. Dyson, and P.E. Wright, *Mechanism of coupled folding and binding of an intrinsically disordered protein.* Nature, 2007. **447**(7147): p. 1021-5.
52. Lu, Q., H.P. Lu, and J. Wang, *Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics.* Physical review letters, 2007. **98**(12): p. 128105.
53. Receveur-Brechot, V., et al., *Assessing protein disorder and induced folding.* Proteins, 2006. **62**(1): p. 24-45.
54. Ganguly, D., W. Zhang, and J. Chen, *Synergistic folding of two intrinsically disordered proteins: searching for conformational selection.* Molecular bioSystems, 2012. **8**(1): p. 198-209.
55. Oldfield, C.J., et al., *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.* BMC genomics, 2008. **9 Suppl 1**: p. S1.
56. Tompa, P., C. Szasz, and L. Buday, *Structural disorder throws new light on moonlighting.* Trends in biochemical sciences, 2005. **30**(9): p. 484-9.
57. Tompa, P. and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions.* Trends in biochemical sciences, 2008. **33**(1): p. 2-8.
58. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nature reviews. Molecular cell biology, 2005. **6**(3): p. 197-208.
59. Silverthorn, *Human Physiology*. Third ed. 2004.
60. Savina, A. and S. Amigorena, *Phagocytosis and antigen presentation in dendritic cells.* Immunological reviews, 2007. **219**: p. 143-56.
61. Dunkelberger, J.R. and W.C. Song, *Complement and its role in innate and adaptive immune responses.* Cell research, 2010. **20**(1): p. 34-50.
62. Wu, J. and L.L. Lanier, *Natural killer cells and cancer.* Advances in cancer research, 2003. **90**: p. 127-56.
63. Janeway, C.A., Jr. , *Immunobiology the immune system in health and disease.* 6th ed. 2005: Garland Science Publishing
64. Packard, T.A. and J.C. Cambier, *B lymphocyte antigen receptor signaling: initiation, amplification, and regulation.* F1000prime reports, 2013. **5**: p. 40.
65. Puri, K.D., J.A. Di Paolo, and M.R. Gold, *B-cell receptor signaling inhibitors for treatment of autoimmune inflammatory diseases and B-cell malignancies.* International reviews of immunology, 2013. **32**(4): p. 397-427.
66. Yang, J. and M. Reth, *Oligomeric organization of the B-cell antigen receptor on resting cells.* Nature, 2010. **467**(7314): p. 465-9.
67. Ollila, J. and M. Vihinen, *B cells.* The international journal of biochemistry & cell biology, 2005. **37**(3): p. 518-23.

68. Yang, J. and M. Reth, *The dissociation activation model of B cell antigen receptor triggering.* FEBS letters, 2010. **584**(24): p. 4872-7.
69. Tolar, P., et al., *The molecular assembly and organization of signaling active B-cell receptor oligomers.* Immunological reviews, 2009. **232**(1): p. 34-41.
70. Sigalov, A.B., *Protein intrinsic disorder and oligomericity in cell signaling.* Molecular bioSystems, 2010. **6**(3): p. 451-61.
71. Minguet, S., E.P. Dopfer, and W.W. Schamel, *Low-valency, but not monovalent, antigens trigger the B-cell antigen receptor (BCR).* International immunology, 2010. **22**(3): p. 205-12.
72. Pierce, S.K. and W. Liu, *The tipping points in the initiation of B cell signalling: how small changes make big differences.* Nature reviews. Immunology, 2010. **10**(11): p. 767-77.
73. Xu, C., et al., *Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif.* Cell, 2008. **135**(4): p. 702-13.
74. Davis, R.E., et al., *Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma.* Nature, 2010. **463**(7277): p. 88-92.
75. Hu, K. and F. Chen, *Identification of significant pathways in gastric cancer based on protein-protein interaction networks and cluster analysis.* Genetics and molecular biology, 2012. **35**(3): p. 701-8.
76. Aivazian, D. and L.J. Stern, *Phosphorylation of T cell receptor zeta is regulated by a lipid dependent folding transition.* Nature structural biology, 2000. **7**(11): p. 1023-6.
77. van der Merwe, P.A. and O. Dushek, *Mechanisms for T cell receptor triggering.* Nature reviews. Immunology, 2011. **11**(1): p. 47-55.
78. Wucherpfennig, K.W., et al., *Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling.* Cold Spring Harbor perspectives in biology, 2010. **2**(4): p. a005140.
79. Kabouridis, P.S., *Lipid rafts in T cell receptor signalling.* Molecular membrane biology, 2006. **23**(1): p. 49-57.
80. Creighton, T., *Proteins Structures and Molecular properties*. second ed. 1993
81. Spirin, A., Swartz, J., *Cell-free Protein Synthesis*. 2008, Weinheim,: WILEY-VCH verlag GmbH & Co.
82. Nirenberg, M. and J.H. Matthaei, *Dependence of Cell-Free Protein Synthesis in E Coli Upon Naturally Occurring or Synthetic Polyribonucleotides.* Proceedings of the National Academy of Sciences of the United States of America, 1961. **47**(10): p. 1588-&.
83. Nirenberg, M.W., *Cell-Free Protein Synthesis Directed by Messenger Rna.* Methods in Enzymology, 1963. **6**: p. 17-23.
84. Devries, J.K. and G. Zubay, *DNA-Directed Peptide Synthesis .2. Synthesis of Alpha-Fragment of Enzyme Beta-Galactosidase.* Proceedings of the National Academy of Sciences of the United States of America, 1967. **57**(4): p. 1010-&.
85. Lederman, M. and G. Zubay, *DNA-Directed Peptide Synthesis .I. A Comparison of T2 and Escherichia Coli DNA-Directed Peptide Synthesis in 2 Cell-Free Systems.* Biochimica et biophysica acta, 1967. **149**(1): p. 253-&.
86. Spirin, A.S., et al., *A Continuous Cell-Free Translation System Capable of Producing Polypeptides in High-Yield.* Science, 1988. **242**(4882): p. 1162-1164.
87. Chekulayeva, M.N., et al., *Continuous-exchange cell-free protein-synthesizing system: Synthesis of HIV-1 antigen Nef.* Biochemical and biophysical research communications, 2001. **280**(3): p. 914-917.
88. Pedersen, A., et al., *Rational improvement of cell-free protein synthesis.* New biotechnology, 2011. **28**(3): p. 218-24.
89. He, M., He, Y., Wang. M, *From DNA to protein: No living cells required.* Process Biochemistry, 2011(46): p. 615-620.

90. Shimizu, Y., T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods, 2005. **36**(3): p. 299-304.
91. Matsuda. T., K.S., Tochio. N., Seki. E., Kigawa. T, *Improving cell-free protein synthesis for stable-isotope labeling.* J Biomol NMR, 2007(37): p. 225-229.
92. Schwarz, D., V. Dotsch, and F. Bernhard, *Production of membrane proteins using cell-free expression systems.* Proteomics, 2008. **8**(19): p. 3933-46.
93. Kim, D.M. and J.R. Swartz, *Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis.* Biotechnology and bioengineering, 2001. **74**(4): p. 309-16.
94. Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnology and bioengineering, 2004. **86**(1): p. 19-26.
95. Katzen, F., G. Chang, and W. Kudlicki, *The past, present and future of cell-free protein synthesis.* Trends in biotechnology, 2005. **23**(3): p. 150-6.
96. Kim, T.W., D.M. Kim, and C.Y. Choi, *Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system.* Journal of biotechnology, 2006. **124**(2): p. 373-80.
97. Isaksson, L., et al., *Highly efficient NMR assignment of intrinsically disordered proteins: application to B- and T cell receptor domains.* PloS one, 2013. **8**(5): p. e62947.
98. Kim, H.C. and D.M. Kim, *Methods for energizing cell-free protein synthesis.* Journal of bioscience and bioengineering, 2009. **108**(1): p. 1-4.
99. Klammt, C., et al., *Cell-free expression as an emerging technique for the large scale production of integral membrane protein.* The FEBS journal, 2006. **273**(18): p. 4141-53.
100. Hatfield, G.W. and D.A. Roth, *Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and Translation Engineering.* Biotechnology annual review, 2007. **13**: p. 27-42.
101. Allert, M., J.C. Cox, and H.W. Hellinga, *Multifactorial determinants of protein expression in prokaryotic open reading frames.* Journal of molecular biology, 2010. **402**(5): p. 905-18.
102. Kurotani, A., et al., *Comprehensive bioinformatics analysis of cell-free protein synthesis: identification of multiple protein properties that correlate with successful expression.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2010. **24**(4): p. 1095-104.
103. Koglin, A., et al., *Combination of cell-free expression and NMR spectroscopy as a new approach for structural investigation of membrane proteins.* Magnetic resonance in chemistry : MRC, 2006. **44 Spec No**: p. S17-23.
104. Ozawa, K., et al., *N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes.* The FEBS journal, 2006. **273**(18): p. 4154-9.
105. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins.* Science, 1989. **244**(4901): p. 182-8.
106. Cellitti, S.E., et al., *In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy.* Journal of the American Chemical Society, 2008. **130**(29): p. 9268-81.
107. Kainosho, M., et al., *Optimal isotope labelling for NMR protein structure determinations.* Nature, 2006. **440**(7080): p. 52-7.
108. Urh, M., D. Simpson, and K. Zhao, *Affinity chromatography: general methods.* Methods in Enzymology, 2009. **463**: p. 417-38.
109. Bornhorst, J.A. and J.J. Falke, *Purification of proteins using polyhistidine affinity tags.* Methods in Enzymology, 2000. **326**: p. 245-54.

110. Aguilar, M.I., *Reversed-phase high-performance liquid chromatography*. Methods in molecular biology, 2004. **251**: p. 9-22.
111. Williamson, M.P., T.F. Havel, and K. Wuthrich, *Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry*. Journal of molecular biology, 1985. **182**(2): p. 295-315.
112. Levitt, M.H., *Spin dynamics Basics of Nuclear Magnetic resonance*. 2002, Chichester: John Wiley & Sons.
113. Keeler, J., *Understanding NMR Spectroscopy*. 2010, Chichester: John Wiley & Sons Ltd.
114. Orekhov, V.Y. and V.A. Jaravine, *Analysis of non-uniformly sampled spectra with multi-dimensional decomposition*. Progress in nuclear magnetic resonance spectroscopy, 2011. **59**(3): p. 271-92.
115. Barna, J.C.J., Laue, E. D., Mayger, M. R., Skilling, J., Worrall, S. J. P., *Exponential sampling, an alternative method for sampling in two-dimensional MR experiments*. Journal of Magnetic Resonance, 1987. **73**: p. 69-77.
116. Kazimierczuk, K. and V.Y. Orekhov, *Accelerated NMR spectroscopy by using compressed sensing*. Angewandte Chemie, 2011. **50**(24): p. 5556-9.
117. Chylla, R.A. and J.L. Markley, *Theory and application of the maximum likelihood principle to NMR parameter estimation of multidimensional NMR data*. Journal of Biomolecular Nmr, 1995. **5**(3): p. 245-58.
118. Gutmanas, A., et al., *Three-way decomposition of a complete 3D 15N-NOESY-HSQC*. Journal of Biomolecular Nmr, 2002. **24**(3): p. 191-201.
119. Hiller, S., et al., *Coupled decomposition of four-dimensional NOESY spectra*. Journal of the American Chemical Society, 2009. **131**(36): p. 12970-8.
120. Korzhnev, D.M., et al., *Backbone dynamics of the channel-forming antibiotic zervamicin IIB studied by 15N NMR relaxation*. FEBS letters, 2001. **495**(1-2): p. 52-5.
121. Orekhov, V.Y., I. Ibraghimov, and M. Billeter, *Optimizing resolution in multidimensional NMR by three-way decomposition*. Journal of Biomolecular Nmr, 2003. **27**(2): p. 165-73.
122. Orekhov, V.Y., I.V. Ibraghimov, and M. Billeter, *MUNIN: a new approach to multi-dimensional NMR spectra interpretation*. Journal of Biomolecular Nmr, 2001. **20**(1): p. 49-60.
123. Wong, L.E., et al., *Automatic assignment of protein backbone resonances by direct spectrum inspection in targeted acquisition of NMR data*. Journal of Biomolecular Nmr, 2008. **42**(2): p. 77-86.
124. Rule. G. S., H.K.T., *Fundamentals of Protein NMR Spectroscopy*. 2006, Dordrecht: Springer.
125. Zimmerman, D.E., et al., *Automated analysis of protein NMR assignments using methods from artificial intelligence*. Journal of molecular biology, 1997. **269**(4): p. 592-610.
126. Jung, Y.S. and M. Zweckstetter, *Mars -- robust automatic backbone assignment of proteins*. Journal of Biomolecular Nmr, 2004. **30**(1): p. 11-23.
127. Bartels, C., et al., *Automated sequence-specific NMR assignment of homologous proteins using the program GARANT*. Journal of Biomolecular Nmr, 1996. **7**(3): p. 207-13.
128. Solyom, Z., et al., *BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins*. J Biomol NMR, 2013. **55**(4): p. 311-21.
129. Jaravine, V.A., et al., *Hyperdimensional NMR spectroscopy with nonlinear sampling*. Journal of the American Chemical Society, 2008. **130**(12): p. 3927-36.
130. Selenko, P. and G. Wagner, *Looking into live cells with in-cell NMR spectroscopy*. Journal of structural biology, 2007. **158**(2): p. 244-53.
131. Kosol, S., et al., *Structural characterization of intrinsically disordered proteins by NMR spectroscopy*. Molecules, 2013. **18**(9): p. 10802-28.

132. Dyson, H.J. and P.E. Wright, *Unfolded proteins and protein folding studied by NMR.* Chemical reviews, 2004. **104**(8): p. 3607-22.
133. Kjaergaard, M. and F.M. Poulsen, *Disordered proteins studied by chemical shifts.* Progress in nuclear magnetic resonance spectroscopy, 2012. **60**: p. 42-51.
134. Gillespie, J.R. and D. Shortle, *Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures.* Journal of molecular biology, 1997. **268**(1): p. 170-84.
135. Gillespie, J.R. and D. Shortle, *Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels.* Journal of molecular biology, 1997. **268**(1): p. 158-69.
136. Bermel, W., et al., *Speeding up sequence specific assignment of IDPs.* Journal of Biomolecular Nmr, 2012. **53**(4): p. 293-301.
137. Bermel, W., et al., *H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins.* Journal of Magnetic Resonance, 2009. **198**(2): p. 275-81.
138. Schanda, P., E. Kupce, and B. Brutscher, *SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds.* Journal of Biomolecular Nmr, 2005. **33**(4): p. 199-211.
139. Theillet, F.X., et al., *Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts.* Nature protocols, 2013. **8**(7): p. 1416-32.
140. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2005. **1751**(2): p. 119-139.
141. Savage, D.F., et al., *Cell-free complements in vivo expression of the E. coli membrane proteome.* Protein science : a publication of the Protein Society, 2007. **16**(5): p. 966-76.
142. Seddon, A.M., P. Curnow, and P.J. Booth, *Membrane proteins, lipids and detergents: not just a soap opera.* Biochimica et biophysica acta, 2004. **1666**(1-2): p. 105-17.
143. Ferndahl, C., et al., *Increasing cell biomass in Saccharomyces cerevisiae increases recombinant protein yield: the use of a respiratory strain as a microbial cell factory.* Microbial cell factories, 2010. **9**: p. 47.
144. Hedfalk, K., et al., *A regulatory domain in the C-terminal extension of the yeast glycerol channel Fps1p.* The Journal of biological chemistry, 2004. **279**(15): p. 14954-60.
145. Bonander, N., et al., *Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield.* Microbial cell factories, 2009. **8**: p. 10.
146. Oberg, F., et al., *Insight into factors directing high production of eukaryotic membrane proteins; production of 13 human AQPs in Pichia pastoris.* Molecular membrane biology, 2009. **26**(4): p. 215-27.
147. Tautermann, C.S., *GPCR structures in drug design, emerging opportunities with new structures.* Bioorganic & medicinal chemistry letters, 2014.
148. Sapienza, P.J. and A.L. Lee, *Using NMR to study fast dynamics in proteins: methods and applications.* Current opinion in pharmacology, 2010. **10**(6): p. 723-30.
149. Nourse, A. and T. Mittag, *The cytoplasmic domain of the T-cell receptor zeta subunit does not form disordered dimers.* Journal of molecular biology, 2014. **426**(1): p. 62-70.
150. Sigalov, A.B., A.V. Zhuravleva, and V.Y. Orekhov, *Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form.* Biochimie, 2007. **89**(3): p. 419-21.
151. Lee, W.Y. and P. Tolar, *Activation of the B cell receptor leads to increased membrane proximity of the Igalpha cytoplasmic domain.* PloS one, 2013. **8**(11): p. e79148.

152. Aghazadeh, B., et al., *Structural basis for relief of autoinhibition of the Dbl homology domain of proto-oncogene Vav by tyrosine phosphorylation.* Cell, 2000. **102**(5): p. 625-33.
153. Fujitani, N., et al., *Structure determination and conformational change induced by tyrosine phosphorylation of the N-terminal domain of the alpha-chain of pig gastric H+/K+-ATPase.* Biochemical and biophysical research communications, 2003. **300**(1): p. 223-9.
154. Futterer, K., et al., *Structural basis for Syk tyrosine kinase ubiquity in signal transduction pathways revealed by the crystal structure of its regulatory SH2 domains bound to a dually phosphorylated ITAM peptide.* Journal of molecular biology, 1998. **281**(3): p. 523-37.
155. Clark, M.R., et al., *The B cell antigen receptor complex: association of Ig-alpha and Ig-beta with distinct cytoplasmic effectors.* Science, 1992. **258**(5079): p. 123-6.