Cellular solid-state NMR on large prokaryotic and eukaryotic membrane protein complexes

Cellulaire vaste stof NMR op grote prokaryotische en eukaryotische membraaneiwitcomplexen

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 21 december 2015 des middags te 2.30 uur

door

Mohammed Kaplan

geboren op 23 oktober 1983 te Mosul, Irak

PROMOTOR:
Prof. dr. M. Baldus

‘There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.’

W. Shakespeare, Hamlet

CONTENTS

CHAPTER ONE

General Introduction

1.1 A brief history of Nuclear Magnetic resonance:

In the last decades, Nuclear Magnetic Resonance (NMR) spectroscopy has become an indispensible tool in chemistry, biology and medicine (where it is also used in the context of Magnetic Resonance Imaging, "MRI"). NMR has a long history dating back to the beginnings of the twentieth century when the quantum-mechanical theory was still being developed. Stern and Gerlach showed experimentally the quantization of electronic angular momentum[1]. A few years later, Uhlenbeck and Goudsmit suggested that the electron also possesses an intrinsic angular momentum and magnetic momentum, the so-called electron spin. To account for the doubling in the optical spectra lines, Pauli and Darwin introduced the idea of a nuclear spin (and the corresponding hyperfine coupling between electron and nuclear spin). After a refinement of their first experiment Stern and Gerlach proved the existence of nuclear spin experimentally (1933). The first to detect an NMR signal was I. Rabi in 1937[2], Also, the Dutch physicist Gorter was close to detect NMR but he was unfortunate to choose LiF which has a very long relaxation time (for this reason, he is sometimes referred to as "*the man who almost discovered NMR*"). Interestingly, K. Zavoisky recorded ^{1}H- NMR signal from water in 1941 (in Kazan, Russia) but he discarded the result for not being reproducible. Later he discovered the electron spin resonance (ESR)[3]. In the 1940's, Bloch and Purcell succeeded in obtaining NMR signals from bulk samples, namely, water and paraffin[4,5]. For their work, they were awarded the Nobel Prize in physics in 1952.

In the last 2 decades, solution-state NMR has been successfully implemented to study at atomic resolution molecules tumbling freely in solution. In spite of considerable technical progress, such studies are intrinsically limited because larger molecules tumble slowly leading to broader lines. Such a size limitation does not exist in case of solid-state NMR (ssNMR), which makes ssNMR the method of choice to study large systems that tumble slowly or are fully immobilized on the NMR time scale. Over the years ssNMR has proven to be a powerful technique in various areas of research including material, pharmaceutical or food science. In addition, complex biomolecular systems such as Amyloid fibrils or biomaterials have been studied. Also, ssNMR methods have been developed to investigate membrane proteins reconstituted in lipid bilayers and even in native membranes. In this thesis, we report on progress to further extend ssNMR to investigate large bacterial molecular complexes in their native cellular environment. Moreover, we present an ssNMR-based

approach to study large eukaryotic membrane proteins in their cellular settings. In these cases, we also make use of the latest developments in the field of hyperpolarized Magnetic Resonance, i.e., high-field Dynamic Nuclear Polarization (DNP)

1.2 Solid-state NMR:

As its name suggests, solid-state NMR (ssNMR) is used to investigate (bio) molecules that remain globally immobilized on the time scale of the NMR experiment. Such preparations can range from microcrystals, fibrils or hydrogels to synthetic lipid bilayers and extend to proteins in their native cellular membranes and even whole cells. However, a major problem in (ss)NMR is the inherent low sensitivity of the technique itself. Due to the small energy gap between the two spin states (in case of spin ½ nuclei), there is only a small difference in population between the low (n lower) and the high (n upper) energy states as stated by Boltzmann's equation:

$$\frac{n_{upper}}{n_{lower}} = e^{-\Delta E/kT}$$

Where ΔE is the energy difference (equals to $h\gamma B$, with h being plank's constant, γ the gyromagnetic ratio and B the external magnetic field), k is Boltzmann's constant and T is the temperature.

A major step to overcome the problem of sensitivity in the ssNMR was the introduction of cross-polarization (CP)[6] (**Fig. 1a**). During CP, magnetization from nuclei with a high gyromagnetic ratio (mostly protons, i.e., ^{1}H) is transferred to less abundant nuclei with a lower gyromagnetic ratio (including ^{13}C or ^{15}N). Another important means to improve sensitivity is the use of Dynamic Nuclear Polarization (DNP) (which will be discussed in point **1.3**) and proton (^{1}H) detection. Proton detection, which was a critically step to establish high sensitivity in solution-state NMR, has been hampered in case of ssNMR by the presence of strong dipolar ^{1}H-^{1}H couplings. However, recent developments in terms of sample labeling and instrumental hardware have greatly improved the potential of proton detected experiments in ssNMR[7–9].

In addition to the low sensitivity associated with NMR, the limited molecular motion of molecules in their solid state results in interactions that are usually averaged out in case of freely tumbling molecules. As a result, anisotropic line shapes (also called powder line shapes) are obtained that yield the low-resolution spectra in ssNMR. The two major interactions that are dominant in ssNMR (excluding the quadrupolar interactions) are homo – and heteronuclear dipole-dipole interactions (DD) and the chemical shift anisotropy (CSA). For example, the heteronuclear dipolar interaction between two spins can be represented by the Hamiltonan operator:

$$H_{IS\ (hetero)} = -d(3\cos^2\theta - 1)I_z S_z$$

Where θ is the angle between the external magnetic field and the vector connecting the two spins, and d represents the dipolar coupling constant:

$$d = \left(\frac{\mu_0}{4\pi}\right) \frac{\hbar \gamma_I \gamma_S}{r_{IS}^3}$$

Because of the orientation-dependence, static ssNMR spectra are orientation dependent that are mathematically described by terms such as $(3\cos^2\theta - 1)$ term in the equation above. In addition, u_0 is the permeability of free space, r is the distance between the two spins and γ represents the gyromagnetic ratios of I and S spins. Note that the strongest dipolar coupling exists between protons because of the highest gyromagnetic ratios, for example compared to ^{13}C or ^{15}N nuclei.

An important milestone in the history of ssNMR was the introduction of Magic angle spinning (MAS)[10,11] (**Fig. 1b**). It can be shown mathematically, that spinning of the ssNMR sample at an angle of 54.7° strongly reduces the influence of anisotropic interactions such as dipolar or CSA interactions resulting in high-resolution spectra (**Fig. 1b & Fig. 2**). The combination of CP and MAS ssNMR (a.k.a. CPMAS) was a very important milestone in introducing high-resolution ssNMR applications in life and material science.

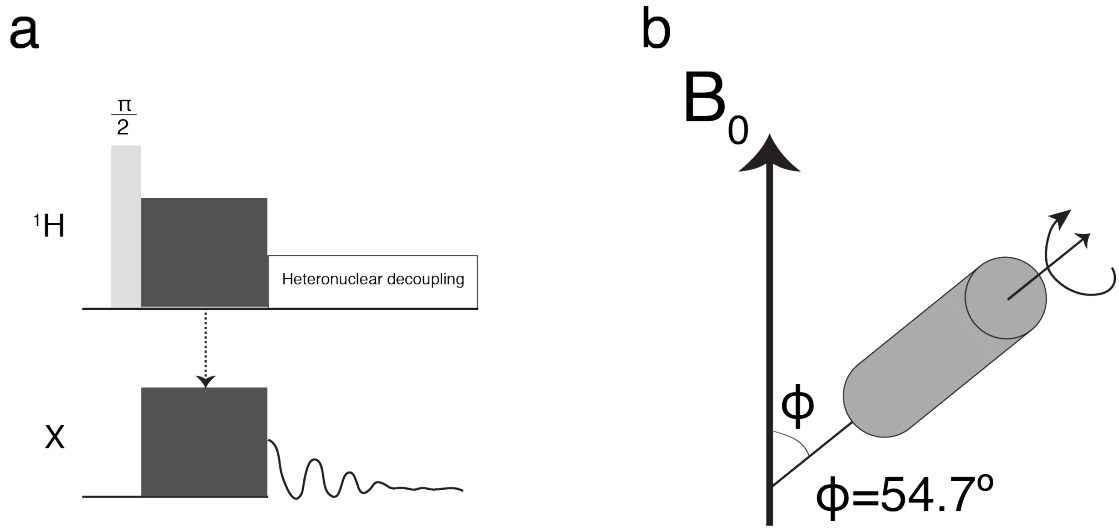


Fig. 1: a) Cross-polarization experiment, which transfers magnetization from protons (high gyromagnetic ratio) to other nuclei like ^{15}N and ^{13}C. Note that the same experiment can also be used to transfer polarization from a spin X to another spin species Y. b) Magic Angle Spinning (MAS) of the ssNMR rotor containing the sample of interest helps to reduce the anisotropic interactions leading to higher resolution spectra.

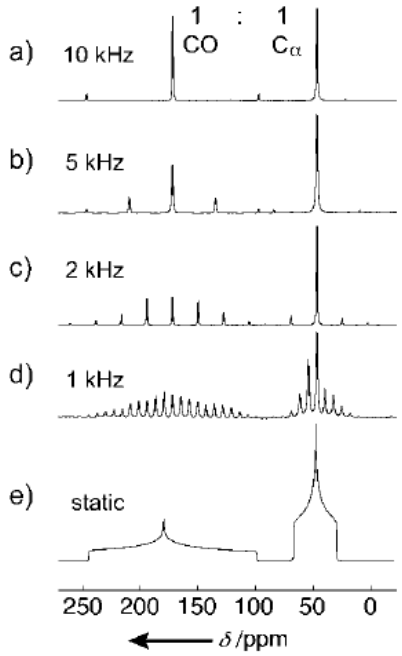


Fig. 2: MAS at increasing spinning speeds improves the spectral resolution. Note that at intermediate spinning speeds, the so-called spinning sidebands that are separated in frequency by the MAS rate (in Hz) appear in the spectrum. Figure adapted from ref [12].

1.3 Dynamic Nuclear Polarization (DNP):

Dynamic Nuclear Polarization (DNP), which was first suggested by Overhauser[13] and then experimentally proven by Carver and Slichter[14] in the 1950's, involves the transfer of

magnetization from the electrons (which have a gyromagnetic ratio 660 times higher than that of a proton) to the nuclei. At present, the following two mechanisms are most often used to transfer magnetization from an electron to a nucleus, these being: 1- The solid effect: which is an interaction between one electron and a nucleus, and 2- The cross effect which involves the interaction between two electrons and a nucleus (See ref [15]). The combination of DNP with MAS and low (approx. 100 K) temperatures (also called LT-DNP MAS) usually involves the use of biradicals (**Fig. 3**) that currently provide the most efficient signal enhancement under cross-effect conditions.

For practical applications, the sample is firstly washed with an appropriate buffer containing the (bi)radical[16] that introduces the electrons to the system (**Fig. 3**) and a cryoprotectant (usually glycerol-d8). Thereafter, the sample is irradiated with microwaves produced by a gyrotron in order to polarize the electrons, then this polarization is transferred to the nucleus by one of the previously mentioned two mechanisms. (**Fig. 4**)

In theory an enhancement of 660 times should be obtained. However, in practice much lower enhancement factors are obtained depending on the nature of the sample and the radical that is used.

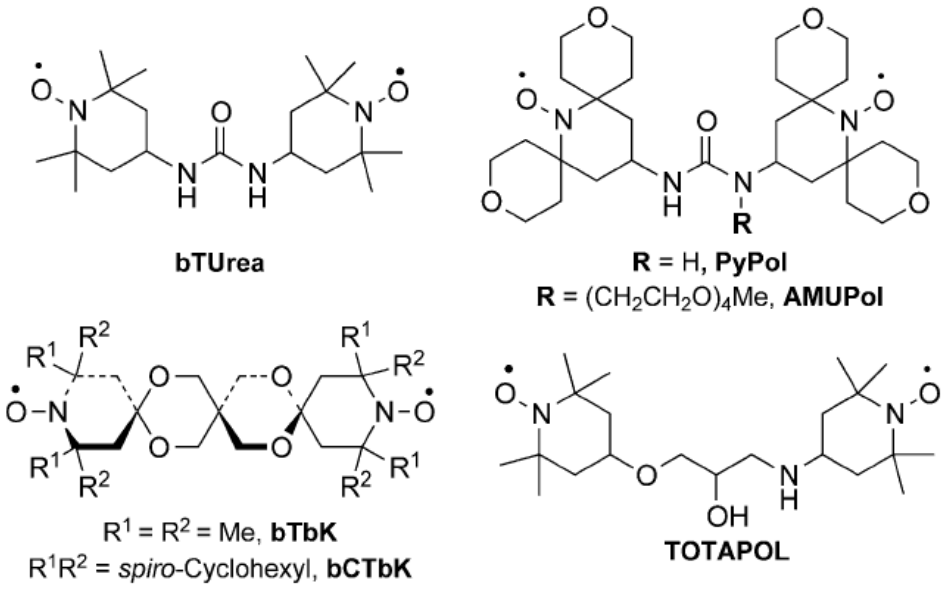


Fig. 3: An example of different biradicals used for DNP. In this thesis, AMUPol was used for cellular DNP measurements. Figure is adapted from ref.[16]

DNP has been applied successfully to investigate a wide variety of systems, like small silica nanoparticles[17] and nanoporous silica material[18], small peptides[19] and pharmaceutical formulations[20], small proteins in native membranes[21] even whole cells[22,23]. Here, we exploited DNP to investigate the bacterial Type IV secretion system core complex, a one-megadalton complex, in its native cellular environment[24]. Moreover, we used DNP, for the first time, to scrutinize a large eukaryotic membrane protein, namely, epidermal growth factor receptor (EGFR), in its native settings.

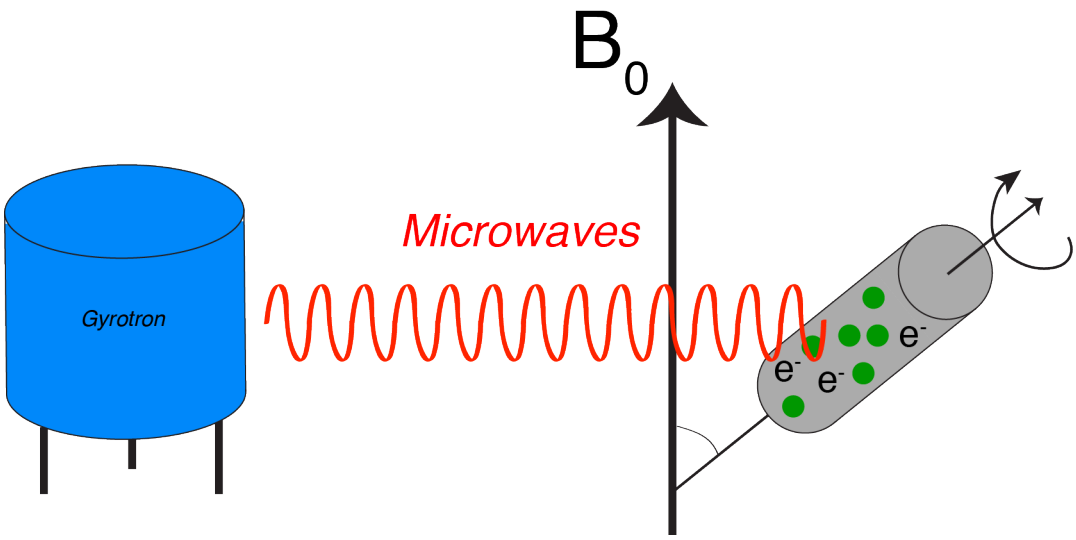


Fig. 4: An illustration of DNP setup. The sample (green circles) is at the MAS in the ssNMR with the DNP juice that contains the radicals (e^-). A gyrotron will provide microwaves continuously during the experiment, which leads to polarization of the electrons, which in turn will be transferred to the nuclei in the sample.

1.4 Cellular DNP-supported solid-state NMR:

The ultimate aim of structural biology is to be able to investigate the structure and dynamics of biomolecules at atomic resolution in their native environment. The influence of the natural environment on the protein structure and function has been demonstrated for many proteins (e.g., for T4SScc[25] and for EGFR see the difference between ref.[26] and ref.[27]). Liquid-state NMR has been successful to investigate small proteins inside their cellular environment (in-cell NMR[28–31]) . Recently, solid-state NMR has been successfully implemented to study small membrane proteins in their native cellular environment[32–35] . Moreover, in our group of

ssNMR in Utrecht, a 150 amino acid protein, namely, PagL, was also studied in the bacterial cell envelope and whole cells[22,36] (**Fig. 5**).

Obtaining atomic-level information has, hitherto, required data acquired *in-vitro* using different techniques like crystallography, NMR or electron microscopy. Cellular ssNMR can be used to examine the validity of the models obtained by these techniques *in-situ* **(Fig. 6)**. In addition to this, it can be used to obtain high-resolution new structural data and insight into the dynamics of the investigated system while still in its native environment.

In this thesis, we extend the application of cellular DNP-supported solid-state NMR to study large membrane protein complexes that span the whole bacterial cell envelope (bacterial type IV secretion system core complex, referred to as T4SScc) in its native environment. Our approach allowed us to confirm an already existing crystal structure of half of the complex in cellular settings. Moreover, we could obtain new data about the structure and dynamics of the other part of the complex, which has thus far remained elusive for high-resolution studies. For the first time, we could also extend cellular ssNMR to scrutinize the dynamics and structure of a large eukaryotic membrane protein in its native cellular environment (epidermal growth factor receptor).

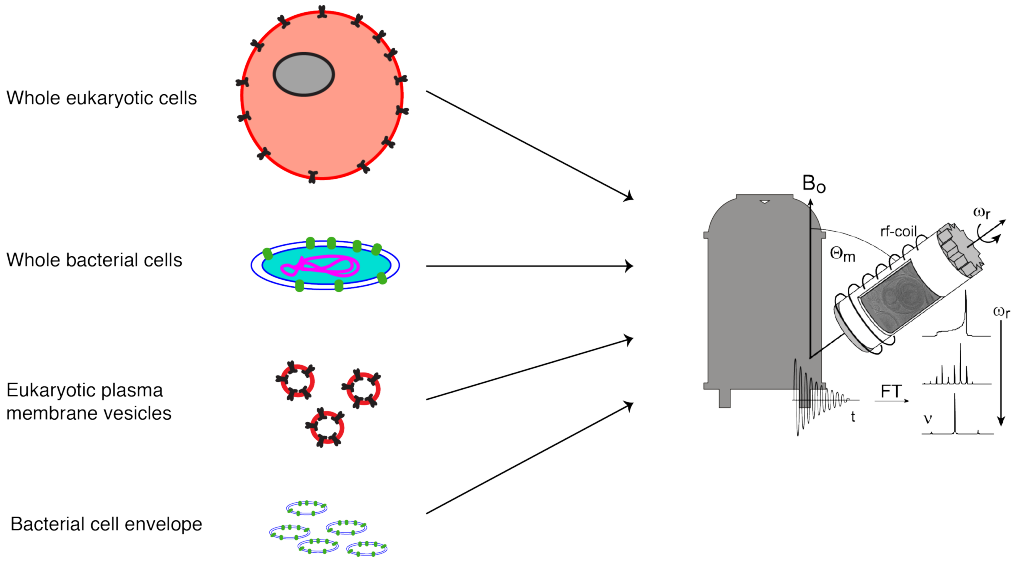


Fig. 5: In cellular ssNMR, proteins (both prokaryotic and eukaryotic) are studied in their native cellular environment. This ranges from whole cells (bacterial or eukaryotic) to isolated membrane vesicles and cell envelopes.

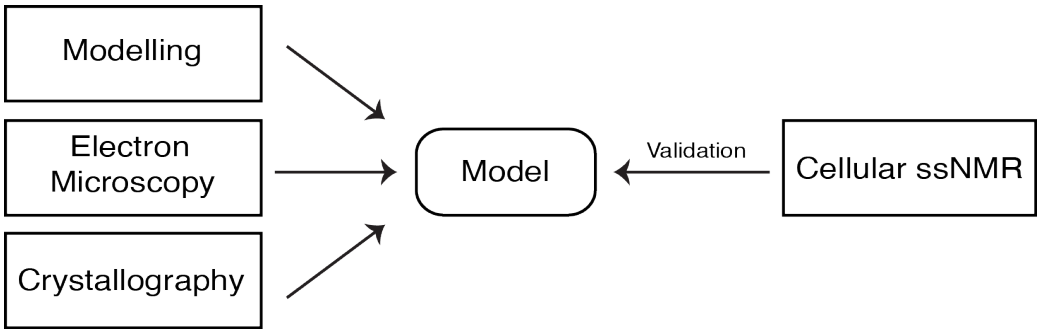


Fig. 6: Schematic representation of how cellular ssNMR can be used to validate models obtained from different *in-vitro* methods. (adapted from ref.[24]).

1.5 Biological systems studied in this thesis

1.5.1 Braun's Lipoprotein (Lpp):

Braun's Lipoprotein (Lpp) is a small protein first described by Braun and co-workers[37] . A crystal structure of Lpp has been described showing the protein adapting a helical conformation[38] (**Fig. 7a**). Lpp consists of 58 amino acids and its sequence lacks histidine, glycine, tryptophan and proline. It is attached to the peptidoglycan via the ε-NH_2 group of its C-terminal lysine and anchored to the outer membrane through its lipidated N-terminal cysteine (N-acyl diglyceridecysteine, **Fig. 7b**). In total, there are ~ 7.5 x 10^5 copies of Lpp per cell. It is believed that only one third of Lpp is attached to the peptidoglycan (see ref.[39] and references therein). Recently, it has been suggested that Lpp can be a receptor for cationic antimicrobial peptides[40] .Lpp mutant cells can still grow, albeit with some deformities in the outer membrane indicating that Lpp plays a pivotal role in maintaining the stability of bacterial outer membrane[41,42]. Due it its high abundance in the bacterial cell envelope, Lpp is of special importance for cellular ssNMR studies, which includes *E.coli* membranes. In the first studies of cellular ssNMR performed on PagL, an outer membrane bacterial protein, it was shown that the spectrum of non-induced cells was dominated by Lpp signals[36] . Our results here of investigating T4SScc in bacterial cell envelope indicate also some contribution of Lpp in our spectra. However, optimized labeling schemes can be used to eliminate Lpp signals from the spectra. Finally, being the most abundant endogenous protein in the *E.coli* cell envelope makes Lpp a very suitable model system for method development that involves cellular preparation.

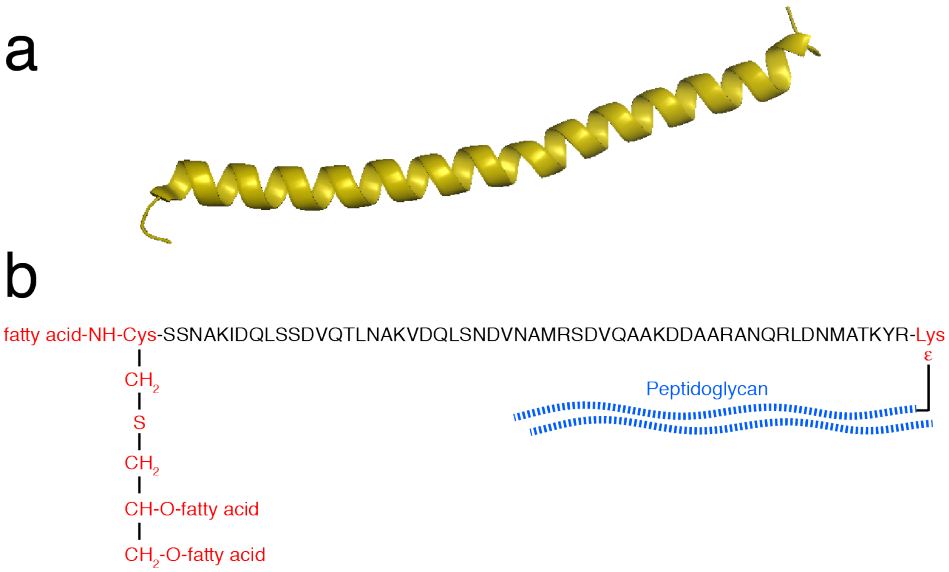


Fig. 7: a) Crystal structure of Braun's lipoprotein (PDB=1EQ7). b) Amino-acid sequence of Lpp showing the lipidated N-terminal cysteine that is anchored to the inner leaflet of the outer membrane. In case of the bound Lpp, the C-terminal lysine is attached to the peptidoglycan.

1.5.2 Type IV secretion system core complex (T4SScc):

Secretion systems are very important for the survival of bacteria. Type IV secretion system (T4SS) is a very versatile bacterial secretion system which is of great medical relevance and responsible for the secretion of wide range of proteins and DNA. A group of T4SScc is involved in the transfer of DNA between bacteria during bacterial conjugation, which is a pivotal step for the spread of antibiotic resistance (see review[43]). Additionally, T4SS is used by different kinds of bacteria like *Helicobacter pylori* and *Legionella pneumophila* to transfer protein effectors into eukaryotic cells[44,45] . Recently, it has also been suggested to use T4SS in DNA-based therapy [46]. T4SS consists of 12 proteins, namely, from VirB1 to VirB11 and VirD4 with VirB7, VirB9 and VirB10 forming the so-called T4SS core complex (T4SScc) which is a one megadalton tetradecamer complex that spans the whole bacterial cell envelope (**Fig. 8**). T4SScc functions to scaffold other proteins of T4SS and is divided into two parts, namely: the outer layer complex (OL) (which contains the c-terminal parts of VirB9 and VirB10 with the full length VirB7), and the inner layer (IL) complex (which consists of the N-terminal parts of VirB9 and VirB10)[43,47]. High-resolution information is available for the OM complex[48] (through crystallography, **Fig. 9**); however, an electron microscopy map is available for the whole T4SScc at two different resolutions[49,50]. In this thesis, we used DNP-

supported ssNMR to investigate T4SScc in bacterial cell envelopes. In our work here, we could confirm the presence of the crystal fold of some parts of the OM complex, moreover, we obtained new information about the structure and dynamics of the hitherto structurally elusive part of T4SScc IL complex.

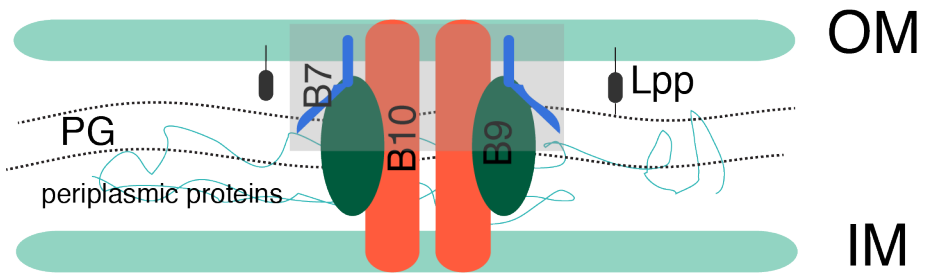


Fig. 8: Cell envelope of gram-negative bacteria showing the outer membrane (OM), the inner membrane (IM) and the peptidoglycan (PG). T4SScc spans the whole cell envelope and consists of B10 (orange), B9 (green) and B7 (blue), with the position of the OL complex indicated in the grey box. Lpp, the most abundant protein in the bacterial cell envelope is also highlighted with the periplasmic proteins.

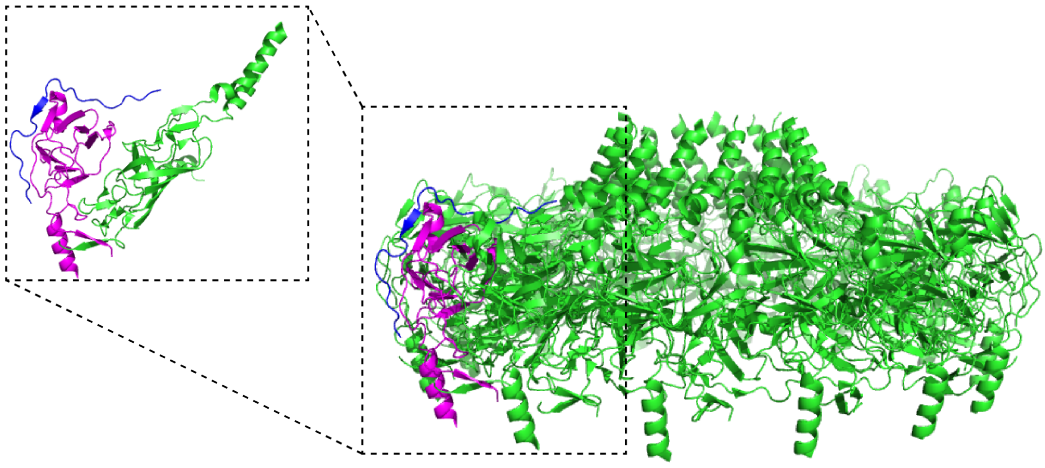


Fig. 9: Crystal structure of the outer layer (OL) T4SScc (PDB=3JQO). The OL part of T4SScc is a tetradecamer that consists of the C-termini of B10 and B9 and whole B7. The black dashed box is a zoom-in of a monomer subunit showing B10 (green), B9 (magenta) and B7 (blue).

1.5.3 Epidermal Growth Factor receptor (EGFR):

The epidermal growth factor receptor (EGFR), also known as Her1 or ErbB1, is a member of the Her (ErbB) family of receptor tyrosine kinase. EGFR is ~ 170 KDa protein that consists

of an extracellular part (domains I-IV), a transmembrane helix, a small juxtamembrane part, a kinase domain and a c-terminal tail (**Fig. 10**). EGFR has been linked with the development of many types of cancers[51] . Upon binding to its ligand (EGF), EGFR is dimerized, phosphorylated and clustered followed by a cascade of events ending with cell proliferation and division[52]. Thus far, crystal and NMR structures have been described for the different domains of EGFR[53–56]. In spite of its importance, high-resolution studies on the full-length receptor in its native environment are still lacking.

Here, we investigated EGFR in its native environment using cellular DNP-supported ssNMR. By labeling and isolating plasma membrane vesicles from A431 cells, we performed ssNMR experiments on the receptor in its native environment which yielded new insights into the dynamics of EGFR. Using specifically labeled EGFR samples, where only certain nuclei are the so-called "NMR active", we could furthermore introduce atomic probes located in different domains of the receptor allowing us to investigate the local dynamics and structure of the protein. The method presented here can be used to investigate other eukaryotic proteins in their native environment using ssNMR/DNP.

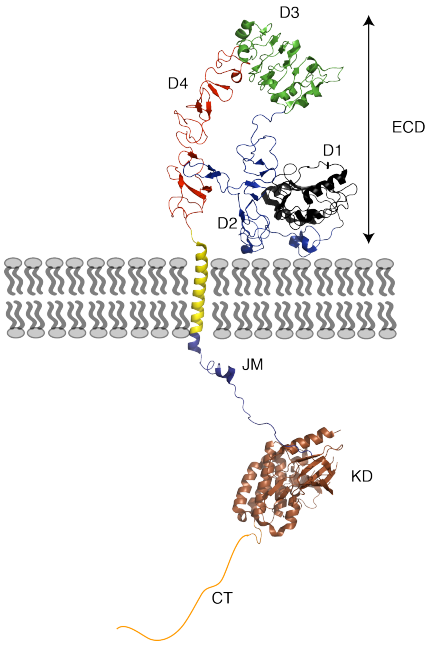


Fig. 10: EGFR consists of an extracellular domain (ECD), which in turn is divided into four domains (D1-D4), a transmembrane domain, a juxtamembrane domain (JM), a kinase domain (KD) and a C-terminal tail (CT). This figure was made by combining the available crystal structures of the ECD and KD, and the NMR structure of the TM and JM.

Scope of this thesis: In **chapter 1** of this thesis, a general introduction about NMR and DNP is given. The idea of cellular ssNMR is also introduced in this chapter. Finally, the biological systems studied in this thesis are described.

In chapter 2: we show that we could express T4SScc in a bacterial strain lacking the presence of OmpF/A in its cell envelope, which made T4SScc the most abundant protein in the isolated cell envelope. We could also produce different specifically labeled samples which allowed us to confirm parts of an already existing crystal structure of the upper half of the protein. Moreover, we could obtain new structural data and insight into the dynamics of the structurally elusive part of the machinery, the so-called inner layer (IL) complex. All these investigations were performed while T4SScc is in the bacterial cell envelope.

In chapter 3: we apply the idea of cellular ssNMR/DNP to investigate the epidermal growth factor receptor (EGFR) in membrane vesicles isolated from human cancer cells (A431). Investigating the fully [^{13}C, ^{15}N] labeled sample shed light on the global dynamics of the protein. Additionally, by making specifically [^{13}C, ^{15}N] labeled samples, we could obtain information on the local dynamics and structure of the protein while it is in the isolated membrane vesicles.

In chapter 4: the combination of proton detection and cellular ssNMR is presented in this chapter. Fractionally deuterated and perdeuterated *E.coli* cell envelopes expressing T4SScc were prepared and proton detected experiments on the fractionally deuterated sample were performed. The preliminary results speak in favor of a folded T4SScc which shows the feasibility of such approach.

In chapter 5: we present general protocols of how to prepare membrane protein samples eligible for ssNMR studies, both for purified proteins and for proteins in cellular environment.

In chapter 6: a general discussion of the data of the previous chapters is presented with future perspectives taking into account the new developments in terms of sample preparation, ssNMR instrumentation (the installation of 950 MHz in Utrecht), and the development of new DNP radicals with possible suggestions for future research.

Finally, a short summary of the thesis is presented.

References:

1. Stern, O. D a s m a g n e t i s c h e M o m e n t des Silberatoms . v-- ~ V 4kT. *Zeitschrift für Phys.* **9,** 353–355 (1922).

2. Rabi., I. I. A New Method of Measuring. *Phys. Rev.* **53,** 318 (1937).

3. Pfeifer, H. A short history of nuclear magnetic resonance spectroscopy and of its early years in Germany. *Magn. Reson. Chem.* **159,** 154–159 (1999).

4. Bloch, F. Nuclear Induction. *Phys. Rev.* **70,** 460 (1946).

5. Purcell, E. M. Resonance Absorption by Nuclear Magnetic Moments in a Solid. 37 (1946).

6. A. Pines, M. G. Gibby, Waugh, J. S. Proton-enhanced Nuclear Induction Spectroscopy. *Chem. Phys. Lett.* **15,** 373–376 (1972).

7. Reif, B., Jaroniec, C. P., Rienstra, C. M., Hohwy, M. & Griffin, R. G. 1H-1H MAS correlation spectroscopy and distance measurements in a deuterated peptide. *J. Magn. Reson.* **151,** 320–7 (2001).

8. Chevelkov, V., Rehbein, K., Diehl, A. & Reif, B. Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew. Chem. Int. Ed. Engl.* **45,** 3878–81 (2006).

9. Sinnige, T., Daniëls, M., Baldus, M. & Weingarth, M. Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J. Am. Chem. Soc.* **136,** 4452–5 (2014).

10. ANDREW, E. R., BRADBURY, A. & EADES, R. G. Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed. *Nature* **182,** 1659 (1958).

11. Lowe, I. J. Free Induction Decays of Rotating Solids. *Phys. Rev. Lett.* **2,** 285–287 (1959).

12. Laws, D. D., Bitter, H.-M. L. & Jerschow, A. Solid-state NMR spectroscopic methods in chemistry. *Angew. Chem. Int. Ed. Engl.* **41,** 3096–3129 (2002).

13. Overhauser, A. W. Polarization of Nuclei in Metals. *Phys. Rev.* **92,** 411–415 (1953).

14. Slichter, T. R. C. and C. P. Polarization of Nucler Spins in Metals. *Phys. Rev.* **92,** 212–213 (1953).

15. Ni, Q. Z. *et al.* High frequency dynamic nuclear polarization. *Acc. Chem. Res.* **46,** 1933–41 (2013).

16. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chem. Int. Ed. Engl.* **52,** 10858–10861 (2013).

17. Akbey, Ü. *et al.* Dynamic nuclear polarization of spherical nanoparticles. *Phys. Chem. Chem. Phys.* **15,** 20706–16 (2013).

18. Lesage, A. *et al.* Surface Enhanced NMR Spectroscopy by Dynamic Nuclear Polarization. 15459–15461 (2010).

19. Koers, E. J. *et al.* Dynamic nuclear polarization NMR spectroscopy: revealing multiple conformations in lipid-anchored peptide vaccines. *Angew. Chem. Int. Ed. Engl.* **52,** 10905–8 (2013).

20. Rossini, A. J. *et al.* Dynamic Nuclear Polarization Enhanced NMR Spectroscopy for Pharmaceutical Formulations. (2014).

21. Specker, E., Rossum, B. Van, Linden, A. H., Lange, S. & Franks, W. T. Neurotoxin II Bound to Acetylcholine Receptors in Native Membranes. 19266–19269 (2011).

22. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51,** 2998–3001 (2012).

23. Takahashi, H. *et al.* Solid-state NMR on bacterial cells: selective cell wall signal enhancement and resolution improvement using dynamic nuclear polarization. *J. Am. Chem. Soc.* **135,** 5105–10 (2013).

24. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12,** 5–9 (2015).

25. Low, H. H. *et al.* Structure of a type IV secretion system. *Nature* 1–18 (2014). doi:10.1038/nature13081

26. Wang, Z. *et al.* Mechanistic insights into the activation of oncogenic forms of EGF receptor. *Nat. Struct. Mol. Biol.* **18,** 1388–93 (2011).

27. Okimoto, R. A. *et al.* Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **350,** 2129–2139 (2004).

28. Banci, L., Barbieri, L., Luchinat, E. & Secci, E. Visualization of redox-controlled protein fold in living cells. *Chem. Biol.* **20,** 747–52 (2013).

29. Serber, Z., Corsini, L., Durst, F. & Dötsch, V. In-cell NMR spectroscopy. *Methods Enzymol.* **394,** 17–41 (2005).

30. Selenko, P. & Wagner, G. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* **158,** 244–53 (2007).

31. Selenko, P. & Wagner, G. NMR mapping of protein interactions in living cells. *Nat. Methods* **3,** 80–81 (2006).

32. Miao, Y. *et al.* M2 proton channel structural validation from full-length protein samples in synthetic bilayers and E. coli membranes. *Angew. Chem. Int. Ed. Engl.* **51,** 8383–6 (2012).

33. Fu, R. *et al.* In situ structural characterization of a recombinant protein in native Escherichia coli membranes with solid-state magic-angle-spinning NMR. *J. Am. Chem. Soc.* **133,** 12370–3 (2011).

34. Etzkorn, M. *et al.* Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **46,** 459–62 (2007).

35. Jacso, T. *et al.* Characterization of Membrane Proteins in Isolated Native Cellular Membranes by Dynamic Nuclear Polarization Solid-State NMR Spectroscopy without Purification and Reconstitution. *Angew. Chemie* **124,** 447–450 (2012).

36. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4863–8 (2012).

37. Braun, V. & Rehn, K. Chemical Characterization, Spatial Distribution and Function of a Lipoprotein (Murein-Lipoprotein) of the E. coli Cell Wall. The Specific Effect of Trypsin on the Membrane Structure. *Eur. J. Biochem.* **10,** 426–438 (1969).

38. Shu, W., Liu, J., Ji, H. & Lu, M. Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. *J. Mol. Biol.* **299,** 1101–12 (2000).

39. Lugtenberg, B. E. N. & Alphen, L. V. A. N. MOLECULAR ARCHITECTURE AND FUNCTIONING OF THE OUTER MEMBRANE OF E S C H E R I C H I A C O L I AND OTHER GRAM-NEGATIVE BACTERIA I . Introduction IA . Scope of this review The cytosol of a bacterial cell is surrounded by a complex cell envelope which usuall. **737,** 51–115 (1983).

40. Chang, T.-W., Lin, Y.-M., Wang, C.-F. & Liao, Y.-D. Outer membrane lipoprotein Lpp is Gram-negative bacterial cell surface receptor for cationic antimicrobial peptides. *J. Biol. Chem.* **287,** 418–28 (2012).

41. Hirota, Y., Suzuki, H., Nishimura, Y. & Yasuda, S. On the process of cellular division in Escherichia coli : A mutant of E . coli lacking a murein-lipoprotein * Biochemistry : **74,** 1417–1420 (1977).

42. Suzuki, H., Nishimura, Y., Yasuda, S., Nishimura, A. & Yamada, M. Murein-Lipoprotein of Escherichia coli : A Protein Involved in the Stabilization of Bacterial Cell Envelope. **9,** 1–9 (1978).

43. Waksman, G. & Orlova, E. V. Structural organisation of the type IV secretion systems. *Curr. Opin. Struct. Biol.* **17,** 24–31 (2014).

44. Backert, S. & Meyer, T. F. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr. Opin. Microbiol.* **9,** 207–17 (2006).

45. Ninio, S. & Roy, C. R. Effector proteins translocated by Legionella pneumophila: strength in numbers. *Trends Microbiol.* **15,** 372–80 (2007).

46. Llosa, M., Schröder, G. & Dehio, C. New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol.* **20,** 355–9 (2012).

47. Waksman, G. & Fronzes, R. Molecular architecture of bacterial type IV secretion systems. *Trends Biochem. Sci.* **35,** 691–8 (2010).

48. Chandran, V. *et al.* Structure of the outer membrane complex of a type IV secretion system. *Nature* **462,** 1011–5 (2009).

49. Fronzes, R. *et al.* Structure of a type IV secretion system core complex. *Science* **323,** 266–8 (2009).

50. Rivera-Calzada, A. *et al.* Structure of a bacterial type IV secretion core complex at subnanometre resolution. *EMBO J.* **32,** 1195–204 (2013).

51. Arteaga, C. L. & Engelman, J. a. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25,** 282–303 (2014).

52. Yarden, Y. The EGFR family and its ligands in human cancer : signalling mechanisms and therapeutic opportunities. **37,** 3–8 (2001).

53. Ferguson, K. M. *et al.* EGF Activates Its Receptor by Removing Interactions that Autoinhibit Ectodomain Dimerization. *Mol. Cell* **11,** 507–517 (2003).

54. Lu, C. *et al.* Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor. *Mol. Cell. Biol.* **30,** 5432–43 (2010).

55. Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277,** 46265–72 (2002).

56. Endres, N. F. *et al.* Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell* **152,** 543–556 (2013).

CHAPTER TWO

Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR

Published in Nat Methods 2015 Jul; 12(7): 649-52

Mohammed Kaplan[1], Abhishek Cukkemane[1#], Gydo C.P. van Zundert[1], Siddarth Narasimhan[1], Mark Daniëls[1], Deni Mance[1], Gabriel Waksman[2], Alexandre M.J.J. Bonvin[1], Rémi Fronzes[3], Gert E. Folkers[1,4] and Marc Baldus[1,4]

[1]NMR Spectroscopy, Utrecht University, 3584 CH Utrecht, The Netherlands.

[2]Institute of Structural and Molecular Biology, University College London and Birkbeck, London WC1E 7HX, United Kingdom

[3] Groupe à 5 ans, Biologie structurale de la secretion bacterienne, Unité mixte de recherche CNRS-institut Pasteur 3528, Institut Pasteur, 75015 Paris, France

2.1 Abstract

Studying biomolecules at atomic-resolution in their native environment is the ultimate aim of structural biology. We investigated the bacterial type IV secretion system core complex (T4SScc) by cellular DNP-based solid-state Nuclear Magnetic Resonance spectroscopy to validate a structural model generated by combining *in-vitro* and *in-silico* data. Our results indicate that T4SScc is well folded in the cellular setting revealing protein regions that so far remained elusive for *in-vitro* studies.

2.2 Introduction

Super high-resolution light microscopy[1] and electron tomography[2] have provided unprecedented insights into cellular organization on the nanometer range. Obtaining information at atomic level thus far required modelling using *in-vitro* data provided by X-ray crystallography, Nuclear Magnetic Resonance (NMR) or electron microscopy (EM). In-cell solution-state NMR can track structure provided that molecular units tumble rapidly[3,4]. Solid-state NMR spectroscopy (ssNMR) has been used to study small proteins embedded in native membranes[5–8] and we studied[9,10] a 150 amino acid membrane protein using cellular preparations. However, cellular ssNMR experiments on larger molecules or complexes pose additional challenges ranging from sample preparation to data interpretation. Here, we show that the combination of dedicated labeling schemes, cellular ssNMR and Dynamic Nuclear Polarization (DNP)[11] allows to *directly* examine the validity of structural models of membrane-associated complexes in a cellular setting (**Fig. 1a**).

We studied the type IV secretion system core complex (T4SScc), a one megadalton protein machine consisting of 14 copies of 3 proteins (VirB7, VirB9 and VirB10, **Fig. 1b**)[12]. It is part of a larger machinery (T4SS) that spans the periplasm and is embedded in both the inner and outer membrane of Gram-negative bacteria. The complex transports various substrates including plasmid DNA during bacterial conjugation or effector proteins into eukaryotic cells[13]. Structural information on T4SScc has been gathered on purified complexes using EM[12,14] and atomic structures have been obtained[15] for the outer layer (OL) built by VirB7 and the C-terminal regions of VirB9 and VirB10 (**Fig. 1b**). Recent work suggests that the core complex also largely maintains its structure in T4SS constructs comprising T4SScc components and the inner membrane components VirB3, VirB4, VirB5, VirB6 and VirB8,

connected to the inner membrane complex via the inner layer (IL) of T4SScc and a flexible region called the stalk[16]. By cysteine scanning studies[17] it became evident that the IL of T4SScc contains N-terminal protein segments of VirB10 that inserts in the inner membrane, which is located at a unique position to regulate substrate transfer across the cell envelope. However, atomic information on the IL has remained elusive. In fact, a comparison of the available EM data suggests that the N-terminal region of VirB10 is compact or unstructured in purified variants of the T4SScc and must adopt a more extended conformation reaching in the inner membrane in larger T4SS complexes and/or when embedded in the cellular envelope[16].

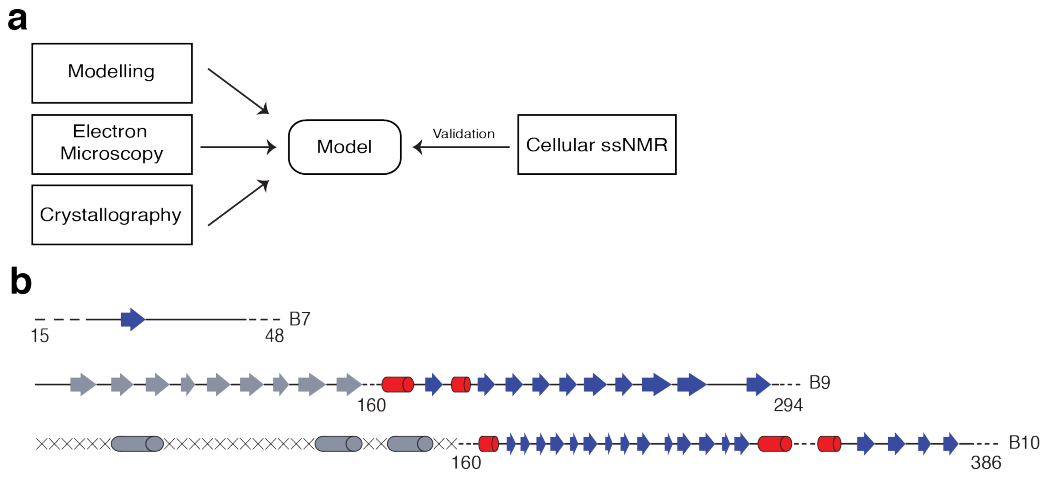


Fig. 1: **a**) Scheme to use cellular ssNMR as a validation tool of structural models obtained by *in-vitro* and *in-silico* techniques in their native environment. **b**) Schematic view of T4SScc (VirB7, top, VirB9, middle, VirB10, bottom) where red and blue symbols represent helical and extended conformations of the crystallized part (PDB=3JQO). Grey regions and black crosses reflect consensus and non-consensus regions of different prediction servers, respectively.

2.3 Materials and methods

Sample preparation. BL21 WT and BL21 double mutant[9] cells were transformed with pKM101 following earlier work[12] and plated on LB agar plates enriched with glucose (4 g/L) and containing chloramphenicol (35 mg/L). Next day, a fresh colony of BL21WT or BL21dm was inoculated in a 2 ml LB culture at 37° C until reaching OD_{600}~1. Thereafter, the cells were centrifuged and transferred into 25 ml of unlabeled M9 medium. At OD_{600}~1, cells were centrifuged again and transferred into 250 ml of unlabeled M9 medium. When reaching OD_{600}~1 the cells were centrifuged for 10 minutes at 2,000g and transferred into 250 ml of ^{13}C,^{15}N labeled M9 medium in case of uniformly labeled samples and then induced by tetracycline (200 µg/L, see Ref.[12]). After induction, samples were incubated at 20° C over night. 5-6 hours after induction, another 200 µg/L of tetracycline and 35 mg/L chloramphenicol was added. In case of GSLV-T4SScc sample, 200 mg/L of each labeled amino acid was added to unlabeled M9 medium at OD_{600}~1, 20 minutes after induction with tetracycline. In TV-T4SScc sample, in addition to the labeled amino acids, all other amino acids were added (as unlabeled at 200 mg/L) to the medium after induction to decrease scrambling. Cell lysis and cell envelope samples were prepared as described earlier[9]. The T4SScc was further purified as in Ref.[12] and visualized on a Tecnai T12 120 kV BioTwin electron microscope after negative staining with 2% uranyl acetate.

For DNP measurements, the sample was washed twice with a buffer containing 20 mM AMUPol in 32.5% D_2O, 12.5% H_2O, 10% 100mM HEPES (pH=7) and 45% glycerol d8 for GSLV-T4SScc and 22.5% glycerol d8 for TV-T4SScc. 50 µl of this buffer was used in each washing step. Samples were centrifuged after washing (100,000 *g*, 25 minutes).

Solid-state NMR & DNP experiments. For NMR measurements, a standard-bore 700 MHz as well as wide-bore 800 MHz/ 527 GHz DNP and 400 MHz/263 GHz DNP systems (Bruker Biospin) were used. We filled cellular envelope preparations containing approximately 1 mg of ^{13}C,^{15}N labeled T4SScc into standard 3.2 mm rotors. For all DNP measurements, samples were cooled down to 100 K in a 3.2 mm sapphire rotor. The DNP enhancement was measured by overlaying HC CP/MAS spectra recorded with and without microwave irradiation. Two and three-dimensional NC correlation spectra were recorded using SPECIFIC-CP ^{15}N-^{13}C transfers. Homonuclear (^{13}C,^{13}C) transfers were established using PARIS or spin-diffusion blocks. ^{1}H decoupling using SPINAL64 was employed during evolution and detection periods.

Modeling and structural analysis. NMR signals were predicted using crystal structures of the outer membrane complex crystal structure (PDB=3JQO) and modelled inner membrane complex. The secondary structure of the N-termini of B9 and B10 were predicted using JPred3 and PSIPRED. Signals stemming from Braun's lipoprotein (Lpp) were predicted using the Lpp crystal structure (PDB=1EQ7). Chemical shift predictions for the crystal part of T4SScc was made using ShiftX and NMR correlations were derived by FANDAS[18]. Analysis of the NMR/DNP spectra was performed using Sparky.

The atomic model of the outer membrane complex of T4SScc (PDB=3JQO) was fitted into the EM density (EMD=2232)[14] as a rigid body. The predicted transmembrane helix in the N-terminal part of VirB10 was modeled by imposing a helical conformation on the amino acid sequence and then manually docked into the cryo-EM density using UCSF Chimera. For the non-crystalline N-terminal part of VirB9, the already docked atomic model (PDB=2YPW)[14] in the EM density (EMD=2232) was used.

2.4 Results and Discussion

To obtain structural information of T4SScc in its cellular setting, we co-expressed all three subunits in *E.coli* BL21 (DE3) wild type (WT, **Fig.2a**) as well as in Omp A/F deficient cells[9] (BL21 double mutant (dm), **Fig. 2a**) following earlier procedures[12]. In both cases, cell envelope fractions contained T4SScc components and EM studies after further purification revealed intact protein complexes (**Fig. 2b**). Furthermore, previous work[15] as well as our experimental data presented below strongly suggested that the complex was correctly folded in the cellular envelope. For cellular ssNMR studies we prepared uniformly, selectively [^{13}C,^{15}N] (Gly, Ser, Leu, Val) as well as (^{13}C,^{15}N -Thr, ^{15}N-Val) labeled T4SScc in the cell envelope of BL21dm cells (referred henceforth to as U, GSLV and TV labeled-T4SScc, respectively). Specific amino acid labeling was confirmed by 2D ssNMR (**Fig. 3**). For spectral evaluation, we used a hybrid approach (**Fig. 1a**) with a T4SScc structural model combining previous EM results and X-ray data available for the OL[14,15]. For T4SScc regions lacking secondary structural information, we employed[14] structure prediction servers. Following earlier work[9,10], a comparison of ssNMR data of cell envelopes without and with protein expression (**Fig. 4**) confirmed that T4SScc and Braun's lipoprotein (Lpp) are dominant. Subsequently, we adapted the NMR software package FANDAS[18] to predict spectra of T4SScc, Lpp and of the other dominating cellular envelope compounds[9]

Lipopolysaccharides (LPS), Peptidoglycans (PG) and lipids (phosphatidylethanolamine: PE) that all contribute to the ssNMR spectrum of U[^{13}C,^{15}N] labeled cellular envelope preparations (**Fig. 5**).

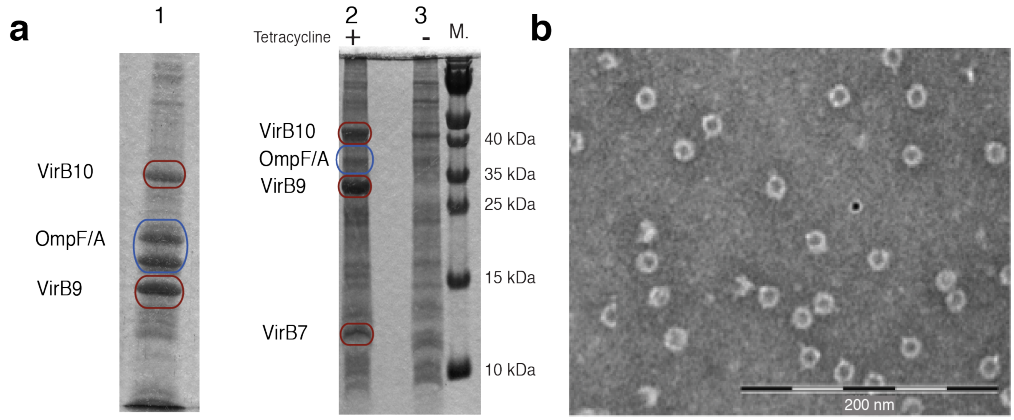


Fig. 2: **a**) Lane 1: 10% Tris/Glycine SDS-PAGE of cell envelope of BL21 WT expressing T4SScc. Lanes 2 and 3: 17.5% Tris/Tricine SDS-PAGE of cell envelope of BL21dm expressing T4SScc. M refers to the molecular weight marker. **b**) Electron Microscopy data of T4SScc after purification

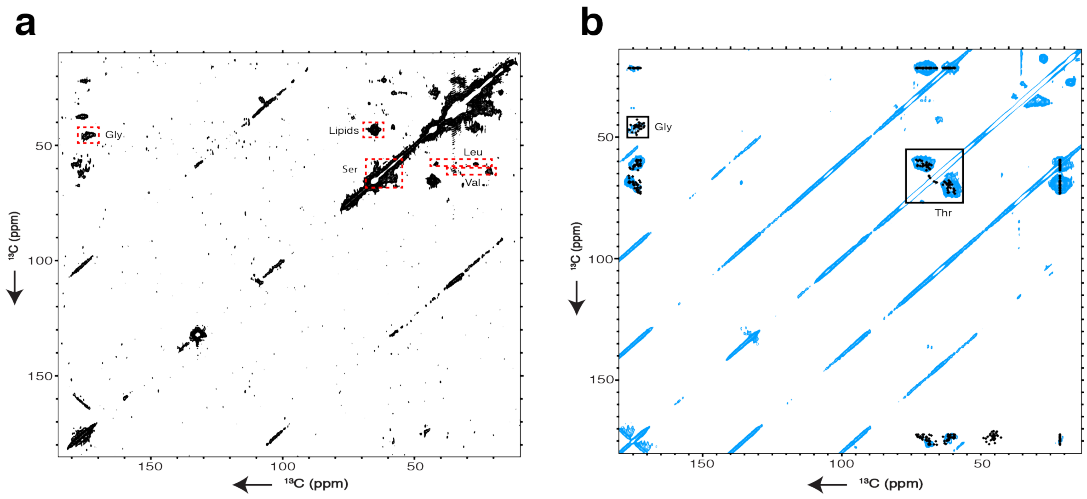


Fig. 3: Two-dimensional ssNMR ^{13}C-^{13}C spin diffusion data on BL21dm cell envelopes **a**) expressing GSLV labeled T4SScc recorded at 700 MHz and b) expressing TV labeled T4SScc recorded at 800 MHz DNP. See tables for further spectroscopic information. In **a**) intra-residue correlations of Leu, Ser and Valine are indicated along with cross peaks consistent with labeled lipid signals. In **b**), Black crosses represent FANDAS predictions for TV-T4SScc.

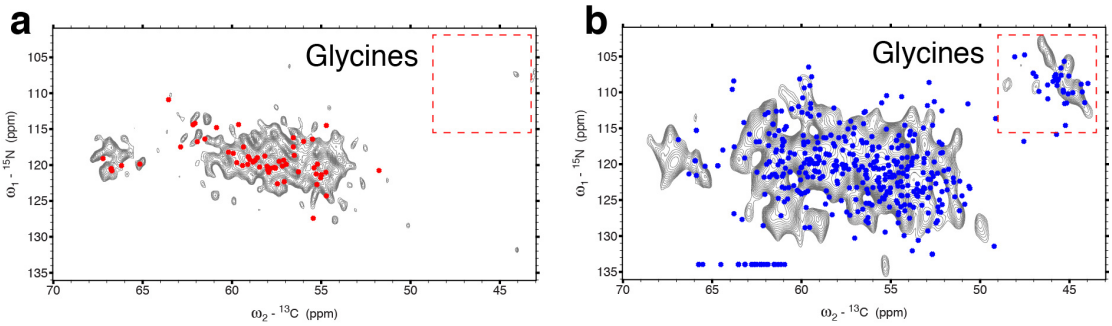


Fig. 4: 2D NCa data obtained on cell envelope prepared from full ^{13}C,^{15}N labeled BL21dm cells without (**a**) and (**b**) with expression of T4SScc. The dashed red box represents the spectral region typical for NCa for Glycine correlations. Red and blue crosses reflect spectroscopic predictions on the basis of the Lpp crystal structure and the T4SScc model.

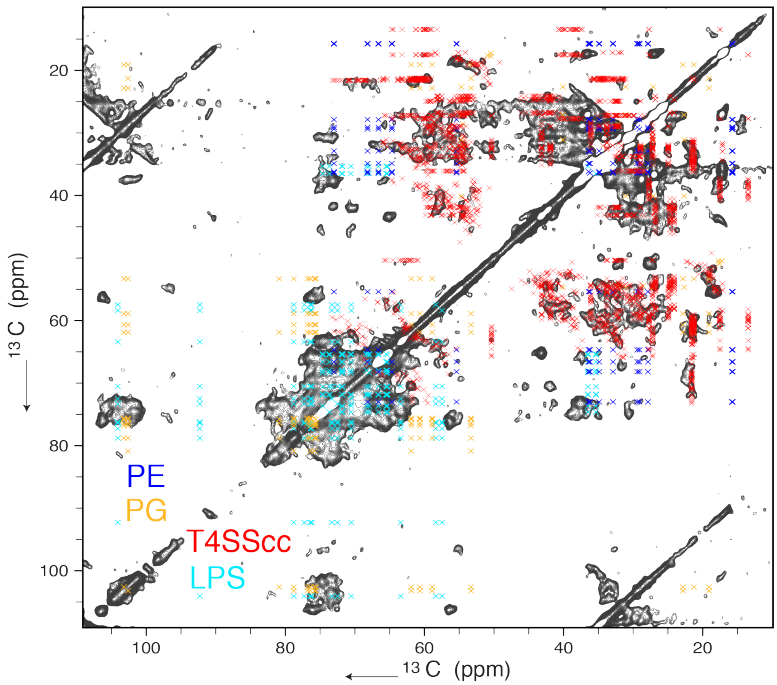


Fig. 5: 2D carbon-carbon ssNMR correlation experiment of U-labeled T4SScc with FANDAS predictions for T4SScc (red), LPS (cyan), PE (blue) and PG (brown). We note that Lpp signals would superimpose with T4SSc signals.

We exploited the well-established correlation between NMR resonance frequencies and protein secondary structure (see, e.g., Ref.[19]) to obtain structural information on the backbone fold of cell-embedded T4SScc. Optimized amino acid selective labeling was used to

maximize spectral dispersion. For example, preparing T4SScc using [^{13}C,^{15}N] Thr and [^{15}N] Val should give rise to seven sequential correlations in β-strand (blue), α-helical (red) as well as random-coil (rc) regions (black) of the entire 1 MDa complex (**Fig. 6a**). Five of these correlations stem from the previously crystallized OL segments (**Fig. 6a**) and two sequential correlations originate from the structurally elusive regions in VirB9 and VirB10. For GSLV-T4SScc, we expected 51 sequential correlations with 31 from the OL (**Fig. 6b**) and 20 originating from N-terminal VirB9 and VirB10 (**Fig. 6b**). Lpp sequential correlations do not contribute to our spectral analysis presented for these samples.

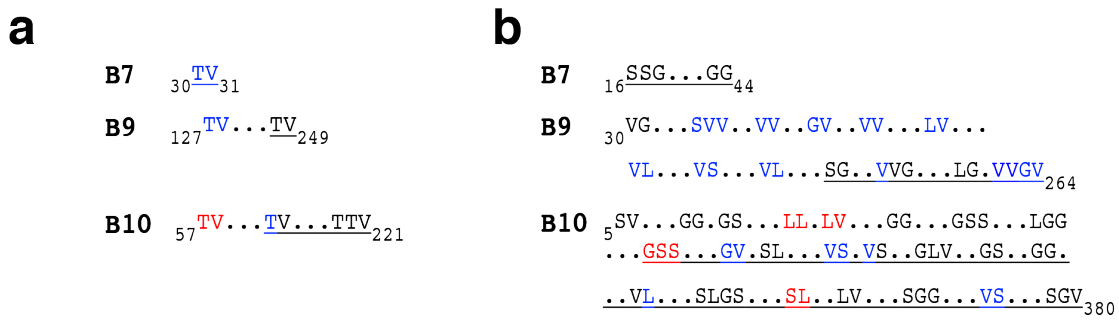


Fig. 6: (**a**) and (**b**) Predicted sequential correlations in TV-T4SScc and GSLV-T4SScc respectively, with letters indicating correlations in random coil (black), β-strand (blue) and α-helix (red). Underlined segments reflect crystallized T4SScc subdomains. Single dots represent (1-4) residues, double dots (5-8) residues, and triple dots more than 8 residues.

We prepared ssNMR samples amenable to DNP studies at 400 and 800 MHz using AMUPol[20] (**Fig. 7**). We observed a DNP enhancement factor of ~60, enabling 3D intra-residue (NCACX) and inter-residue (NCOCX) experiments at 400 MHz. At 800 MHz, the DNP enhancement factor of ~15 allowed for 2D NCACX and NCOCX experiments for cross-validation. For TV-T4SScc, a combined evaluation of these data sets confirmed the presence of resonances of α, β and rc conformations (**Fig. 8a,b & c).** Using FANDAS and standard deviations[21] for protein secondary structure predictions, we identified sequential strips connected via a unique ^{15}N chemical shift along ω_1 in (ω_2,ω_3) intra-residue 2D Ca-Cx and Co-Cx planes in our 3D data sets (**Fig 8a**), respectively. The chemical shifts were also consistent with a 2D NCOCX experiment at 800 MHz (**Fig. 8b)** and led to tentative assignments for all sequential pairs (**Tables 1** and **2**). The increased spectral resolution of the 800 MHz data revealed several resolved correlations in the Cβ region and allowed us to

assess the spectral line width. For correlations such as ^{202}ThrVal^{203} (B10) in the OL or ^{30}ThrVal^{31} located in B7, we determined ^{13}C and ^{15}N resonance line width of ~1.5 ppm and 3 ppm, respectively. Only two correlations relating to α-helical elements predicted for ^{57}ThrVal^{58} and ^{220}ThrVal^{221} in B10 exhibited larger line width, especially in the ^{15}N dimension, consistent with variable dynamics within the T4SScc.

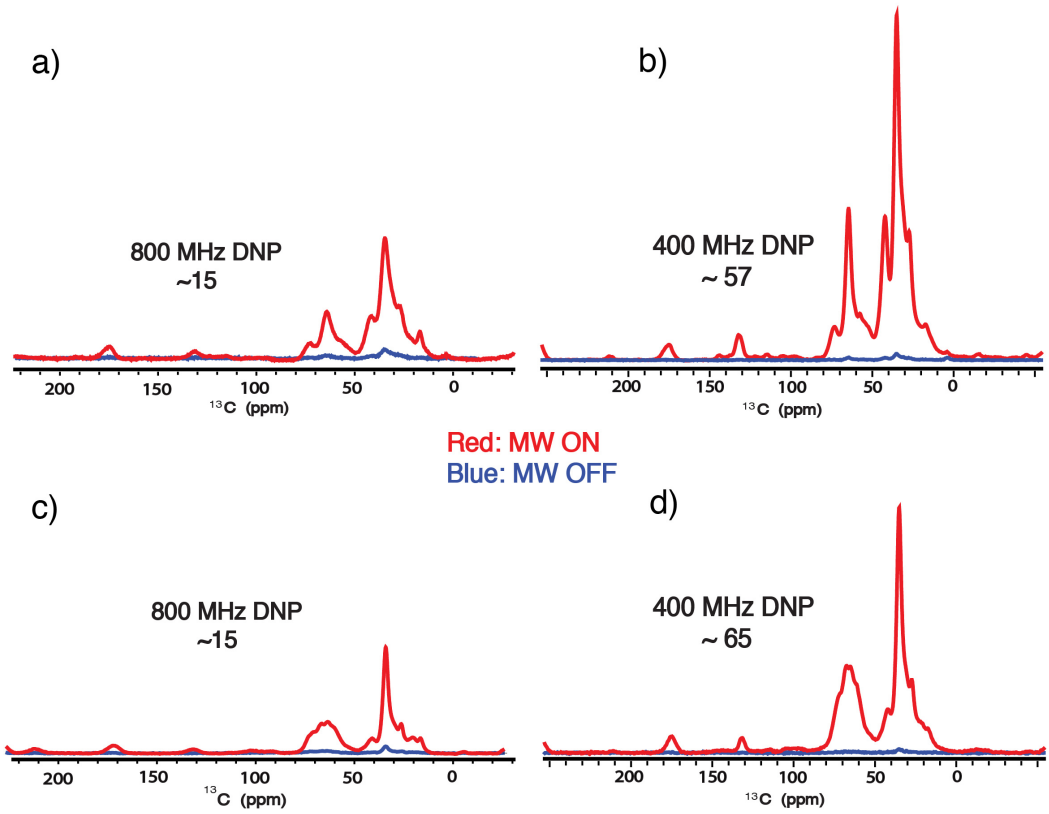


Fig. 7: Signal enhancements using DNP at 800 MHz (**a**, **c**) and 400 MHz (**b**, **d**) of the GSLV labeled (**a** and **b**) and TV labeled (**c** and **d**) BL21dm cell envelope expressing T4SScc after addition of 20 mM AMUpol. In **a**), a 1D ^{13}C CP spectrum measured on 800 MHz DNP with microwave irradiation off (blue) and on (red). An enhancement factor of 15 was obtained. In **b**) a 1D ^{13}C CP spectrum of the same sample was measured on a 400 MHz DNP machine (Bruker Biospin) with microwave irradiation off (blue) and microwave irradiation on (red). Here an enhancement factor of (57) was obtained. **c**) and **d**) represent equivalent experiments for the TV labeled sample where an enhancement factor of 15 was obtained at 800 MHz DNP and 65 at 400 MHz DNP.

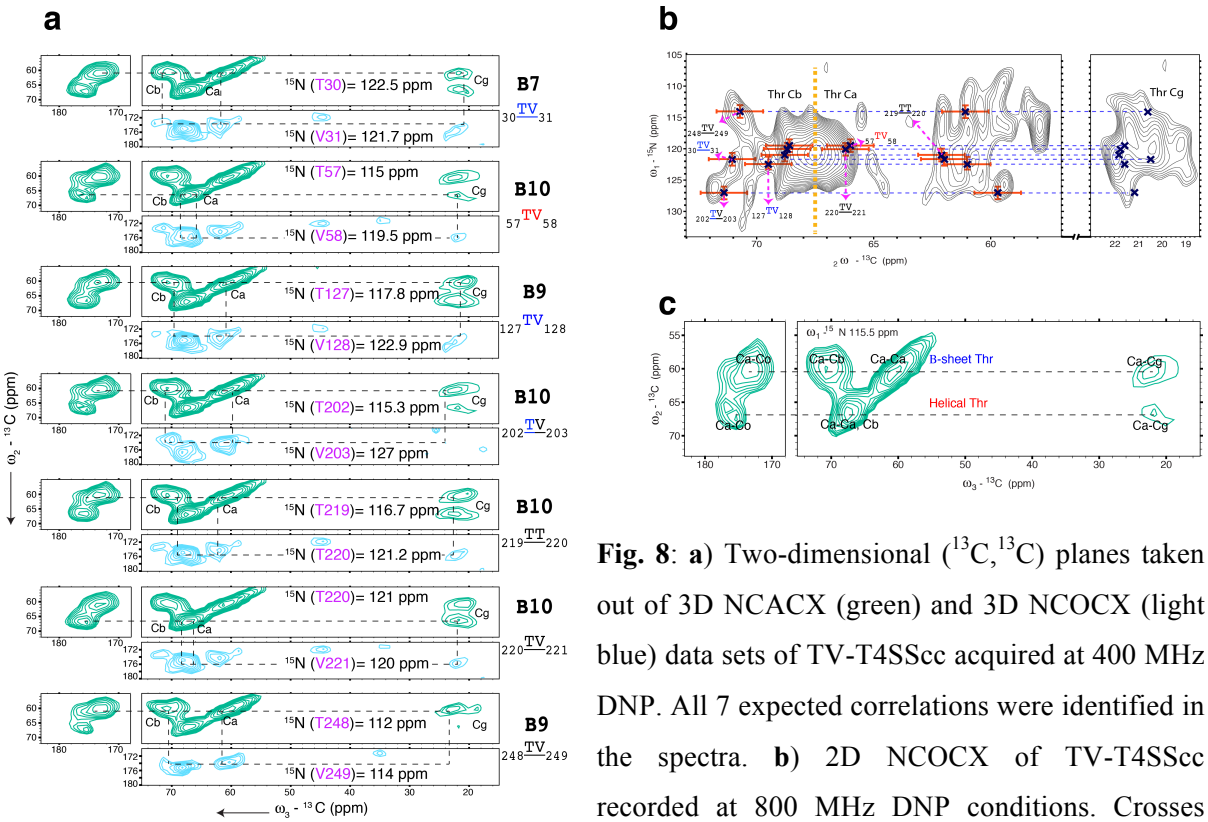


Fig. 8: **a**) Two-dimensional (^{13}C,^{13}C) planes taken out of 3D NCACX (green) and 3D NCOCX (light blue) data sets of TV-T4SScc acquired at 400 MHz DNP. All 7 expected correlations were identified in the spectra. **b**) 2D NCOCX of TV-T4SScc recorded at 800 MHz DNP conditions. Crosses represent the seven correlations expected in this labeling scheme. Additional correlations can be explained by scrambling to glycine. Red error bars for carbon are 2 ppm and for nitrogen 3 ppm, respectively. For clarity, the dashed vertical line separates Cb (resonating above 67.5 ppm) and Ca (resonating below 67.5 ppm) ^{13}C resonance frequencies. **c**) Example of an ω_2-ω_3 plane in a 3D NCACX experiment at 400 MHz DNP conditions on TV-T4SScc revealing intra-residue Threonine correlations that exhibit secondary Ca/Cb chemical shifts typical for β-sheets (blue) and α-helices (red). Dashed lines highlight carbon backbone as well as side chain correlations.

Next, we investigated the validity of our structural model by comparing structural predictions to 3D data on GSLV-T4SScc (**Fig. 9)** as in Fig. 8a. Again we could identify ω_2-ω_3 planes leading to sequential correlations in the ^{261}ValValGlyVal^{264} VirB9 stretch of the OL forming a β-strand in crystals in agreement with the observed chemical shifts. NMR data and predictions were also consistent with the detection of the only GlySer pair occurring in an α-helix in (^{175}GlySer^{176}) in VirB10. We expected the sequence (^{40}KAFVILMALLALVFIGITV^{58}) in the N-terminal region of B10 to adopt an α-helical conformation (**Fig. 6**). Indeed, we could identify in our 3D data correlations consistent with LeuLeu (only found in the N-terminal B10) or LeuVal contacts (**Fig. 9**), in line with the identification of the unique α-helical ^{57}ThrVal^{58} pair for N-terminal VirB10 (**Fig. 8a** and **b**). SsNMR experiments at higher temperatures speak against a sizable increase in protein

motion for the N-terminal half of VirB10 (**Fig. 10**) and thus suggest membrane association for this N-terminal VirB10 stretch.

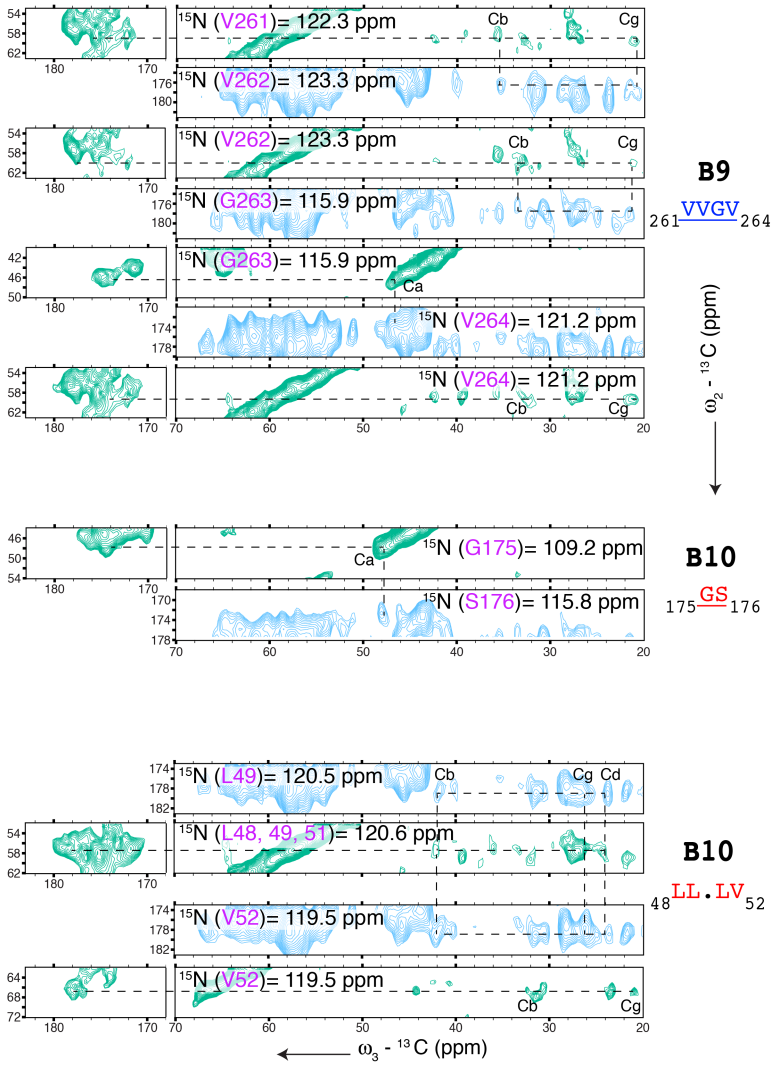


Fig. 9: Two-dimensional (^{13}C,^{13}C) planes taken out of 3D NCACX (green) and 3D NCOCX (light blue) data sets of GSLV-T4SScc acquired at 400 MHz DNP. Dashed lines highlight carbon backbone as well as side chain correlations.

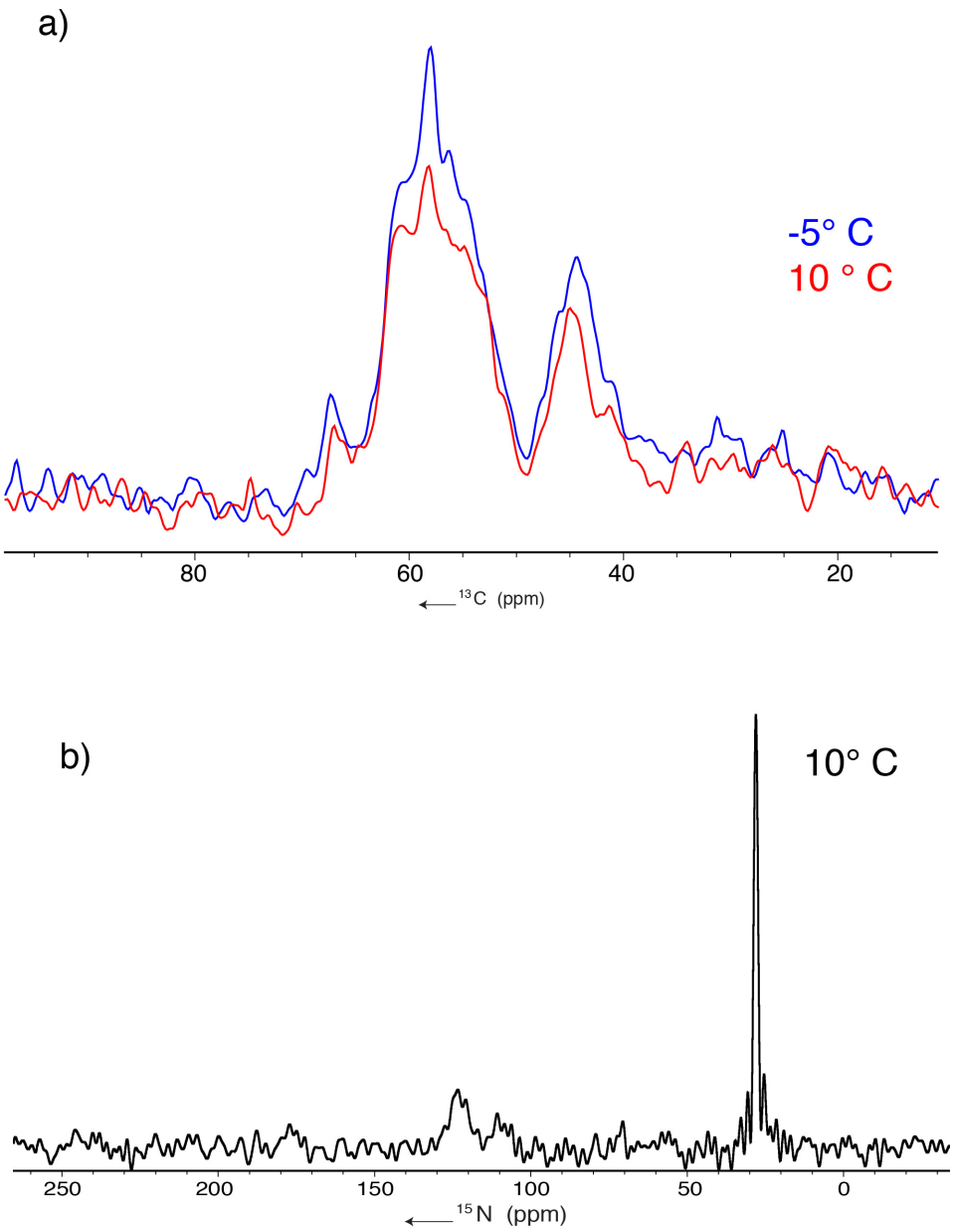


Fig. 10: **a**) 1D NCA experiment on BL21dm cell envelope expressing GSLV labeled T4SScc at -5° C (blue spectrum) and 10° C (red spectrum). No significant change in the spectrum shape or intensity is observed between the two temperatures. This indicates that, at high temperatures, T4SScc is still largely rigid. Taking into consideration that about 25% of the signal in this sample is originating from the N-terminal part of B10, this implies a stable N-terminus of B10 at this temperature. **b**) 1D 90° pulse on ^{15}N performed on the same sample mentioned in **a**) at 10° C. A prominent sharp lipid peak is visible in the spectrum (~36 ppm), consistent with mobile lipids, and a small broad protein signal is seen at ~120 ppm. This also indicates a stable T4SScc in the cell envelope at high temperatures.

Tables:

Table 1: Chemical shifts of the identified correlations of TV-T4SScc with secondary structure information determined from standard chemical shift tables. We note that these tentative assignments were found within the error margins of the predicted correlations.

	N (NCOCX) (±3 ppm)	**N (NCACX) (±3 ppm)**	**Ca (±2 ppm)**	**Cb (±2 ppm)**	**Secondary Structure**
B7					
^{30}TV^{31}	121.7	122.5	62	71	β sheet
B9					
^{127}TV^{128}	122.9	117.8	61	69.5	β sheet
^{248}TV^{249}	114	112	61.6	70.2	r.c.
B10					
^{57}TV^{58}	119.5	115	65.9	68.6	α-helix
^{202}TV^{203}	127	115.3	59.7	71.4	r.c.
^{219}TT^{220}	121	116.7	62.1	68.8	r.c.
^{220}TV^{221}	120	121	66.2	68.7	r.c./ α-helix

Table 2 Chemical shifts of the identified correlations of GSLV-T4SScc with secondary structure information determined from standard chemical shift tables. We note that these tentative assignments were found within the error margins of the predicted correlations

	N (NCOCX) (±3 ppm)	**N (NCACX) (±3 ppm)**	**Ca (±2 ppm)**	**Cb (±2 ppm)**	**Secondary Structure**
B9					
^{261}VV^{262}	123.3	122.3	59.5	35.5	β sheet
^{262}VG^{263}	115.9	123.3	60.5	33.5	β sheet
^{263}GV^{264}	121.2	115.9	46.2		β sheet
B10					
^{175}GS^{176}	115.8	109.2	48		α-helix
^{48}LL^{49}	120.5	120.6	57.3	42	α-helix
^{51}LV^{52}	119.5	120.6	57.3	42	α-helix

2.5 Conclusions

Taken together, our analysis provided strong support that the 3D fold of the T4SScc OL seen in *in-vitro* crystals is maintained in the cellular preparation (**Fig. 11**). At the same time, these

findings underlined the validity of our chemical-shift analysis to probe protein structure in a cellular environment. Moreover, our observed correlations in the IL part of T4SScc were consistent with a ß-strand arrangement for ^{127}ThrVal^{128} in VirB9 providing the first experimental atomic evidence that the N-terminal region of VirB9 adopts a distinct conformation at the periplasmic surface of the cell-embedded core complex. Also, we could identify a series of sequential correlations (^{48}LLALVFIGITV^{58}) of the IL which suggest that VirB10 is inserting in the inner membrane as an α-helix allowing the core complex to span the entire cell envelope (**Fig. 11**). Our results also indicate that VirB9 and VirB10 are involved in forming a stable scaffold that allows for assembly of the entire complex in the cellular envelope. At the same time, differences in ssNMR line width seen in our study suggest structural flexibility that may be needed for accommodation of the pilus and substrate secretion[22].

Existing approaches rely on the molecular modeling of *in-vitro* data into low-resolution cellular structures. Combination with cellular ssNMR data such as shown here allows for direct experimental validation of atomic models in a cellular environment. In the case of T4SScc, such information may help to engineer systems as *in-vivo* gene-delivery vectors with applications in DNA vaccination and therapeutic gene therapy in human cells[23]. Furthermore, our approach may be applicable to study molecular complexes in their native environment such as occurring during gene transcription, protein folding or cellular signaling.

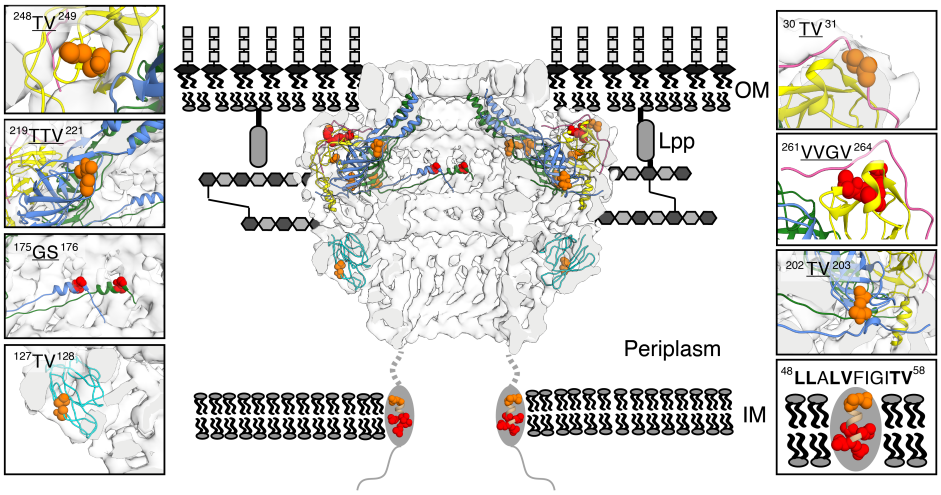


Fig. 11: Summary of residue-specific ssNMR probes and their location in reference to the cellular envelope including the T4SScc electron microscopy map with the outer membrane complex fitted inside (PDB = 3JQO). N-terminus of VirB9 is also docked into the map. Identified residues in our spectra are shown as orange balls for the TV-T4SScc and red balls for the GSLV-T4SScc. Boxes represent zoom-ins of the identified correlations. See Ref.[16] for further information on the cellular location of the T4SScc subcomponents.

References

1. Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* **143,** 1047–58 (2010).

2. Leis, A., Rockel, B., Andrees, L. & Baumeister, W. Visualizing cells at the nanoscale. *Trends Biochem. Sci.* **34,** 60–70 (2009).

3. Serber, Z., Corsini, L., Durst, F. & Dötsch, V. In-cell NMR spectroscopy. *Methods Enzymol.* **394,** 17–41 (2005).

4. Banci, L., Barbieri, L., Luchinat, E. & Secci, E. Visualization of redox-controlled protein fold in living cells. *Chem. Biol.* **20,** 747–52 (2013).

5. Etzkorn, M. *et al.* Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **46,** 459–62 (2007).

6. Fu, R. *et al.* In situ structural characterization of a recombinant protein in native Escherichia coli membranes with solid-state magic-angle-spinning NMR. *J. Am. Chem. Soc.* **133,** 12370–3 (2011).

7. Jacso, T. *et al.* Characterization of Membrane Proteins in Isolated Native Cellular Membranes by Dynamic Nuclear Polarization Solid-State NMR Spectroscopy without Purification and Reconstitution. *Angew. Chemie* **124,** 447–450 (2012).

8. Miao, Y. *et al.* M2 proton channel structural validation from full-length protein samples in synthetic bilayers and E. coli membranes. *Angew. Chem. Int. Ed. Engl.* **51,** 8383–6 (2012).

9. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4863–8 (2012).

10. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51,** 2998–3001 (2012).

11. Ni, Q. Z. *et al.* High frequency dynamic nuclear polarization. *Acc. Chem. Res.* **46,** 1933–41 (2013).

12. Fronzes, R. *et al.* Structure of a type IV secretion system core complex. *Science* **323,** 266–8 (2009).

13. Low, H. H. *et al.* Structure of a type IV secretion system. *Nature* 1–18 (2014). doi:10.1038/nature13081

14. Rivera-Calzada, A. *et al.* Structure of a bacterial type IV secretion core complex at subnanometre resolution. *EMBO J.* **32,** 1195–204 (2013).

15. Chandran, V. *et al.* Structure of the outer membrane complex of a type IV secretion system. *Nature* **462,** 1011–5 (2009).

16. Trokter, M., Felisberto-rodrigues, C., Christie, P. J. & Waksman, G. Recent advances in the structural and molecular biology of type IV secretion systems. *Curr. Opin. Struct. Biol.* **27,** 16–23 (2014).

17. Jakubowski, S. J. *et al.* Agrobacterium VirB10 domain requirements for type IV secretion and T pilus biogenesis. *Mol. Microbiol.* **71,** 779–94 (2009).

18. Gradmann, S. *et al.* Rapid prediction of multi-dimensional NMR data sets. *J. Biomol. NMR* **54,** 377–87 (2012).

19. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–23 (2009).

20. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chem. Int. Ed. Engl.* **52,** 10858–10861 (2013).

21. Seidel, K., Etzkorn, M., Schneider, R., Ader, C. & Baldus, M. Comparative analysis of NMR chemical shift predictions for proteins in the solid phase. *Solid State Nucl. Magn. Reson.* **35,** 235–42 (2009).

22. Waksman, G. & Orlova, E. V. Structural organisation of the type IV secretion systems §. *Curr. Opin. Struct. Biol.* **17,** 24–31 (2014).

23. Llosa, M., Schröder, G. & Dehio, C. New perspectives into bacterial DNA transfer to human cells. *Trends Microbiol.* **20,** 355–9 (2012).

CHAPTER THREE

Activation changes global and local dynamics of the full-length Epidermal Growth Factor Receptor in native cell membranes

Mohammed Kaplan[1], Siddarth Narasimhan[1], Cecilia de Heus[2], Klaartje Houben[1], Dusan Popov-Celeketic[2], Deni Mance[1], Eugene A. Katrukha[2], Purvi Jain[2], Willie JC Geerts[3], Gert Folkers[1], Lukas C. Kapitein[2], Paul MP van Bergen en Henegouwen[2], and Marc Baldus[1]

[1]*NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Department of Chemistry, Utrecht University, 3584 CH Utrecht, The Netherlands*

[2] *Cell Biology, Department of Biology Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands*

[3]*Biomolecular Imaging, Bijvoet Center for Biomolecular Research, Department of Chemistry, Utrecht University, 3584 CH Utrecht, The Netherlands,*

Submitted manuscript

3.1 Abstract

The epidermal growth factor receptor (EGFR) belongs to a family of transmembrane proteins that are involved in vital cellular processes such as cell proliferation or differentiation representing important targets in cancer therapy[1,2]. Detailed structural information has been obtained for various EGFR segments such as the extracellular region (ECR)[3,4], the tyrosine kinase domain[5] (KD) or for constructs containing the transmembrane (TMD) and juxtamembrane (JMD) domain in membrane mimetics[5-7.] However, a unified structural view that describes the ligand-induced functional coupling between the ECR and the intracellular domains of the full-length receptor in a native membrane environment has remained elusive[8,9]. Here we have developed a solid-state Nuclear Magentic Resonance (NMR)-based approach to directly examine structural and dynamical properties of the full-length EGFR in native membrane vesicles before and after activation by EGF. Our results suggest that the ligand-free state of EGFR is characterized by a rigid KD and point to a dynamic ECR and C-terminal domain. Ligand binding not only restricts the overall motion of EGFR, but also reduces local backbone fluctuations that are present both in the ECR and the intracellular tyrosine kinase domain. We propose an allosteric activation mechanism in which structural flexibility at both global and local scale is critical for ligand binding and receptor dimerization in EGFR and, potentially, other receptor tyrosine kinases.

3.2 Introduction

Solution-state NMR has become a powerful tool to elucidate the influence of protein dynamics for allosteric regulation in soluble enzymes[10,11] and protein receptors embedded in membrane mimetics[12-14]. Solid-state NMR (ssNMR), on the other hand, can give detailed structural insight into the role of the bilayered membrane for protein structure in synthetic[15] or native bacterial membranes[16] and can track ligand-induced conformational changes in transmembrane proteins[17]. In addition, ssNMR can probe changes in local or overall protein dynamics by the observed scaling of the dipolar couplings at ambient temperature[18,19] and by tracking ssNMR line width variations due to backbone fluctuations at low temperatures[20].

Here we have developed an ssNMR-based approach to probe the structure and dynamics of the full-length EGFR before and after activation by its natural ligand, the epidermal growth

factor (EGF) in membrane vesicles derived from eukaryotic cells. Using previous structural information about EGFR domains[3-6] , and assuming the C-terminal (CT) domain to be unstructured, we monitored EGFR structure and dynamics in membrane vesicles derived from eukaryotic cells at global and residue-specific levels.

3.3 Materials and methods

Cell Culture

A431 Cells obtained from ATCC (CRL-1555, LGC Standards, Germany) and EGFR negative cells NIH 3T3 clone 2.2 murine fibroblasts were cultured in Dulbecco's modified eagle's medium (DMEM: Gibco, invitrogen, Paisley, UK) containing 10% (v/v) fetal calf serum (FCS), L- glutamine, penicillin and streptomycin at 37 ºC with an atmosphere containing 5% CO_2.

Phosphorylation assay

Phosphorylation of EGFR was induced either by adding 8 nM EGF to the cells in medium before membrane vesicles were prepared or to membrane vesicles in a phosphorylation buffer for 10 min at 37^0 C. Proteins were separated by SDS-PAGE and blotted onto PVDF-membrane. The membrane was incubated with R-α-phosphoEGFR (Y1068) (Cell Signaling Technology, Danvers, Massachusetts) and M-α-Actin followed by G-α-R800 (Li-Cor) and G-α-mouse700 (Li-Cor). To detect EGFR the blot was first stripped with stripping buffer and then blocked and incubated with R-α-EGFR (C74B9, Cell signaling technology)) followed by G-α-R800. The detection was performed with the Odyssey imaging system (Li-COR) and bands were quantified using Odyssey software.

Preparation of a ^{13}C, ^{15}N labeled medium to label eukaryotic cells

For isotope labeling, we adapated published procedures using a combination of dialyzed fetal calf serum and labeled amino acid mixtures obtained from algae extracts to produce a [^{13}C, ^{15}N] enriched medium. 1 liter of DMEM without amino acids was supplemented with 2 g/L glucose and 1g of a ^{13}C, ^{15}N algal amino acid mixture (Cortecnet). Due to the absence of

certain amino acids in this mixture, unlabeled Trp (16 mg/L), Cys (62 mg/L) and Gln (2 mM) were added. In addition, 10% of dialyzed fetal calf serum was added to the medium.

1 g of labeled algal mixture contains the following: ASX: 8.8%, THR: 3.2%, SER: 4.6%, GLU: 8.2%, PRO: 4.0%, GLY: 8.3%, ALA: 11.9%, VAL: 7.0%, MET: 2.0%, ILE: 6.0%, LEU: 12.2%, TYR: 3.9%, PHE: 5.4%, HIS: 1.2%, LYS: 5.8%, TRP: 0.0%, ARG: 5.9%, CYS: 0.0%

^{13}C, ^{15}N labeling of eukaryotic cells:

A431 cells were cultured in the labeled medium described above on Corning cell culture dishes (150 mm x 25 mm). Cells cultured in the first week (2-3 passages) in the labeled medium were not used to prepare the samples to ensure full incorporation of labeled substance in the cells. Once the plates were ~ 80-90% confluent, cells were incubated with PBS containing 2 mM EGTA at 37° C for 15 min, after which the cells were manually scraped. Subsequently, cells were spun at 500xg for 10 min at 4° C. The cell pellet was resuspended in PBS and spun again at 500xg for 10 minutes at 4° C and used to prepare the membrane vesicles as described below. To fill a 3.2 mm rotor with [^{13}C, ^{15}N] labeled A431 membrane vesicles around 20 (150 mm X 25 mm) plates were used.

Membrane vesicle preparation:

Cells were re-suspended with homogenizer buffer (10 mM Tris pH 7.4, 250 mM sucrose, 1 mM EDTA. Phosphatase inhibitors (100 uM sodium-orthovanadate) and protease inhibitors (Complete, Roche) added freshly and cells were lysed by 10 passages through a syringe (21gx1.5; 0.2x40mm). Subsequently, cells were spun at 1000xg at 4°C for 10 min to remove unbroken cells, nuclei and cell debris from the supernatant. This was repeated until no pellet was visible anymore. The supernatant was subsequently spun at 150,000xg for 30 min at 4° C to collect membrane vesicles. The vesicles were resuspended in 10 mM Hepes pH 7.4 supplemented with protease and phosphatase inhibitors.

Preparation of [^{13}C, ^{15}N] A431 vesicles with EGF for ssNMR/DNP measurements:

Isolated A431 membrane vesicles were spun down at 124,000x*g* for 25 min at 4° C, and the pellet was resuspended in phosphorylation buffer (20 mM HEPES pH 7, 10 mM $MgCl_2$, 3 mM $MnCl_2$, 1 mM DTT, with protease and phosphatase inhibitors. To this buffer, 1 mM ATP was added. The vesicles were incubated with 8 nM EGF at 37° C for 10 min. Subsequently, the vesicles were washed three times with 10 mM HEPES pH 7.4 buffer containing protease and phosphatase inhibitors. For DNP samples, the sample was washed once with 10 mM HEPES pH 7.4 (containing protease and phosphatase inhibitors) and then two times with DNP buffer: 20 µl AMUPol (in D_2O), 20 µl H_2O, 20 µl D_2O, 40 mg glycerol-d8 (equivalent to 30 ul) and 10 µl 100 mM HEPES. For each washing step, 50 ul of the buffer was used.

Nanobody labeling and dSTORM imaging

Vesicles derived from A431 cells were labeled in suspension with 10 nM anti-EGFR (7D12) nanobodies conjugated to A647 for 1 hr at 4°C. Non-bound nanobodies were removed by centrifugation at 75,000xg for 40 min and vesicles were attached to glass slides for microscopical analysis. dSTORM microscopy was performed as recently described[21] on a Nikon Ti microscope equipped with a 100x Apo TIRF oil objective (NA. 1.49), a Perfect Focus System and an additional 2.5x Optovar to achieve an effective pixel size of 64 nm. Evanescent laser illumination was achieved using a custom illumination pathway with a 15 mW 405 nm diode laser (Power Technology) and a 40 mW 640 nm diode laser (Power Technology). Fluorescence was detected using a water-cooled Andor DU-897D EMCDD camera and ET series Cy5 filter (Chroma Technology). All components were controlled by µmanager software[22] . For dSTORM imaging of Alexa Fluor 647, the sample was continuously illuminated with 640 nm. In addition, the sample was illuminated with 405 nm light at increasing intensity to keep the number of fluorophores in the fluorescent state constant. Between 5000 and 10000 frames were recorded per acquisition with exposure time of 40 ms.

Purified vesicles from A431 (EGFR-positive) and 3T3 fibroblasts (EGFR negative) were incubated with 10 nM of anti-EGFR or non-specific nanobodies fused to Alexa647 fusion for one hour at 4°C. Four flow chambers with an approximate volume of 5 µL each were made

with stripes of double-sided tape between a plasma-cleaned 22x22 mm coverslip and the microscope slide. These chambers were filled with four consecutive 10 times dilutions of labeled vesicles and incubated for 5 minutes at RT. The chamber was washed with 25 µL of imaging buffer to remove non-stuck vesicles and sealed using vacuum grease. The composition of the imaging buffer was 100mM MEA, 5% w/v glucose, 700 µg/ml glucose oxidase and 40 µg/ml catalase in PBS buffer. The dilution containing optimal density of vesicles on the coverslip was chosen for the imaging (~ 20-30 vesicles per 100 μm^2). The signals from 3T3 fibroblasts vesicles labeled with anti-EGFR nanobody and A431 vesicles labeled with non-specific nanobody were not distinguishable from the noise.

Analysis of the dSTORM localization and rendering was performed as described earlier. In short, we used custom written ImageJ plugin for the single-molecule localization (https://github.com/ekatrukha/DoM_Utrecht). Each spot was fitted with asymmetric two-dimensional Gaussian PSF and only fits with a calculated width within ±30% of the measured PSF's standard deviation were accepted. Localizations within one pixel distance in a number of successive frames were considered to arise from the same molecule. In this case the weighted mean was calculated for each coordinate, where weights were equal to inverse squared localization errors. The resulting table with molecule coordinates was used to render the final localization image with 5 nm pixel size. Each molecule was plotted as a 2D Gaussian of the integrated intensity equal to one and with standard deviation in x and y equal to the localization precision. The noise arising from non-specific localization was suppressed using local density based filtering. Only particles having more than 150 neighbors in the circle of 250 nm radius were kept in the reconstruction.

Cryo-electron microscopy

For the preparation of thin vitrified specimens of the A431 vesicles, a 3 ul drop of sample was placed on the surface of a glow discharged Quantifoil micromachined holey carbon (R 2/2) TEM grid (Quantifoil Micro Tools GmbH, Jena, Germany) held by the Vitrobot mark IV tweezer (FEI, Eindhoven, The Netherlands). Before introducing the sample into the Vitrobot, the environmental chamber of the Vitrobot was equilibrated at room temperature (22°C) and humidity was set at 100 %. Blotting conditions were chosen so that a 10-500 nm liquid specimen film spanning R 2/2 um holes of the QF were formed when excess sample was

removed by the blotting filter paper in the Vitrobot. The specimen was released and fell through the opening shutter and into liquid ethane at its freezing point, where the thin specimen films were vitrified. The vitreous specimen was transferred under liquid nitrogen into a Gatan 626 single tilt liquid nitrogen cryo holder (Gatan GmbH, Munich, Germany) and into a Tecnai20 LaB6 electron microscope (FEI, Eindhoven, The Netherlands), where the specimen temperature was maintained below -165°C. An Eagle 4kX4k CCD camera (FEI, Eindhoven, The Netherlands) was used under normal and low-dose conditions to record micrographs of the vesicles, which was done in mTif format with a nominal under focus of 3 nm. Vesicle diameter was measured using the IMOD software package[23] .

Heat maps of amino acid distribution in EGFR

The available structures (See table 1 in the SI) were first split according to individual domains (DI-IV, TM, KD & CT). The parts of the domains that were elusive from the available structures were built in random coil conformation using PyMol. Secondary structure assignments were made for these structures by supplying the .pdb files to the software: STRuctural IDEntification (STRIDE)[24], which uses the phi and psi angles of the residues to assign secondary structures. The secondary structure assignments were simplified by sorting them into three categories based on their similarity: Alpha Helix, Beta Sheet and Random Coil (See table below). This data was then use to calculate the amino acid distribution per secondary structure over the whole protein.

STRIDE Assignment	Simplified Assignment
Alpha Helix	Alpha Helix
3_{10} Helix	
Strand	Beta Sheet
Bridge	
Turn	Random Coil
Coil	

Solid-state NMR & DNP experiments.

For NMR measurements, a standard-bore 700 MHz as well as wide-bore 800 MHz/ 527 GHz DNP and 400 MHz/263 GHz DNP systems (Bruker Biospin) were used. We filled fully [^{13}C,

^{15}N] or MFTL labeled A431 membrane vesicles obtained from around 20 (150 mm x 25 mm) plates into standard 3.2 mm rotors. For all DNP measurements, samples were cooled down to 100 K in a 3.2 mm sapphire rotor. DNP samples were prepared using AMUpol[25] and buffers as described above (see Preparation of [^{13}C, ^{15}N] A431 vesicles with EGF for ssNMR/DNP measurements). The DNP enhancement was measured by overlaying HC CP/MAS spectra recorded with and without microwave irradiation. Two and three-dimensional NC correlation spectra were recorded using SPECIFIC-CP[26] ^{15}N-^{13}C transfers. Homonuclear (^{13}C,^{13}C) transfers were established using PARIS[27] or spin-diffusion[28] blocks. ^{1}H decoupling using SPINAL64 was employed during evolution and detection periods except in HC HETCOR (Supplementary Fig. 17) where GARP[29] decoupling was employed at 10 KHz.

3.4 Results

We used A431 cells, known to exhibit a high expression level of EGFR[30], to produce membrane vesicles amenable to electron microscopy (EM), super-resolution microscopy (dSTORM) and ssNMR/DNP studies (**Fig. 1 a, b &c**, see Material and Methods). The isolated vesicles are of 50-250 nm size as seen by cryo-Electron microscopy experiments (**Fig. 2**). Super-resolution light microscopy (dSTORM) experiments using anti-EGFR nanobodies (**Fig. 1c,** See Material and Methods) verified localization of EGFR along the membrane of isolated vesicles (**Fig. 1c**) and we observed ligand-induced phosphorylation (**Fig. 1d**), confirming high-level expression of functionally active EGFR.

To determine the changes of EGFR dynamics in native membrane environment upon addition of EGF by ssNMR, we prepared [^{13}C, ^{15}N] labeled EGFR-containing vesicles by growing A431 cells on a medium containing a [^{13}C, ^{15}N] labeled algae mixture (**Fig. 1a**). The validity of this approach was confirmed by examining the effect of variations in temperature and the addition of EGF in the resulting 1D and 2D ssNMR spectra (**Figs. 3 to 7**).

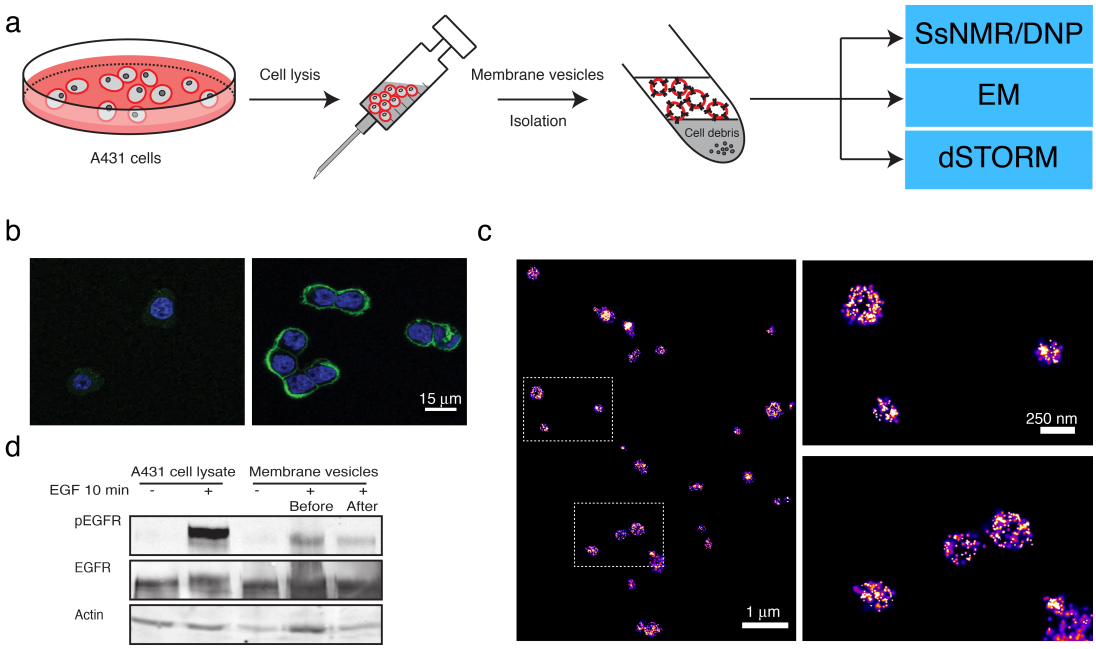


Fig. 1: a) schematic presentation of the preparation of A431 membrane vesicles. For EM/dSTORM studies cells were grown on DMEM medium while for ssNMR studies, A431 cells were cultured in [^{13}C, ^{15}N] labeled DMEM medium (~ 20 plates were required for 1 ssNMR/DNP sample, see Materials and Methods). Cells were scratched from the plates and lysed by passing them through a syringe for few times. After removal of the unbroken cells and cell debris by spinning the cell lysate at low speed, the membrane vesicles were spun down at high speed and loaded into an ssNMR rotor. b) Confocal microscopy of A431 cells (right) and EGFR negative cells NIH 3T3 clone 2.2 murine fibroblasts (left) incubated with Alexa488 tagged EGF (in green). Blue represents DAPI staining of the nuclei. c) dSTORM reconstruction of A431 vesicles stained with anti-EGFR Alexa647 conjugated nanobody (left) and two magnified 3-µm^2 areas (right). Scale bars indicate 1 µm and 250 nm, respectively. d) Phosphorylation assay of A431 plasma membrane vesicles to detect phosphorylated EGFR (pEGFR) with a specific antibody (see Material and Methods). A431 cells were incubated at 37° C with (+) or without (-) EGF for 10 minutes. For membrane vesicles samples, either A431 cells were incubated at 37° C for 10 minutes with EGF (+), followed by vesicle preparation (sample referred to as "before") or vesicles were first prepared from A431 cells, after which they were incubated at 37° C without (-) or with (+) EGF (sample referred to as "after") as described in detail in the M&M.

While signal intensities in frozen samples in the absence or presence of EGF were very comparable (**Fig. 6**), ssNMR intensities differed at ambient temperature with a clear increase in signal after addition of EGF (**Fig. 7**), clearly indicating a ligand-induced rigidification of EGFR. The observed changes were in

agreement with chemical shift predictions based on previous structural studies indicating that our ssNMR spectra were dominated by the signals stemming from well-folded EGFR (**Fig. 8** & table 1). Moreover, we observed that ssNMR signals from our vesicular samples remained constant during extended measurement periods consistent with intact protein preparations (**Fig. 9**). Next to the well-folded protein signals, we also observed flexible random coil signals from unstructured protein regions (such as the EGFR C-ter), (**Fig. 10**) and other molecules, like lipids and sugars (**Figs. 8 & 10**).

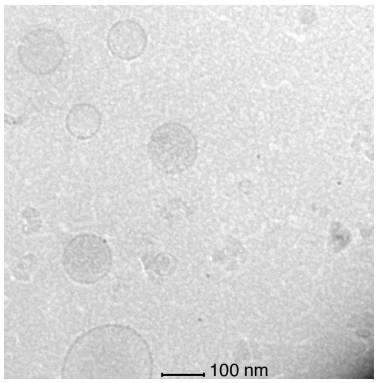


Fig. 2: Cryo-EM of A431 membrane vesicles. Sample was observed without chemical fixation or contrast. Membrane vesicles shown here have the same size as that observed with dSTORM.

Although spectral overlap precluded an analysis of the entire 1210 amino acid receptor, two-dimensional (^{13}C,^{13}C) (DQSQ) (**Fig. 3a**) and (^{15}N,^{13}C) (NCa) spectra at ambient temperature provided sufficient spectral resolution to investigate conformational changes in the Thr/Ser segment (**Fig. 3b**) as well as the spectral regions containing Ala (**Fig. 3c**), Pro (**Fig. 3d**) and Gly (**Fig. 3e**) signals. Using standard secondary chemical-shift values[31] we could distinguish spectral regions characteristic for alpha-helical (red boxes), random-coil (rc, black boxes) and beta-strand (blue boxes) for backbone Ca (dashed line) and Cb (solid line) resonances and we estimated, based on the available structures, the number of signal contributions for each subdomain from high (red) to low (green). Before EGF binding, we observed in the above mentioned regions mainly signals compatible with the TM and KD domains, evidenced by α-helical Ser and Thr signals (abundant in KD and KD & TM, respectively). For Thr we also observed β-strand/rc chemical shifts, but these are randomly distributed over the different EGFR sub-domains (**Fig 3b**). In addition, Ala signals in β-strand conformation, which are most abundant in the ECR (Domain I), were largely absent before addition of EGF (**Fig 3c**).

Interestingly, Ala β-strand signals appeared after addition of EGF, suggesting that the ECR rigidifies upon ligand binding. We also saw a large increase in Gly signals after addition of EGF, of which more than half is found in the ECR. Taken together our analysis suggested that before ligand binding only the KD and the TM exhibit rigidity and that addition of EGF leads to an overall reduction in receptor dynamics both for the ectodomain (domains I-IV) as well the intracellular CT domain. This latter notion would also be supported by a strong increase in signal intensities in Pro (**Fig. 3d**), Gly (**Fig. 3e**) and Asn/Asp (**Fig. 11 & 12**) regions that are largely stemming from the CT and the ECR. Note that such a global analysis of our ssNMR spectra does not allow us to draw general conclusions about the much smaller TM and JM domains.

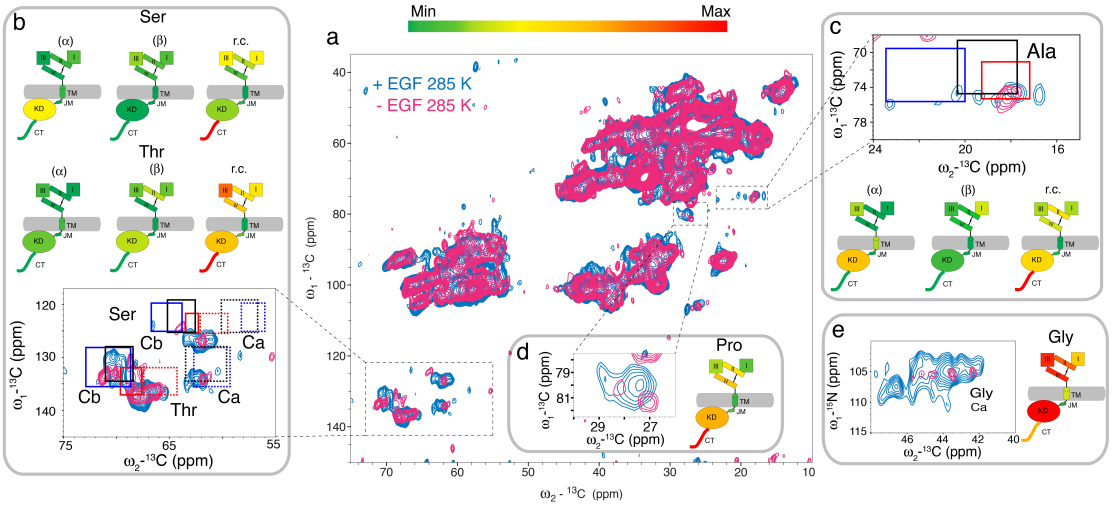


Fig. 3: a) Schematic view of EGFR domains highlighting the sequential correlations expected in [^{13}C-F, M and ^{15}N-L, T] labeled sample. Color-coding stands for specific backbone conformations as described in the main text. The effect of receptor activation as seen by 1D and 3D DNP-ssNMR on MFTL EGFR. 1D (b) and 2D (c,d) slices (b) of the 3D NCOCX experiments with EGF (brown) and without EGF (grey). c) The identified methionines (Met1, Met2, see Table 2 in SI) in our 3D DNP NCOCX spectra with EGF (brown) and without EGF (grey) are indicated in the met 2D slices. Red, black and blue boxes represent the chemical shift values of helical, random coil and beta-sheet Cb. Dotted boxes are for Ca with the same color coding indicated for the dotted boxes. Dotted purple lines connect the identified Ca, Cb correlations as a guide to the eye. Dotted black lines connect the two Cb correlations of the identified methionines through the different slices. It is noteworthy that some contributions of methionine Cg cannot be excluded as they resonate in a close spectral region to the methionine Cb. d) is the same as mentioned in c) but for phenylalanines.

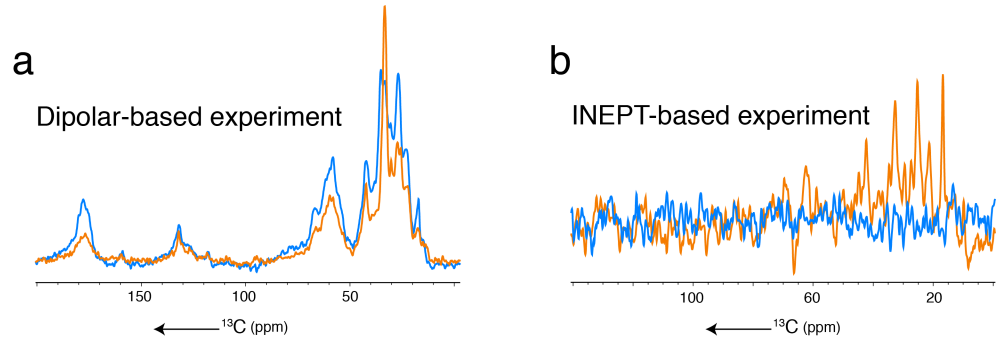


Fig. 4: a) HC CP of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles without EGF at 285K (Orange) and at 253 K (Blue). b) INEPT-based experiment of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles without EGF at 285 K (Orange) and at 253 K (Blue).

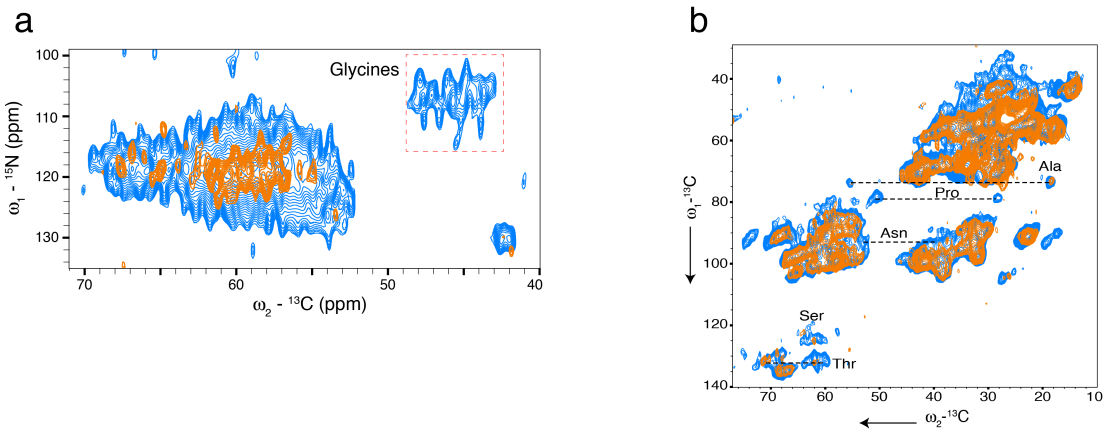


Fig. 5: a) NCa of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles without EGF at 285 K (Orange) and at 253 K (Blue). b) DQSQ of the same sample.

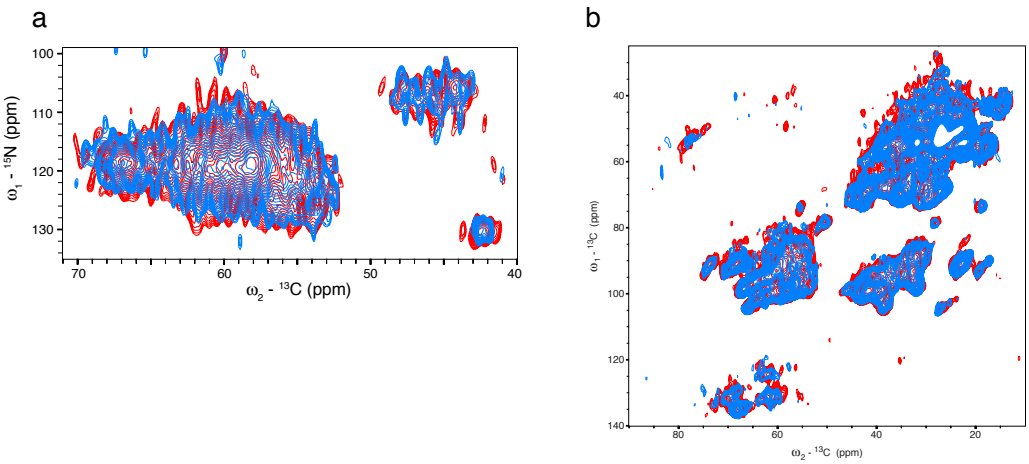


Fig. 6: a) NCa experiment of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles with EGF (red) and without EGF (blue) performed at 253 K. b) DQSQ of the same samples of a).

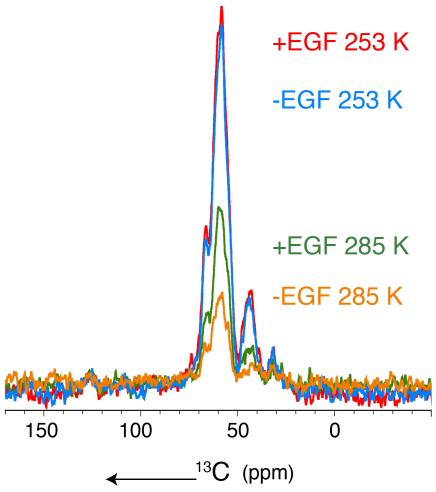


Fig. 7: First increment of NCa of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles without EGF (blue at 253 K and orange at 285 K) and with EGF (red at 253 K and green at 285 K).

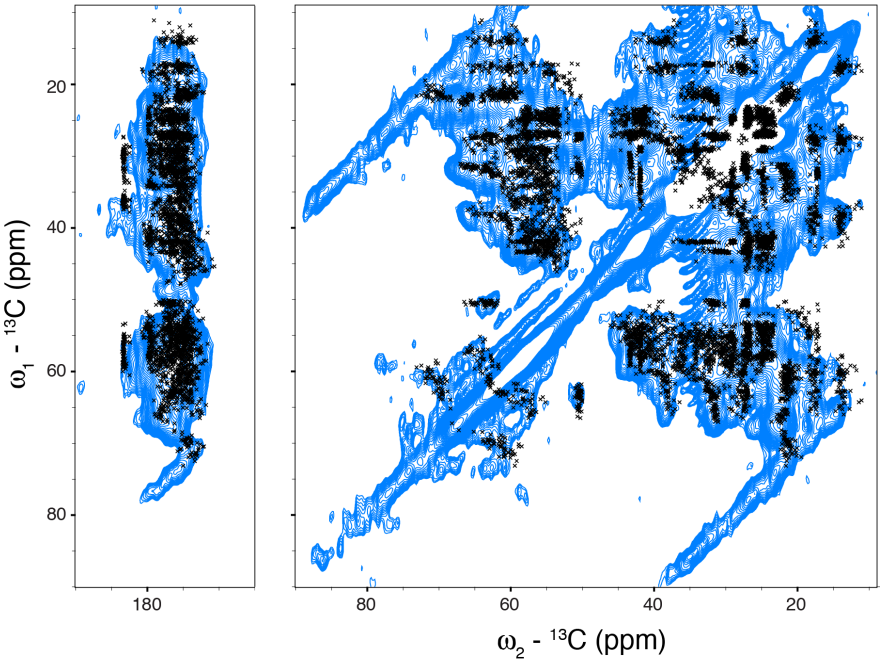


Fig. 8: CC PARIS experiment of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles without EGF performed at 253 K. Black crosses represent FANDAS predictions of EGFR based on the different available crystal structures (see table 1). Note that the peaks at ~ 70 ppm are stemming from lipids.

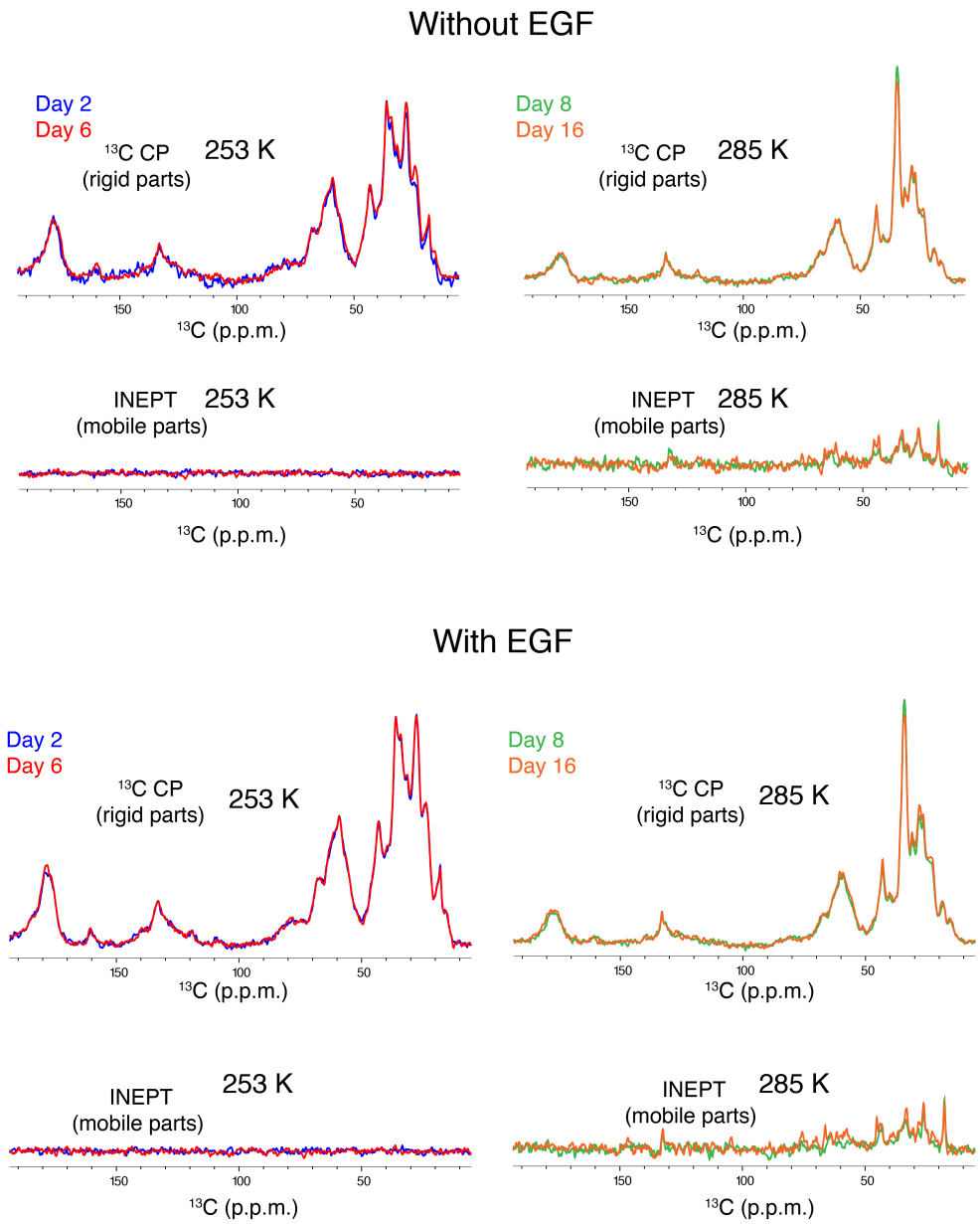


Fig. 9: 1D HC CP and INEPT on fully [^{13}C, ^{15}N] labeled A431 vesicles with and without EGF performed during the course of 2D experiments recording at the 700 MHz spectrometer. At the end of measurements (day 16), both samples showed the same profile as in the beginning of the measurements.

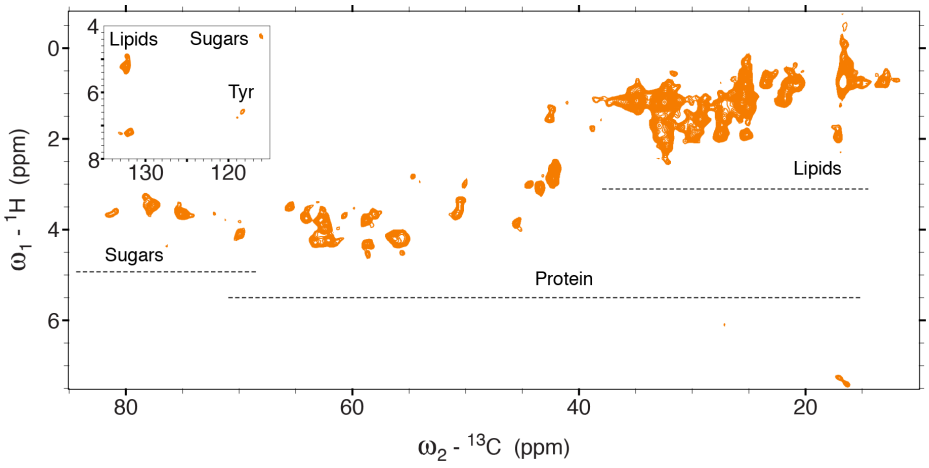


Fig. 10: 2D INEPT experiment of fully [^{13}C, ^{15}N] labeled A431 membrane vesicles without EGF performed at 285 K showing mobile molecular components.

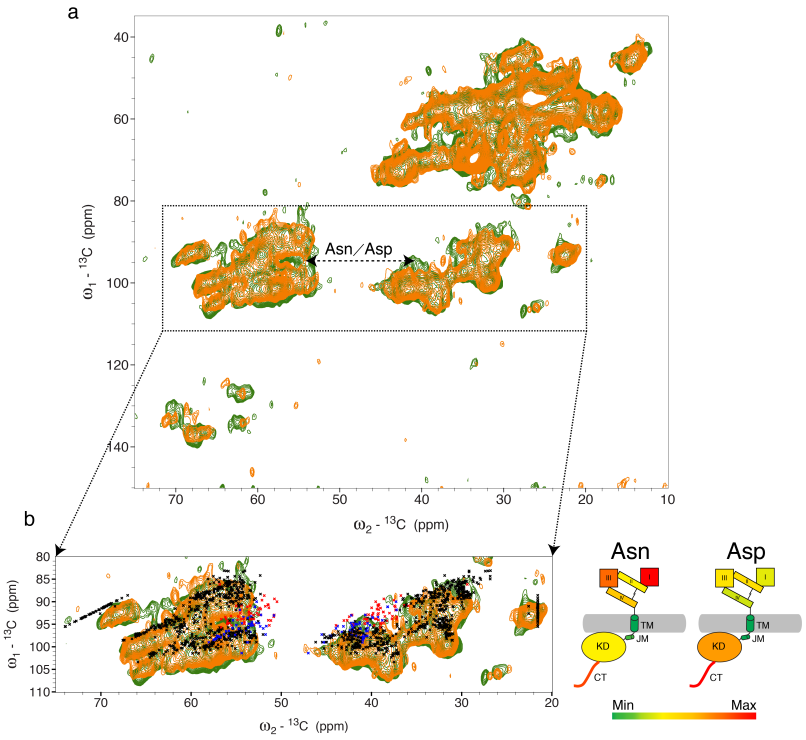


Fig. 11: a) DQSQ of fully [^{13}C, ^{15}N] labeled A431 plasma membrane vesicles without EGF (orange) and with EGF (green) at 285 K with the black arrow highlighting the region where Asn/Asp are prominent. b) Zoom-in of the same experiment in (a) with FANDAS predictions of EGFR made based on the crystal structures (Table 1 in the SI). Red and blue crosses represent Asp and Asn predictions respectively. Black crosses represent the prediction of the other amino acids. Schematic representations of EGFR are color-coded according to the distribution of Asn and Asp amino acids in the different domains of EGFR with red being the highest and green the lowest numbers as indicated in color bar code.

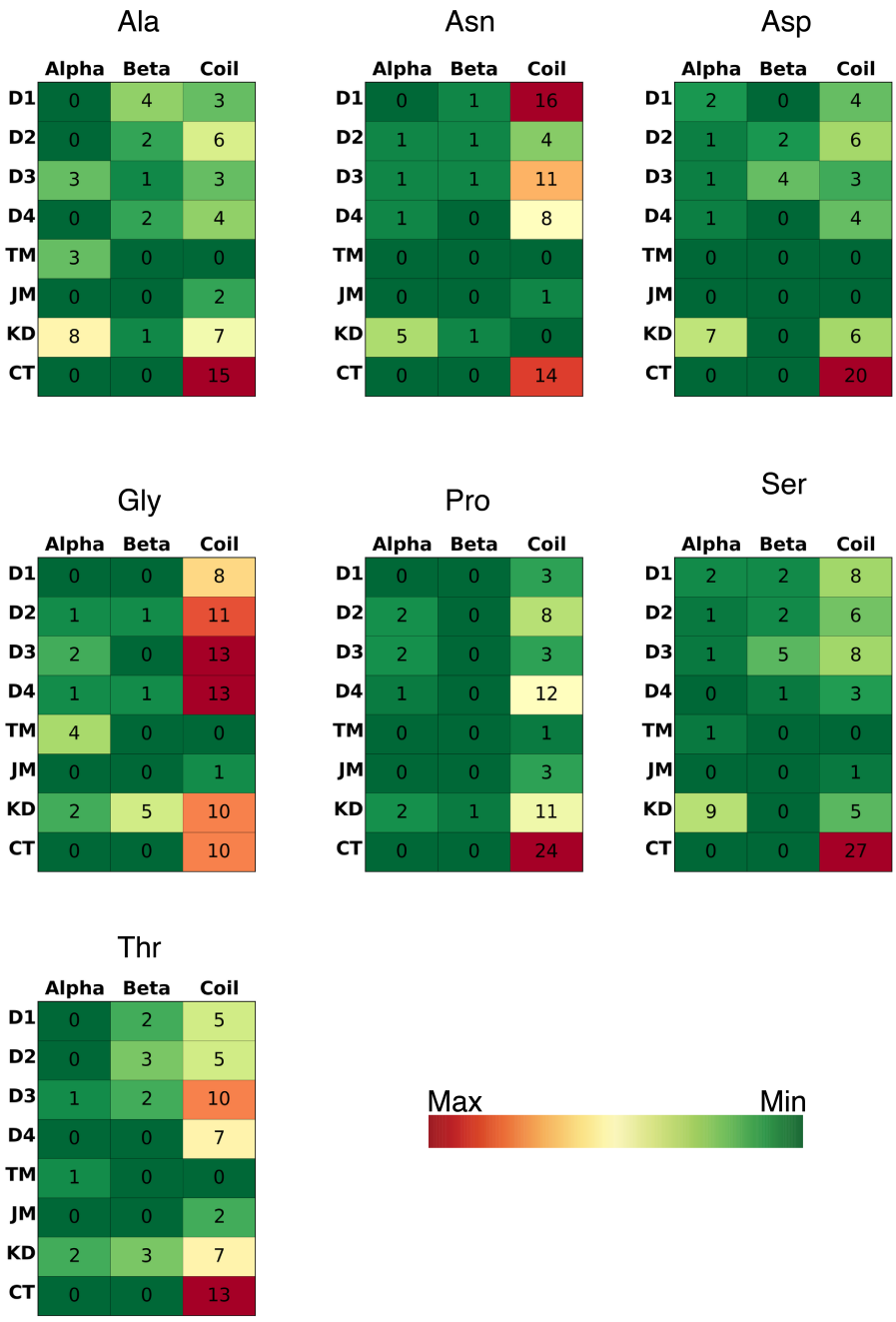


Fig. 12: Heat maps of the Ala, Asn, Asp, Gly, Pro, Ser and Thr distribution in EGFR with red being the maximal and green the minimal numbers (see M&M).

In order to obtain site-specific information of the different domains of EGFR, we produced specifically labeled A431 membrane vesicles with ^{13}C-Met, ^{13}C-Phe, ^{15}N-Thr and ^{15}N-Leu (MFTL-labeled EGFR, **Fig. 13**). As shown before[16], this strategy introduced atomic probes that lead to residue-specific sequential correlations in inter-residue ssNMR, the so called NCOCX (see, e.g., Ref.[16]), experiments. This approach generates in total nine sequential residue pairs located in extracellular domains DI- DIII, the intracellular tyrosine kinase domain (KD) and the C terminus (**Fig. 14a**). Due to a limited signal to noise ratio at higher temperatures (**Supplementary Fig. 1**), we resorted to DNP experiments which significantly increase NMR signal intensity via electron polarization[32] (see Material and Methods, **Supplementary Fig. 2**). The increased sensitivity allowed us to perform and cross validate 2D and 3D NCOCX experiments at 400 and 800 MHz DNP conditions (**Fig. 14 b, c, d, Figs. 15-17**). Again, we made use of standard spectral regions expected for α -helix (red), β-sheets (blue) and random-coil (black) regions. Interestingly, the addition of EGF significantly increased spectral resolution (**Fig. 14b**). In the corresponding 2D slices of the 3D data (**Fig. 14c**, brown, dashed lines), we could identify a set of two Met (CO,Cb) and (CO,Ca) correlations with a dominant peak pair in α-helical conformation (possibly stemming from the KD) and a weaker correlation consistent with rc/ α-helical shifts (ECR domain II, for tentative CS values, table 2). In the Phe spectral region (**Fig. 14 d**, brown), we found evidence for sequential correlations in β–strand (**Fig. 14 d**, top, possibly stemming from DIII) conformation. In addition, the data were compatible with strong rc (CT domain and DI) and α-helical (3 sequential contacts in DI) backbone conformations (**Fig. 14d**, middle and lower panels). Before EGF binding (**Fig. 14c,b**, grey) our ssNMR spectra exhibited broader peaks and suggested the presence of possibly 2 helical correlations in the Met region.

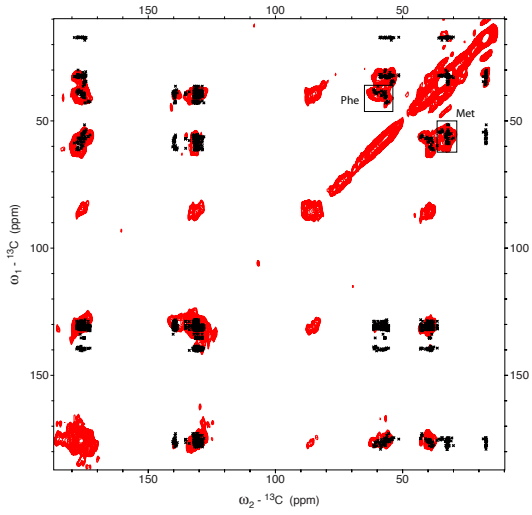


Fig. 13: CC PARIS of MFTL-labeled A431 membrane vesicles recorded on the 700 MHz spectrometer (performed at 253 K). Black crosses represent Met and Phe FANDAS predictions of EGFR.

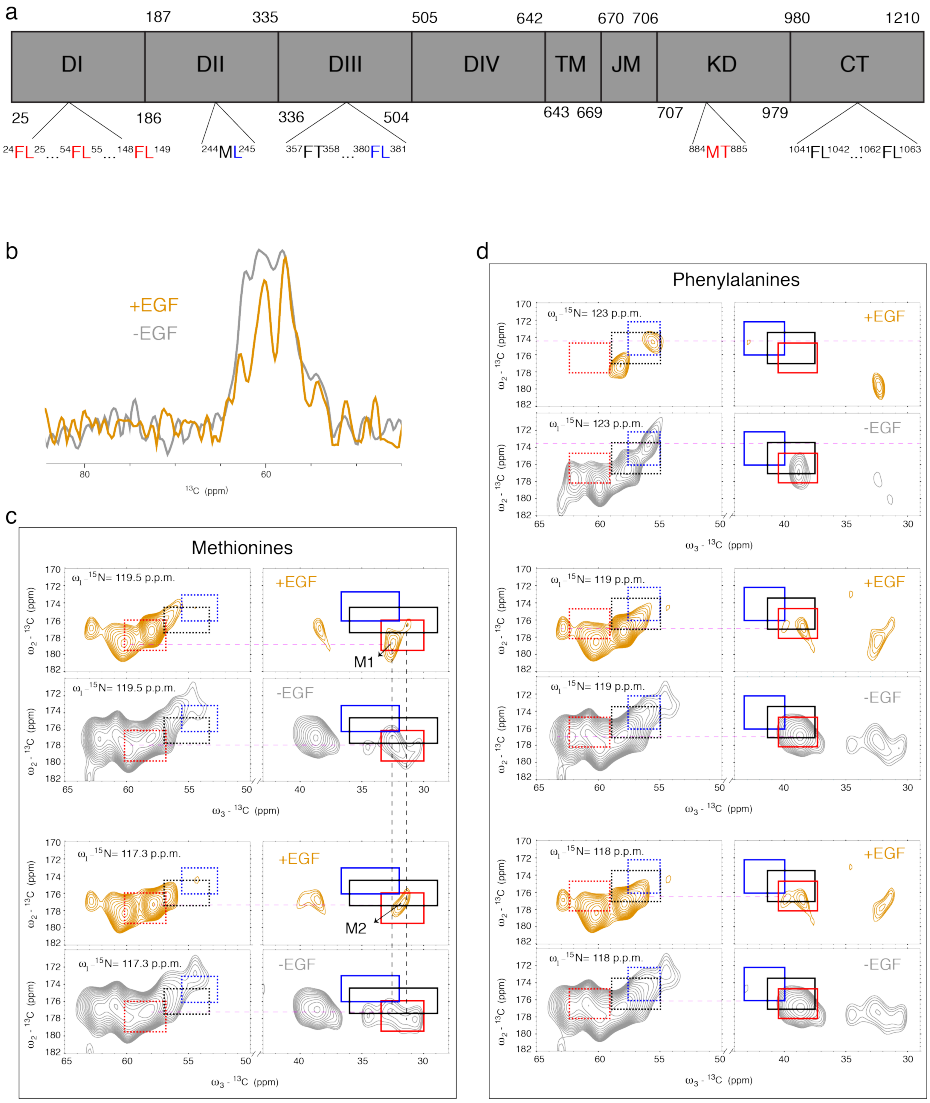


Fig. 14: a) Schematic view of EGFR domains highlighting the sequential correlations expected in [^{13}C-F, M and ^{15}N-L, T] labeled sample. Color-coding stands for specific backbone conformations as described in the main text. The effect of receptor activation as seen by 1D and 3D DNP-ssNMR on MFTL EGFR. 1D (b) and 2D (c,d) slices (b) of the 3D NCOCX experiments with EGF (brown) and without EGF (grey). c) The identified methionines (Met1, Met2, see Table 2) in our 3D DNP NCOCX spectra with EGF (brown) and without EGF (grey) are indicated in the met 2D slices. Red, black and blue boxes represent the chemical shift values of helical, random coil and beta-sheet Cb. Dotted boxes are for Ca with the same color coding indicated for the dotted boxes. Dotted purple lines connect the identified Ca, Cb correlations as a guide to the eye. Dotted black lines connect the two Cb correlations of the identified methionines through the different slices. It is noteworthy that some contributions of methionine Cg cannot be excluded as they resonate in a close spectral region to the methionine Cb. d) is the same as mentioned in c) but for phenylalanines.

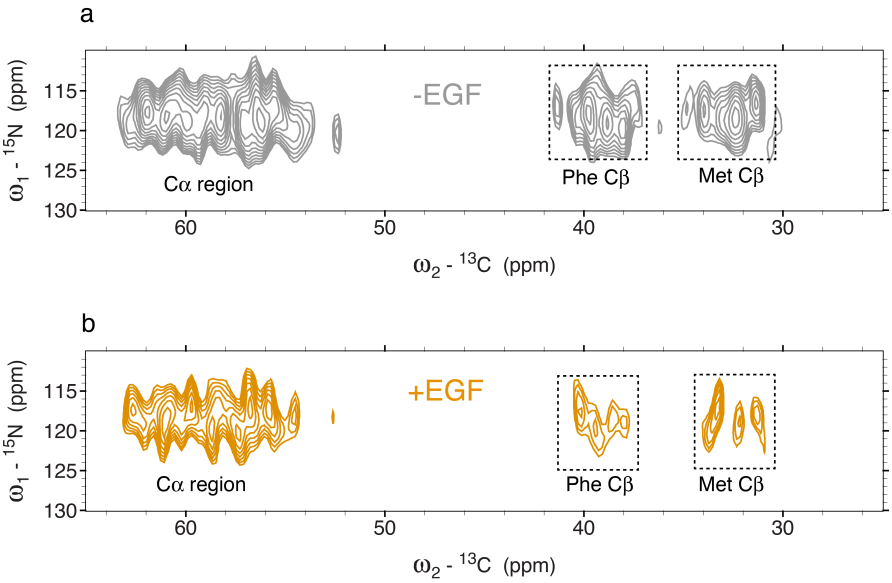


Fig 15: 2D NCOCX of MFTL-labeled A431 membrane vesicles without EGF (a) and with EGF (b) performed on 800 MHz DNP field.

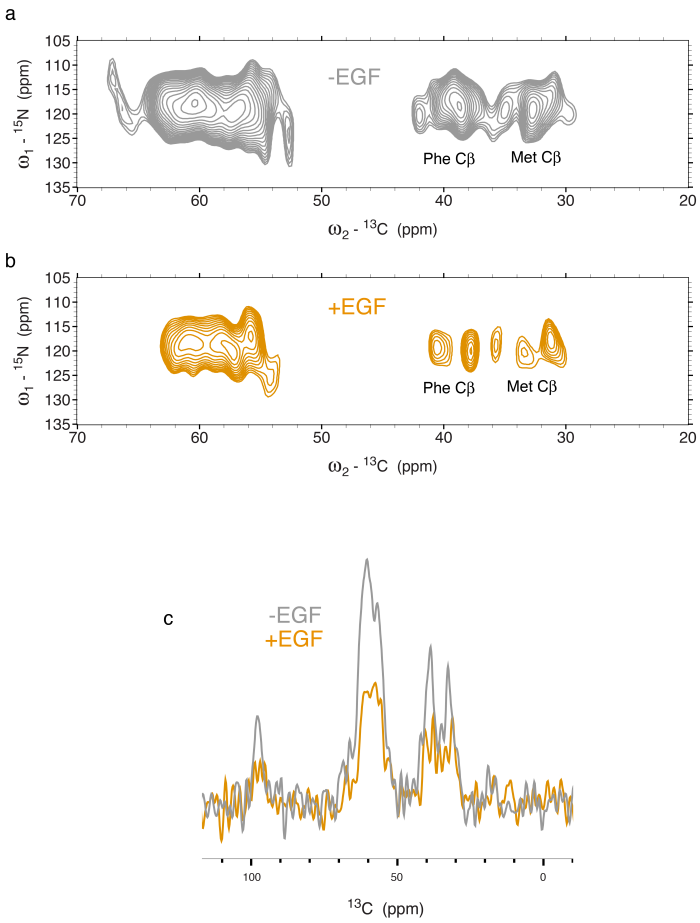


Fig. 16: a) and b) 2D NCOCX of MFTL-labeled A431 membrane vesicles without EGF (a) and with EGF (b) performed on the 400 MHz DNP field. c) 1D slice (row 74) extracted from the 2D NCOCX experiments shown in a) and b).

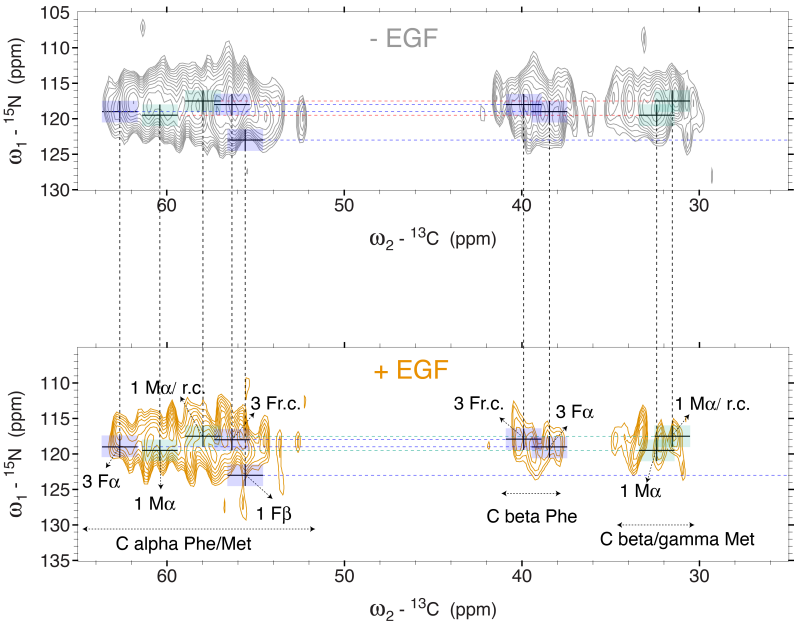


Fig. 17: 2D NCOCX of MFTL-labeled A431 membrane vesicles without EGF (blue) and with EGF (green) performed on 800 MHz DNP field. The crosses represent the chemical shifts identified in the 3D experiments performed on the 400 MHz DNP field with the line width in the ^{13}C dimension is 2 ppm and for ^{15}N dimension 3 ppm. Red boxes highlight the Met correlations and the blue boxes highlight the Phe correlations.

3.5 Conclusions:

Taken together, our ssNMR analysis of isotope-labeled intact preparations at low temperature DNP (LT-DNP) and ambient temperatures suggests that the observed spectroscopic changes are largely due to alterations in receptor dynamics as a result of EGF binding or temperature increase. Before activation, our data is in accordance with a highly dynamic ECR and CT, and a rigid KD, in line with earlier studies suggesting autoinhibitory interactions of the KD and the N-terminal portions of the intracellular JM region with the membrane surface[7,33] (**Fig. 18a & b**). The fluctuations (local and global) of the ECR in the absence of EGF preclude strong interactions of the ECR with the membrane, which is in disagreement with recent MD studies on the EGFR ECR domain[13]. We propose that the presence of the natural composition of the cell membrane and receptor glycosylation in our preparations[34] contribute to the receptor dynamics. Indeed, we observed in our ssNMR experiments additional mobility in other endogenous cellular components including lipids

and sugars suggesting the presence of receptor dynamics at different time scales in our samples (**Fig. 10**).

Our results suggest a model for EGFR activation in which the ECR is present on the cell surface as an ensemble of different conformers including the inactive/closed conformation seen by X-ray studies[3]. Such a notion would for example explain the existence of multiple α-helical correlations in our DNP data. The open conformation (in which the autoinhibitory tether between domain II and IV is released (**Fig. 18a**) enables the ectodomain to form inactive dimers, which could previously be detected for very short time periods[35] in line with the dynamic state of the ectodomain. EGF binding can occur to all conformers, including the open conformation. In this model, EGF does not induce a conformational change of the receptor, but rather stabilizes the open conformation, which is preceding receptor dimerization. Such a close relationship between receptor activation and protein stabilization involves the ECR, the C-terminal region as well as the KD domain (with specific residues probed in our DNP experiments using MTFL-labeled EGFR shown in **Fig. 18b**), the latter observations supporting earlier work on oncogenic KD mutants[36,37]. A reduction in global as well as local dynamics would be compatible with the prevalence of active dimers (**Fig. 18b**) and higher oligomers[35,38] and the observed reduction of the CT mobility may be triggered by binding to the KD activation loop or to different signaling proteins, including adaptor proteins and enzymes.

Our presented approach extends the size and complexity of dynamic protein assemblies that can be studied by NMR from small enzymes[10,11] and protein receptors embedded in membrane mimetics[12-14] to large protein clusters in a eukaryotic cell membrane environment. Analogous to emerging signal-transduction mechanisms across cell membranes[12], the concept of an allosteric regulation in which a reduction in receptor dynamics may be sufficient to shift the conformational equilibrium from inactive monomers and inactive dimers to EGF activated or ligand inhibited EGFR populations may also help to dissect the details of negative cooperativity of EGF binding in a natural plasma membrane environment[39]. SsNMR experiments may also help to further refine structure and dynamics of such membrane-embedded EGFR populations, including the D4 region containing glycolysated sites critical for ligand binding[40] and the C-terminal domain of EGFR. Such studies may also provide critical insight into the role of the plasma membrane and receptor dynamics in related

eukaryotic growth factor receptor kinases that play key roles in regulating cellular processes such as proliferation, differentiation or cell survival[8,9].

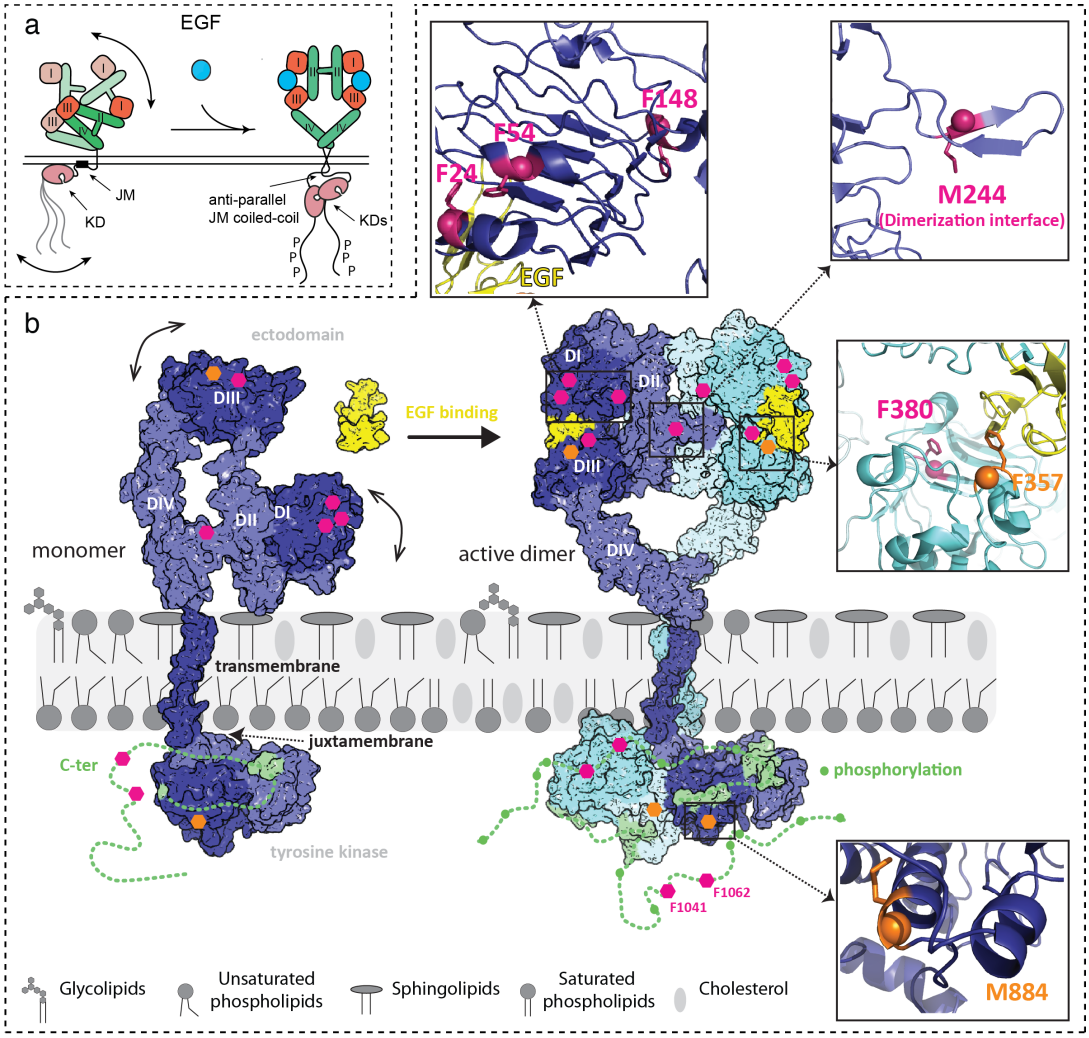


Fig. 18: a) Generic Model of EGFR activation via conformational selection in the ECR. b) At high temperatures (285 K), the unbound receptor exhibits dynamics in both the ECR and CT. Upon binding to the ligand EGF (shown in yellow), the receptor dimerizes and exhibits less dynamics both on global and local scale. Residues probed by ssNMR in the MFTL sample are highlighted in orange (MT & FT residue pairs) and magenta (ML & FL pairs) and zoom-ins show local protein structure. ^{13}C residues (M) and (F) contain side-chains in stick representation and the ^{15}N residues (T and L) are represented as a sphere on the backbone N.

Tables:

Table 1: PDB file codes for the structures used to make EGFR predictions. For parts where no crystal structure is available (some parts of domain IV in the extracellular domain and most of the c-terminal tail) a random coil chemical shift was used.

Domain in EGFR	**PDB ID of the structure used**
Extracellular (Inactive)	1NQL
Extracellular (Active)	1IVO
Transmembrane helix	2M20
Juxtamembrane part	2M20
Kinase domain	1M14
C-terminal tail	1M14

Table 2: Chemical shifts of the two identified methionines in the DNP data, labeled as 1 and 2 in figure 14 b.

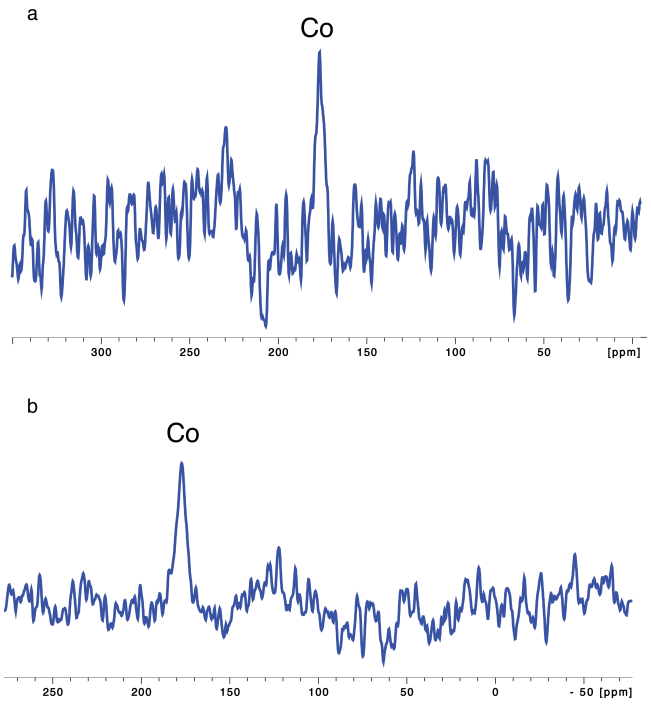
Methionine 1

Ca	Cb	CO	N
60	32.5	178.5	119.5

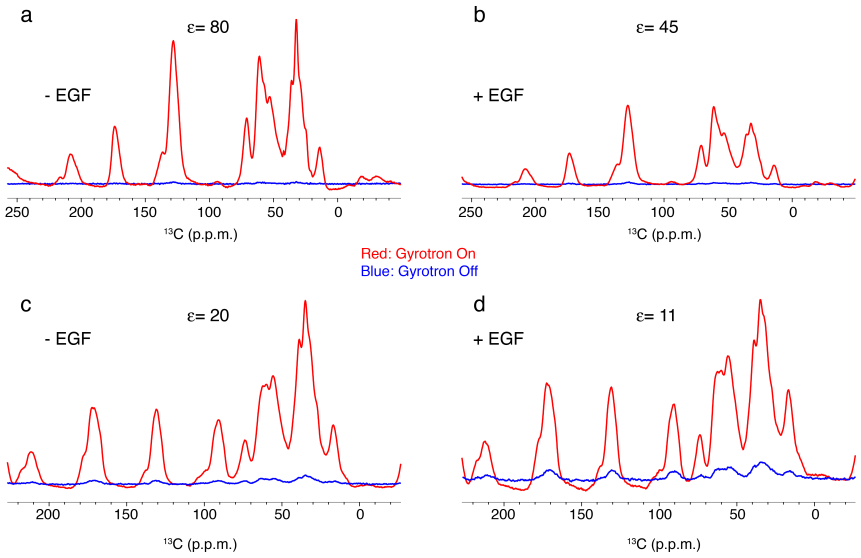
Methionine 2

Ca	Cb	CO	N
57.8	31.9	177.5	117.3

Supplementary Figures:



Supplementary Fig. 1: a) 1D NCo of MFTL-labeled A431 membrane vesicles measured on the 700 MHz spectrometer at (253 K) with 11,000 scans. b) NCOCX of the same sample as in a) measured at the same temperature with 56,000 scans.



Supplementary Fig. 2 DNP enhancements of the MFTL-labeled A431 membrane vesicles. Red spectra are with microwaves on while blue spectra are with microwaves off. a) and b) are for the MFTL-labeled membrane vesicles without EGF and with EGF respectively recorded on the 400 MHz DNP field. c) and d) are for the same samples recorded on the 800 MHz DNP field.

References:

1. Yarden, Y. The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities. *Eur. J. Cancer* **37 Suppl 4,** S3–8 (2001).

2. Arteaga, C. L. & Engelman, J. A. ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. *Cancer Cell* **25,** 282–303 (2014).

3. Ferguson, K. M. *et al.* EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Molecular Cell* **11,** 507–517 (2003).

4. Ogiso, H. *et al.* Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains. *Cell* **110,** 775–787 (2002).

5. Stamos, J. Structure of the Epidermal Growth Factor Receptor Kinase Domain Alone and in Complex with a 4-Anilinoquinazoline Inhibitor. *J. Biol. Chem.* **277,** 46265–46272 (2002).

6. Lu, C. *et al.* Structural Evidence for Loose Linkage between Ligand Binding and Kinase Activation in the Epidermal Growth Factor Receptor. *Molecular and Cellular Biology* **30,** 5432–5443 (2010).

7. Endres, N. F. *et al.* Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. *Cell* **152,** 543–556 (2013).

8. Bessman, N. J., Freed, D. M. & Lemmon, M. A. Putting together structures of epidermal growth factor receptors. *Curr. Opin. Struct. Biol.* **29,** 95–101 (2014).

9. Kovacs, E., Zorn, J. A., Huang, Y., Barros, T. & Kuriyan, J. A Structural Perspective on the Regulation of the Epidermal Growth Factor Receptor. *Annu. Rev. Biochem.* **84,** 739–764 (2015).

10. Kern, D. & Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13,** 748–757 (2003).

11. Kerns, S. J. *et al.* The energy landscape of adenylate kinase during catalysis. *Nat Struct Mol Biol* **22,** 124–131 (2015).

12. Nygaard, R. *et al.* The Dynamic Process of beta2-Adrenergic Receptor Activation. *Cell* **152,** 532–542 (2013).

13. Arkhipov, A. *et al.* Architecture and Membrane Interactions of the EGF Receptor. *Cell* **152,** 557–569 (2013).

14. Sounier, R. *et al.* Propagation of conformational changes during μ-opioid receptor activation. *Nature* **524,** 375–378 (2015).

15. Matsushita, C. *et al.* Transmembrane helix orientation influences membrane binding of the intracellular juxtamembrane domain in Neu receptor peptides. *Proceedings of the National Academy of Sciences* **110,** 1646–1651 (2013).

16. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat Meth* **12,** 649–652 (2015).

17. Lange, A. *et al.* Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. *Nature* **440,** 959–962 (2006).

18. Hong, M., Zhang, Y. & Hu, F. Membrane Protein Structure and Dynamics from NMR Spectroscopy. *Annu. Rev. Phys. Chem.* **63,** 1–24 (2012).

19. Etzkorn, M. *et al.* Complex formation and light activation in membrane-embedded sensory rhodopsin II as seen by solid-state NMR spectroscopy. *Structure* **18,** 293–300 (2010).

20. Koers, E. J. *et al.* NMR-based structural biology enhanced by dynamic nuclear polarization at high magnetic field. *J Biomol NMR* **60,** 157–168 (2014).

21. Mikhaylova, M. *et al.* Resolving bundled microtubules using anti-tubulin nanobodies. *Nat. Commun.* **6,** 7933 (2015).

22. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using μManager. *Curr. Protoc. Mol. Biol.* **Chapter 14,** Unit14.20 (2010).

23. Kremer, J. R. Computer Visualization of Three-Dimensional Image Data Using IMOD. **76,** 71–76 (1996).

24. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Genet.* **23,** 566–579 (1995).

25. Sauvée, C. *et al.* Highly Efficient, Water-Soluble Polarizing Agents for Dynamic Nuclear Polarization at High Frequency. *Angew. Chem. Int. Ed. Engl.* **52,** 10858–10861 (2013).

26. Baldus, M., Petkova, A. T., Herzfeld, J. & Griffin, R. G. Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. *Mol. Phys.* **95,** 1197–1207 (1998).

27. Weingarth, M., Demco, D. E., Bodenhausen, G. & Tekely, P. Improved magnetization transfer in solid-state NMR with fast magic angle spinning. *Chem. Phys. Lett.* **469,** 342–348 (2009).

28. Bloembergen, N. On the interaction of Nuclear Spins in a crystalline lattice. *Physica* **15,** 386–426 (1949).

29. Shaka, A. ., Barker, P. . & Freeman, R. Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Reson.* **64,** 547–552 (1985).

30. McDonagh, C. F. *et al.* Antitumor Activity of a Novel Bispecific Antibody That Targets the ErbB2/ErbB3 Oncogenic Unit and Inhibits Heregulin-Induced Activation of ErbB3. *Molecular Cancer Therapeutics* **11,** 582–593 (2012).

31. Wang, Y. & Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11,** 852–861 (2002).

32. Ni, Q. Z. *et al.* High frequency dynamic nuclear polarization. *Acc. Chem. Res.* **46,** 1933–1941 (2013).

33. Sengupta, P. *et al.* EGFR Juxtamembrane Domain, Membranes, and Calmodulin: Kinetics of Their Interaction. *Biophys J* **96,** 4887–4895 (2009).

34. Kaszuba, K. *et al.* N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. *Proceedings of the National Academy of Sciences* **112,** 4334–4339 (2015).

35. Hofman, E. G. *et al.* Ligand-induced EGF Receptor Oligomerization Is Kinase-dependent and Enhances Internalization. *J Biol Chem* **285,** 39481–39489 (2010).

36. Shan, Y. *et al.* Oncogenic Mutations Counteract Intrinsic Disorder in the EGFR Kinase and Promote Receptor Dimerization. *Cell* **149,** 860–870 (2012).

37. Sutto, L. & Gervasio, F. L. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proceedings of the National Academy of Sciences* **110,** 10616–10621 (2013).

38. Clayton, A. H. A. Ligand-induced Dimer-Tetramer Transition during the Activation of the Cell Surface Epidermal Growth Factor Receptor-A Multidimensional Microscopy Analysis. *J. Biol. Chem.* **280,** 30392–30399 (2005).

39. Arkhipov, A., Shan, Y., Kim, E. T. & Shaw, D. E. Membrane Interaction of Bound Ligands Contributes to the Negative Binding Cooperativity of the EGF Receptor. *PLoS Computational Biology* **10,** e1003742 (2014).

40. Whitson, K. B. *et al.* Functional Effects of Glycosylation at Asn-579 of the Epidermal Growth Factor Receptor. *Biochemistry* **44,** 14920–14931 (2005).

CHAPTER FOUR

Towards proton detected solid-state NMR studies of large membrane complexes in a cellular environment

4.1 Abstract:

^{1}H-detection can enhance the sensitivity of solid-state NMR spectra. Recently, development in the instrumentation of ssNMR and sample preparation has revolutionized proton detection in ssNMR. Here, we endeavor to implement proton detection on a fractionally deuterated cellular-embedded type IV secretion system core complex (T4SScc). Our preliminary results indicate the feasibility of such approach and speak in favor of a folded T4SScc in cellular membranes.

4.2 Introduction:

The high sensitivity of proton-detected experiments, which is a result of their high gyromagnetic ratio compared to other nuclei, makes them a very tempting target for NMR. Proton detection, which is an established method in solution-state NMR, has long been hampered in solid-state NMR due to the strong ^{1}H-^{1}H dipolar couplings. In recent years, significant improvements in ssNMR instrumentation that allowed fast spinning of the sample at the magic angle and also developments in sample preparation that have been originally introduced from solution-state NMR, opened the possibility to detect protons in solid samples[1,2]. Perdeuteration, which means producing protein samples under conditions completely devoid of protons and subsequently introducing accessible amide protons by buffer exchange helps to decrease the proton density and thus the number and strength of ^{1}H-^{1}H dipolar couplings[3–5]. Additionally, fractionally deuteration, in which some of the side chains protons are still preserved can yield information about side-chain contacts and dynamics, which is pivotal to determine the structure of proteins Mance et al, in preparation). Furthermore, proton clouds, that is to say, producing the sample in fully deuterated environment and adding certain labeled and protonated amino acids can help to obtain long-range contacts between the side chains[6]. Moreover, a recent study showed that proton detection for fully protonated samples is also feasible[7].

Previously, we have shown the feasibility to use solid-state NMR to investigate large prokaryotic and eukaryotic membrane complexes in their native environment, the so-called cellular solid-state NMR[8–10]. However, a big hurdle in cellular samples is their low sensitivity

due to the small amount of sample used (usually less than 1 mg). For this reason, combining cellular samples with proton detection, in addition to the interesting information obtained from these experiments, is a step forward to increase the sensitivity for such cellular ssNMR studies.

Here, we give a description of how to prepare cellular samples amenable to proton detection experiments using the bacterial type IV secretion system core complex (T4SScc, see also chapter 2) as an example. For this purpose, we optimized protocols to produce perdeuterated and fractionally deuterated *E.coli* cell envelopes expressing the T4SScc. Elementary ssNMR data on these samples are also presented in this chapter.

4.3 Sample preparation:

In order to prepare [^{13}C, ^{15}N] cellular samples suitable for proton detection, i.e., perdeuterated and fractionally deuterated, we applied the following steps:

1- Transform Bl21 cells double mutant (BL21dm), which lack the outer membrane proteins OmpA and OmpF with pKM101 plasmid encodeing for T4SScc. Cells were transformed by heat shock at 42° C for 35 seconds and then plated on agar plates enriched with 4 g/L glucose.
2- The agar plates were incubated at 37° C over night.
3- Next day, one colony was used to produce precultures of 2 ml each with LB medium. The precultures were incubated at 37° C for few hours until they reached OD_{600}~ 1.
4- Once the LB precultures are at OD_{600} 1, cells were spun down at 3,000 *g* for 10 minutes at room temperature.
5- Each cell pellet from one preculture was resuspended in 25 ml of unlabeled M9 medium (This M9 contains H_2O).
   Note: make four 25 ml M9 media and keep two of them at 37° C and two at 26° C over night. This is to ensure that next day one (or more) of these media is at OD_{600} ~ 1.
6- Cells were spun down from one 25 ml M9 medium with OD_{600} ~ 1 for 10 minutes at 3000 *g* at room temperature.
7- Cells were resuspended in 100 ml M9 medium which has D_2O and unlabeled protonated glucose. The sample was incubated at 37° C until it reached OD_{600} ~ 1 (usually over night).

8- When the cells reached OD_{600} ~ 1, they were spun down and resuspended in 500 ml of M9 medium which has D_2O with unlabeled deuterated glucose. Subsequently the sample was incubated at 37° C until reaching OD_{600} ~ 1.

9- When the cells reached OD_{600} ~ 1, they were spun down for 10 minutes at 3000 *g* at room temperature. After that, the pellet was resuspended as follows:

   - Half of the pellet was resuspended in 250 ml M9 medium containing D_2O, 1abeled $^{15}NH_4Cl$ and labeled protonated glucose (to produce the fractionally deuterated sample).

   - The other half of the pellet was resuspended in 250 ml M9 medium containing D_2O, 1abeled $^{15}NH_4Cl$ and labeled deuterated glucose (to produce the perdeuterated sample).

10- Samples were incubated at 37° C and after 20 minutes induced with 200 ug/L tetracycline (dissolved in 100% ethanol) and they were shifted to 18° C over night.

   - Note: 6 hours after induction, add another dose of tetracycline and antibiotic (in this case chloramphenicol) to the medium.

Next day, the cells were at OD_{600} ~ 1 (1.2). The starting OD_{600} was ~ 1.

For isolation of the cell envelope from *E.coli* we applied the following procedure:

1- Spin the cells down for 15 minutes at 3000 *g* at 4° C.

2- Resuspend with ice-cold PBS and spin the cells down for 15 minutes at 3000 *g* at 4° C.

3- Resuspend the cells with 10-15 ml of T50E2 buffer (50 mM Tris-HCl, 2mM EDTA, pH=8.5).

   - At this stage, the cells were kept at -20° C over night.

4- Next day, thaw the cells on ice; adjust the final volume to 25 ml by adding T50E2 buffer. Add protease inhibitors and DNAses. Do not add lysozyme, as this will digest the peptidoglycan layer.

5- Lyse the cells with a pre-cold French press (at 8000 psi, 4° C) 2-3 times (until the suspension becomes clear).

6- Remove the unbroken cells and cell debris by spinning the suspension at 1600 *g* for 15 minutes at 4° C. repeat this step until no pellet is seen any more (this can be repeated 5-6 times).

7- To isolate the cell envelope, spin the supernatant of step (6) at 160,000 *g* for 9 minutes at 4° C. After this centrifugation, the supernatant represents the protoplasm and the pellet is the cell envelope.

8- Resuspend the cell envelope pellet with 10 mM HEPES containing protease inhibitors and keep the samples at -20° C.

4.4 SsNMR experiments:

A 1.3 mm NMR rotor was filled with cell envelope of fractionally deuterated BL21dm expressing T4SScc. Measurements were performed on a wide-bored 800 MHz instrument (Bruker Biospin, Germany) machine using a magic angle spinning (MAS) rate of 52 KHz. The sample was estimated to be around 30° C. A reference sample of Ubiquitin sample measured under the same conditions.

4.5 Results:

To check if the T4SScc was expressed under the growth conditions required for perdeuteration and fractional deuteration, SDS-PAGE of whole cells (before and after induction) and of the isolated cell envelopes was run (**Fig. 1**).

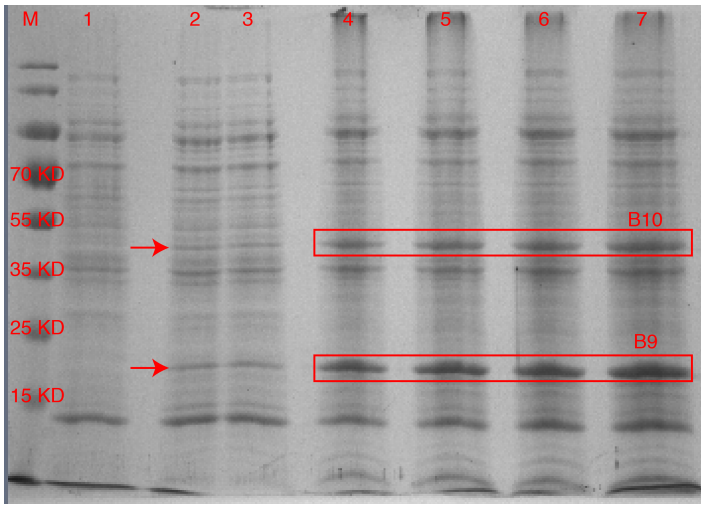


Fig. 1: 10% SDS-PAGE of T4SScc samples prepared for proton detection experiments. M is the marker. The samples are as the following: 1- BL21dm whole cells (before induction), 2- BL21dm whole cells (after induction, fractionally deuterated), 3- BL21dm whole cells (after induction, perdeuterated), 4- Cell envelope of fractionally deuterated induced sample (5 ul of the sample was run on gel), 5- Cell envelope of perdeuterated induced sample (5 ul of the sample was run on gel), 6- Cell envelope of fractionally deuterated induced sample (10 ul of the sample was run on gel), 7- Cell envelope of perdeuterated induced sample (10 ul of the sample was run on gel). The positions of B10 and B9 are highlighted with red arrows and red boxes. Note that B7 is too small to be reliably detected (see chapter 2).

The SDS-PAGE shows high expression of T4SScc in BL21dm cells in both cases, i.e., under the conditions required for perdeuteration and fractional deuteration. However, to improve the purity of the cell envelope samples, they may be washed with PBS two times, to remove any cytoplasmic proteins and contaminants in the samples.

Knowing that T4SScc is expressed in high levels in our cellular samples, we measured preliminary ssNMR proton-detected data (**Figs. 2 & 3**). Figure 2 shows the results of a proton-detected NH experiment suggesting the presence of individually resolved resonances. For further information on the used pulse sequence, see ref.[7]

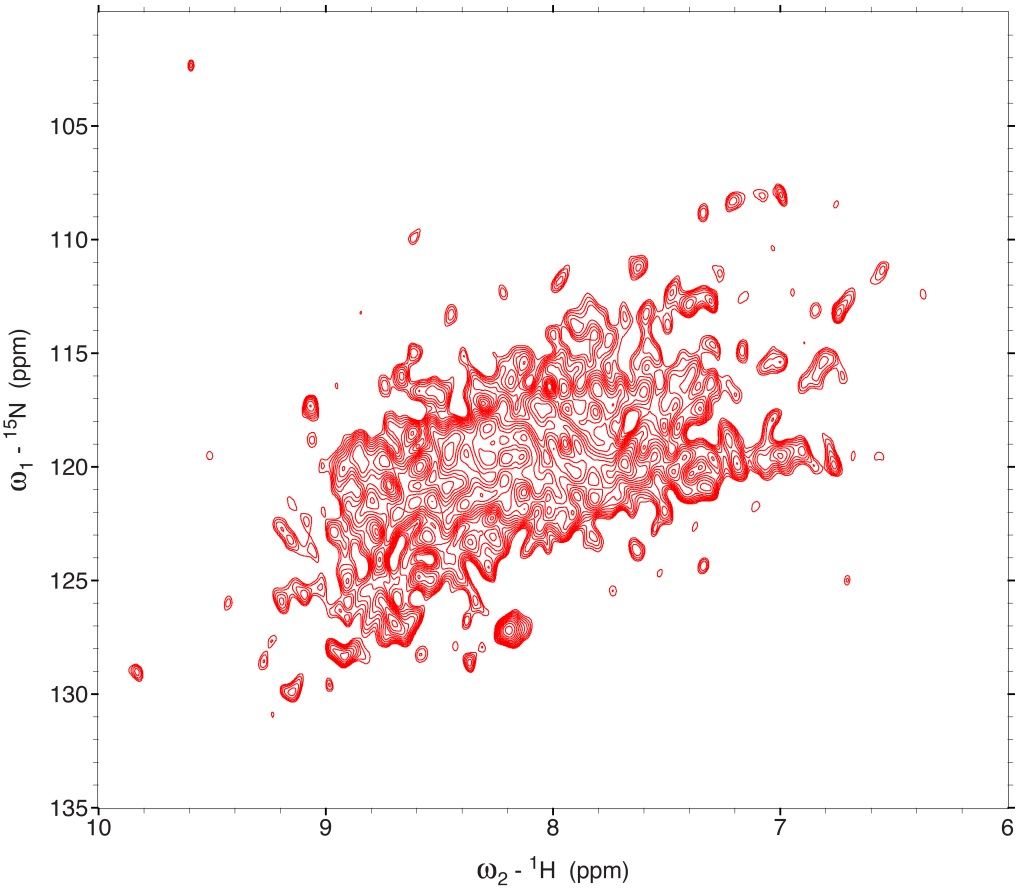


Fig. 2: NH of fractionally deuterated fully [^{13}C, ^{15}N] labeled cell envelope of BL21dm expressing T4SScc recorded at 800 MHz.

In addition, Figure 3 shows the results of a ^{1}H-detected CH experiment, suggesting the presence of resolved correlations with ^{13}C and ^{1}H line width of approximately 1 ppm and 0.2 ppm, respectively. For further information about the pulse schemes, please see (Mance et al., in press).

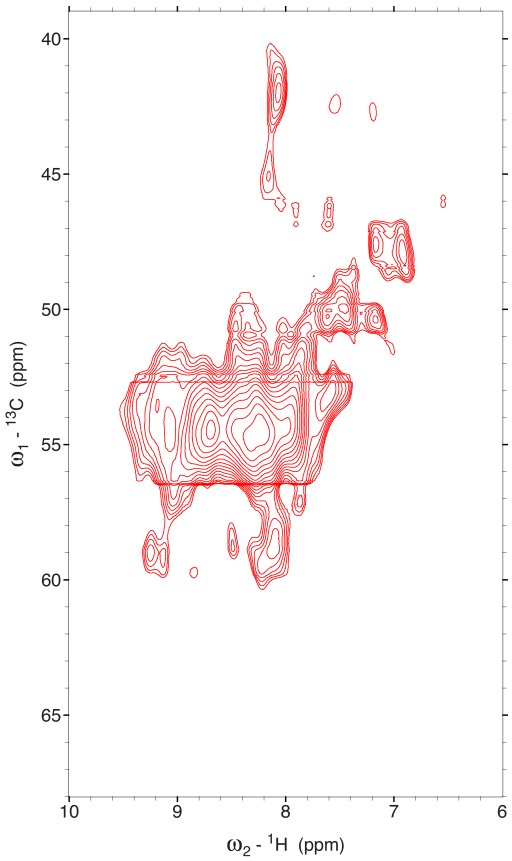


Fig. 3: CH of fractionally deuterated fully [^{13}C, ^{15}N] labeled cell envelope of BL21dm expressing T4SScc recorded at 800 MHz.

Finally, we compare in Figure 4 2D NH correlation data of fractionally deuterated BL21dm cell envelope expressing T4SScc (see Figure 1) with an equivalent spectrum of microcrystalline ubiquitin (Mance et al, to be published) that shows similar spectral dispersion in ^{15}N and ^{1}H dimensions. Since ubiquitin contains protein segments in alpha-helical, random-coil and beta-sheet regions, this observation provides a first indication that also our cell-envelope preparations contain a well-folded T4SScc (**Fig. 4**). However, to confirm that T4SScc is folded further experiments are to be performed. It is worth-mentioning that the peaks S20, G35, I36, A46, G47, E64 in ubiquitin are all located in random coil conformations (in the PDB structure 1UBQ).

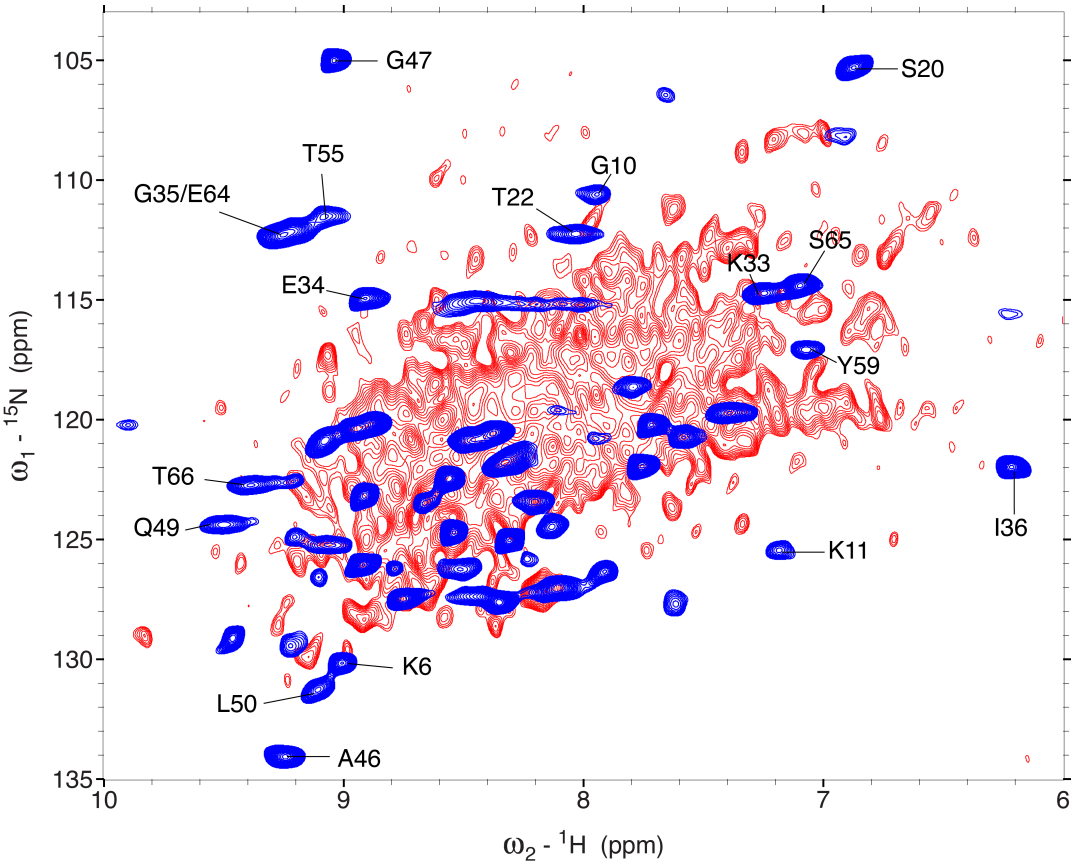


Fig. 4: NH of fractionally deuterated fully [^{13}C, ^{15}N] labeled cell envelope of BL21dm expressing T4SScc (red) and of ubiquitin (blue) recorded at 800 MHz with a selection ubiquitin correlations annotated.

4.6 Conclusions:

Here, we could produce fractionally deuterated and perdeuterated cell envelops of BL21dm cells expressing T4SScc. The fractionally deuterated sample was used to perform ssNMR proton detection studies at the 800 MHz NMR field. We could record NH and CH spectra of the fractionally deuterated samples, which were compatible with the presence a protein signal. Comparing the NH spectrum of the fractionally deuterated T4SScc with that of the purified ubiquitin speaks in favor of the presence of different secondary structure chemical shifts in our NH spectra of T4SScc, which suggests a folded complex in the cell envelope. These experiments provide a promising basis to conduct additional experiments that combine the benefits of using ^{1}H-detected ssNMR experiments with cellular preparations.

References:

1. Reif, B., Jaroniec, C. P., Rienstra, C. M., Hohwy, M. & Griffin, R. G. 1H-1H MAS correlation spectroscopy and distance measurements in a deuterated peptide. *J. Magn. Reson.* **151,** 320–7 (2001).

2. Chevelkov, V., Rehbein, K., Diehl, A. & Reif, B. Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew. Chem. Int. Ed. Engl.* **45,** 3878–81 (2006).

3. Linser, R. *et al.* Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins. *Angew. Chem. Int. Ed. Engl.* **50,** 4508–12 (2011).

4. Ward, M. E. *et al.* Proton-detected solid-state NMR reveals intramembrane polar networks in a seven-helical transmembrane protein proteorhodopsin. *J. Am. Chem. Soc.* **133,** 17434–43 (2011).

5. Knight, M. J. *et al.* Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 11095–100 (2012).

6. Sinnige, T., Daniëls, M., Baldus, M. & Weingarth, M. Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J. Am. Chem. Soc.* **136,** 4452–5 (2014).

7. Weingarth, M., Cruijsen, E. A. W. Van Der, Ostmeyer, J., Lievestro, S. & Baldus, M. Quantitative Analysis of the Water Occupancy around the Selectivity Filter of a K + Channel in Di ff erent Gating Modes. (2014).

8. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51,** 2998–3001 (2012).

9. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4863–8 (2012).

10. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12,** 5–9 (2015).

CHAPTER FIVE

Magic-Angle-Spinning Solid-State NMR of Membrane Proteins

Published in part in: Methods Enzymol 2015;557:307-28

Lindsay A Baker, Gert E Folkers, Tessa Sinnige, Klaartje Houben, Mohammed Kaplan, Elwin AJ van der Cruijsen, Marc Baldus

NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands.

5.1 Abstract

Solid-state NMR spectroscopy (ssNMR) provides increasing possibilities to examine membrane proteins in different molecular settings, ranging from synthetic bilayers to whole cells. This flexibility often enables ssNMR experiments to be directly correlated with membrane protein function. In this chapter, we discuss experimental aspects of such studies starting with protein expression and labeling, leading to membrane protein isolation or to membrane proteins in a cellular environment. We show that optimized procedures can depend on aspects such as the achieved levels of expression, the stability of the protein during purification or proper refolding. Dealing with native membrane samples, such as isolated cellular membranes, can alleviate or entirely remove such biochemical challenges. Subsequently we outline ssNMR experiments that involve the use of magic-angle-spinning and can be used to study membrane protein structure and their functional aspects. We pay specific attention to spectroscopic issues such as sensitivity and spectral resolution. The latter aspect can be controlled using a combination of tailored preparation procedures with solid-state NMR experiments that simplify the spectral analysis using specific filtering and correlation methods. Such approaches have already provided access to obtain structural views of membrane proteins and study their function in lipid bilayers. Ongoing developments in sample preparation and NMR methodology, in particular in using hyperpolarization or proton-detection schemes, offer additional opportunities to study membrane proteins close to their cellular function. These considerations suggest a further increase in the potential of using solid-state NMR in the context of prokaryotic or eukaryotic membrane protein systems in the near future.

5.2 Introduction

Membrane proteins (MPs) are involved in a diverse range of biological functions but pose unique challenges for structural biologists. Their amphipathic, heterogeneous native environment is challenging to mimic *in vitro*, complicating not only isolation of these proteins but also the interpretation of data obtained in non-native settings. The choice of environmental mimetic, such as detergents or synthetic bilayers, can have a significant impact on the structure, function, and stability of MPs. Solid-state NMR spectroscopy (ssNMR) provides a method by which to examine MPs in different membrane systems, ranging from synthetic bilayers to whole cells[1]. This flexibility enables ssNMR experiments

to be directly correlated with complementary approaches probing structure on different length scales or function via functional assays or methods such as electrophysiology.

As with other spectroscopic methods, *sensitivity* is a critical factor for NMR studies on membrane proteins. Unless specific experimental conditions are established (such as low temperature[2]) and/or hyperpolarization methods are used, it is not uncommon for a single ssNMR sample to contain 5-10 mg of the MP of interest including environmental molecules such as water and lipids. At the same time, "NMR-active" nuclei such as (^{13}C or ^{15}N, respectively) must be incorporated into the molecule, usually by adding isotopically labeled molecules to the growth medium. Next to the issue of sensitivity, spectral *resolution* in ssNMR studies on MPs is most easily established by conducting experiments under magic angle spinning (MAS) ssNMR. Under MAS, a randomly packed sample is rotated about an axis at an angle of 54.7 ° relative to the magnetic field (the 'magic angle'), at speeds ranging from 1- 100 kHz, to average out some of the interactions and mimic the effects of molecular tumbling. The rest of this chapter will assume the use of MAS ssNMR, as it is applicable to a wide range of specimens.

Figure 1 summarizes the overall workflow for MAS ssNMR experiments on MPs. Although each step will be discussed in detail below, the choices made at each step have repercussions throughout, so it is worthwhile to consider the experimental path as a whole. In particular, the choice of membrane system for the final sample will depend on the achieved levels of expression, the stability of the protein during purification, and the need for refolding (such as for the purification of over-expressed β-barrels from inclusion bodies discussed below). These considerations underline there is no single general approach that works best for every membrane protein.

With sufficient expression levels of folded protein, correctly inserted into the membrane, native membrane samples, such as isolated cellular membranes, can remove many of the biochemical challenges associated with purifying large quantities of MPs. This approach also has the potential to reduce the volume of cell culture required to produce samples[3], significantly reducing costs. In addition, the choice of membrane system will influence the type of NMR experiments that are best suited to the system – for example, spectroscopic editing of lipid signals from fully isotopically labelled native membrane samples.

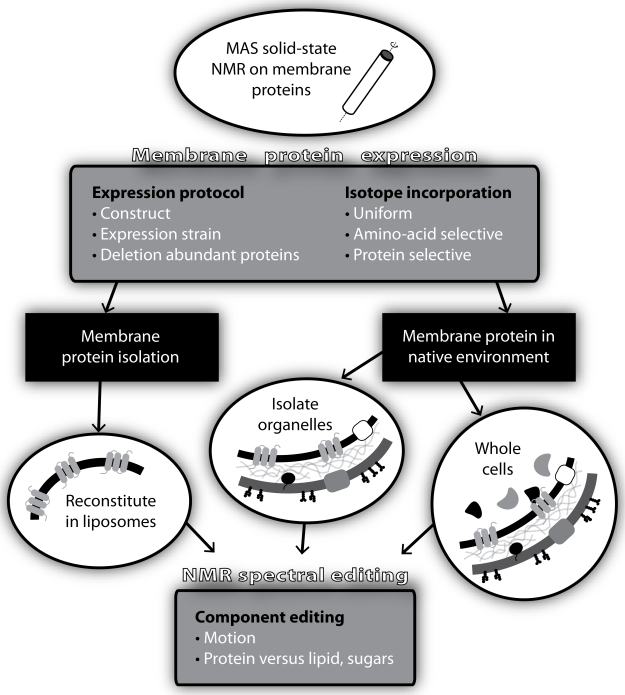


Fig. 1: Experimental pathways for magic angle spinning solid-state NMR of membrane proteins. In general, the first step is to establish expression conditions for the protein of choice. Since NMR requires significant quantities of isotope labeled protein, this step will determine the outcome of subsequent steps. After suitable expression has been obtained, a variety of samples can be prepared, using synthetic or native membranes. Specific NMR experiments can be used to target specific types of molecules in heterogeneous membrane samples.

5.3 Production of recombinant proteins in *Escherichia coli*

Production of recombinant proteins in *E. coli* is a well-established[4] and frequently summarized method for preparation of protein samples for structural and functional studies (see Refs. [5,6]). We will therefore focus on techniques specifically useful for the preparation of membrane protein samples for ssNMR.

5.3.1 Optimization of expression conditions using small scale cultures

The choice of expression conditions can have a significant impact on the yield of membrane proteins, either correctly folded in the native membrane environment or in inclusion bodies. In a benchmark study for a set of soluble proteins, it was observed that relatively subtle changes in expression protocols can easily change expression levels by an order of

magnitude[7]. Since expression parameters are not independent, multi-dimensional optimization of expression conditions is required, and small-scale cultures serve as a convenient way to test many dependent conditions. In our experience, small-scale expression screening experiments serve as a good predictor for large-scale sample production, provided that the conditions used for screening closely mimic the production process[8]. In Table 1 and below, we summarize some of the culture conditions that may have a large effect on recombinant protein expression. For a more detailed protocol for isotope labeling and bacterial growth, see [9,10].

Table 1 Optimization parameters for expression of MPs in *E. coli*

	Starting condition	Possible variations
Bacterial strain	BL21 Rosetta2	BL21, Lemo21, C41/C43, Origami, KRS
Culture medium	M9	LB, SB/TB, autoinduction media
Induction (OD_{600})	0.6–0.8	0.25–2.0
Temperature (°C)	18	16, 20, 28, 37
Induction time (h)	12	1–24

5.3.1.1 Bacterial strain

The choice of bacterial strain can critically affect both expression and solubility. For the expression of eukaryotic proteins in *E. coli,* it is crucial to use strains that overexpress tRNAs that are frequently used in eukaryotes but not in *E. coli* (e.g. Rosetta, EMD Millipore). The so-called Walker strains C41(DE3) and C43(DE3), derived from BL21 cells, often permit expression of membrane proteins in high amounts with lower toxicity[11]. A study by Wagner *et al.* suggested that mutations in these cells affect the expression of the T7 polymerase[12], leading the authors to engineer a BL21-derived strain called Lemo21 (New England Biolabs), which allows for regulation of T7 polymerase activity via controlled expression of T7 lysozyme, its natural inhibitor. However, the gene for T7 lysozyme is under the control of a rhamnose-inducible promotor that is inhibited by glucose, complicating its use when growing in M9 media. We have seen improved expression in LEMO21 cells in M9 cultures for a membrane protein, but only under specific culture conditions.

For cellular preparations, a BL21 strain that is deficient in the genes for the highly abundant outer membrane proteins OmpA and OmpF can be used to reduce background signals[13]. However, the use of this strain is not needed when specific labeling is achieved by treatment with rifampicin (as described in Section 5.3.2).

5.3.1.2 Preculture conditions

Since membrane proteins are often toxic for the host strain, it is important to prevent expression of the protein prior to induction (see also Ref[10]). In most cases, both T7 polymerase and recombinant protein expression are suppressed by the lac repressor. To keep the lac repressor in its active state, 0.5-1% (w/v) glucose can be added during transformation and culturing. Similarly, we found that for toxic proteins, bacteria kept in the exponential growth phase at low densities throughout all precultures exhibit better expression levels during culturing[14].

5.3.1.3 Culture medium

As described above, isotope labeling is required for sample preparation for NMR and expression is normally done in a defined minimal medium. As such, we generally perform expression screening in minimal medium. Although the overall yield per liter in richer media (Luria Broth (LB), Super Broth, Terrific Broth) is often better, the amount of correctly folded protein per cell is not necessarily so. It is possible that the slower growth rate in minimal media enables more effective protein folding, increasing the expression yield.

For MPs, uniformly labeled samples often are too complex and crowded for resonance assignment. More sophisticated labeling schemes have been described previously[15,16], but general strategies include amino acid specific labels and specific ^{13}C incorporation via metabolic precursors. When adding labeled amino acids to the growth media, it is important to consider the metabolic pathways associated with each amino acid. For example, threonine can be converted to isoleucine via α-ketobutyrate; to label threonine but not isoleucine, $^{12}C^{14}N$ isoleucine may be added to the media in addition to $^{13}C^{15}N$ threonine. For a summary of the metabolic interconversion of amino acids, please see (see, e.g., Figure S5, in Ref. [17]),. In our experience, addition of amino acids to M9 cultures can change growth patterns and expression levels, and expression conditions might need to be re-optimized. Amino acid labeling schemes also may use combinations of singly labeled amino acids (i.e. only ^{13}C or

^{15}N labeled) for assignment of specific sequential correlations (i.e. between ^{13}C in residue (i) and ^{15}N in residue (i+1)). Spectra can also be simplified by using specifically labeled metabolic precursors as a carbon source, such as glycerol labeled at either the C2 or C1/3 positions. These strategies results in each amino acid ^{13}C labeled at specific positions[15].

Lastly, introducing deuterons via D_2O and/or appropriate precursors provides additional routes in ssNMR spectroscopy, from the level of spectral editing to ^{1}H detection schemes to the analysis of water molecules (see section 5.7). In general, the doubling time of *E. coli* in D_2O is significantly slower than in H_2O, and an additional preculture in D_2O should be included to give the bacteria time to adapt before induction of protein expression. D_2O can be introduced between the LB and M9 cultures as deuterated LB, or subsequent to the M9 H_2O culture (see below).

5.3.1.4 Expression conditions

Our experience with both soluble[7] and membrane proteins (**Fig. 2**) reveals that both expression levels and amount of folded protein can be optimized by changing the timing of induction according to the growth phase of the bacteria.

Although induction at an OD_{600} of 0.6-0.8 serves as a good starting point, induction earlier or later in the exponential phase, and occasionally in the stationary growth phase, can result in better expression and folding. Optimization of this condition is particularly important for NMR of native membranes, since the amount of folded protein per cell, not the total amount of protein in a culture, is critical. The length of time between induction and harvesting also needs to be optimized, especially for toxic proteins where cultures can stop growing almost immediately after induction. Lastly, growth temperature at induction, as well as the concentration of the induction agent (such as IPTG), should be optimized. Slower growth due to lower temperature can be beneficial for MP folding.

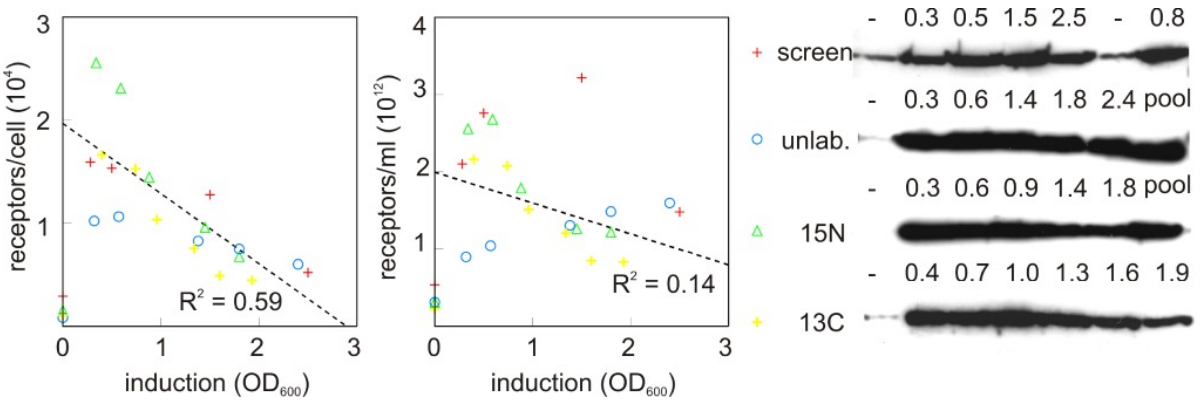


Fig. 2: Effect of growth phase on the expression of membrane protein expression. The GPCR NTR1 (UniProt ID: P20789) was expressed in *E. coli* DH5α and induced at various densities (OD_{600}) and various media: small scale expression screening in unlabeled M9 medium (+), or in 2L flasks containing 500 ml unlabeled M9 medium (o). ^{15}N enriched medium (Δ) or ^{13}C 15N enriched medium (+). The left two panels show the number of receptors as quantified by a radioligand binding assay. The left panel shows the number of receptors per cell, while the middle panel shows the number of receptors per mL of culture. Both the total amount of protein/ml cell culture as well as the amount of protein per cell is negatively affected by the OD_{600} of induction. The right panel shows a Western blot of the total expression of an equivalent volume of cell culture for each condition, as a function of cell density at the time of induction (OD_{600}) as labeled above each row. The label "pool" represents a combined "average" sample from all expression conditions.

5.3.1.5 Evaluation of results

To evaluate small-scale expression screening, SDS-PAGE can be used to compare the total and membrane fractions of a lysed cell culture in a detergent-containing lysis buffer. If expression levels are low, or the MP of interest stains poorly, western blotting can be used. Ideally, a functional assay or ligand-binding assay would be used to quantify the amount of properly folded MP produced, providing a quantitative optimization target.

5.3.1.6 Protocol - Generalized small scale expression tests

1) Transformation – plate on LB +antibiotics+ glucose– overnight
2) Preculture – a few colonies, each in 2 mL LB + antibiotics + glucose – 4-6 h
3) Media change – Once cells have reached OD_{600} of 0.5 -1.0, centrifuge for 3' at 6000 rpm in a benchtop mini centrifuge and remove supernatant.

4) Preculture - Resuspend cells in M9 + antibiotics, and split into 3 x 25 mL cultures. Grow one culture at 20 °C, 30 °C and 37 °C overnight.
5) For subsequent experiments, the precultures with an OD_{600} of ~1.0 can be used to inoculate 50 mL freshly prepared M9 at OD_{600} 0.05 – 0.1.
6) This preculture is then split evenly into 10 x 5.0 mL in 50 ml Grenier tubes, each with a different expression condition.
7) Approximately 30' prior to induction, cultures are transferred to the temperature used for induction, and when the required OD_{600} is reached, expression is started by adding inducers such as IPTG.
8) Harvest cells at different time points after induction by centrifuging at 4000 xg for 15'.
9) Resuspend 100 µL of cell culture in 100 µL SDS sample buffer per OD_{600} 1.0 culture density, for SDS-PAGE analysis. The remainder of the culture can be used for small scale cell envelope preparation, lysis in a suitable buffer or functional characterization.

5.3.2 Selective isotope incorporation

For preparation of NMR samples of membrane proteins in their native cellular environments, it is desirable that only the protein of interest be isotopically labeled, and the cellular background be NMR silent. Rifampicin, an antibiotic that targets the native *E. coli* RNA polymerase [18,19], and the single protein production (SPP) system [20], which makes use of the highly specific endoribonuclease MazF, both inhibit background protein production prior to translation and have been used for NMR in both the solution [21,22] and solid states [3,23]. The SPP method requires replacement of any ACA sequences in the mRNA of the target protein to avoid degradation, while the rifampicin method requires use of a non-bacterial RNA polymerase (such as the T7 system) for expression. A detailed protocol for the use of the SPP system exists[24]. Since the rifampicin procedure is applicable for the generally used T7 expression system, we will focus here on this method. This protocol has been used to produce cellular membrane samples of YidC from *E. coli* (UniProt ID: P25714**)** and KcsA from *S. lividans* (UniProt ID: P0A334) for solid-state NMR spectroscopy[3].

5.3.2.1 Protocol – Selective isotope incorporation for native membrane samples

1) Preculture – 5 mL LB +antibiotics – 8 h at 37 °C
2) Preculture – 50 mL M9 + antibiotics – overnight at 37 °C
3) (If deuteration desired, 50 -100 mL preculture in D_2O M9 +antibiotics) *optional*
4) Culture – 50-500 mL M9 + antibiotics – ^{12}C, ^{14}N, $^{1/2}H$
    a. Inoculate at ~ OD 0.1 and grow to OD ~1.8
    b. Pellet cells (4000 xg, 10-15 m, 20 °C)
    c. Replace media with appropriately isotopically labeled M9+antibiotics
    d. Add IPTG to 1 mM final concentration and grow 15-30'at 25-28°C
    e. Add rifampicin to 100 ug/mL final concentration and grow at 25-28°C overnight, shielded from light *

* NOTE: rifampicin is light sensitive – keep stock solutions and bacterial cultures with it covered in aluminum foil.

For solid-state NMR of membrane proteins, an additional aspect that needs to be taken into account arises for native membrane samples prepared with either rifampicin or the SPP method. Labelling with ^{13}C-glucose results in labeled lipids in the cellular membranes, resulting in high background levels in experiments not involving ^{15}N. The high background levels can be overcome spectroscopically (using magnetization transfer via ^{15}N – see Section 5.7), or by labeling via amino acids instead of glucose and ammonium chloride, or by the use of cerulenin, which inhibits phospholipid biosynthesis[23]

5.4 Isolation of cells and cellular membranes (prokaryotic cells)

5.4.1 From cells to ssNMR samples of MPs

All ssNMR samples begin with a similar isolation protocol. The differences for whole cell, native membrane, purified membranes, and purified and reconstituted samples are indicated in the general protocol below.

5.4.1.1 Protocol – Isolation of cells and membranes

1) Harvest cells by centrifugation at 4000 xg for 15' at 4 °C.

2) Resuspend cells in ~5-10 mL (depending on cell volume) of an ice cold buffer of choice (e.g. 50 mM Tris pH 7.4, 100 mM NaCl, or phosphate-buffered saline (PBS)). Cells may be frozen at -80 °C at this point for several months.

NOTE: Whole cells may be washed with PBS by repeating the centrifugation (step 1) and packed into a ssNMR rotor.

3) Lyse cells with a chilled (4 °C minimum 1 hr prior to lysis) French press (8000 psi) – it normally takes about 4 cycles for complete lysis (can be monitored by measuring OD_{600})
4) Pellet cell debris at 8000 xg for 15 m

   * Optional Step: Isolation of inclusion bodies for further purification by centrifugation at 25 000-100 000 xg for 15 minutes
5) Pellet membranes at ~ 100 000 xg for 1 h

NOTE: At this stage, membranes are ready for protein purification, if desired (See Section 5.6 for a discussion of next steps for purified protein samples).

6) Wash membranes in ~ 20 mL 10 mM phosphate buffer (e.g. pH 6.8)
7) Pellet membranes at ~ 100 000 xg for 1 h

NOTE: At this stage, membranes are ready for separation of inner and outer membranes, if desired (See below for a discussion of next steps for membrane separation).

8) Resuspend in 1 mL 10 mM phosphate buffer (e.g. pH 6.8)
9) Pellet membranes at ~ 125 000 xg for 2-3 h
10) Remove supernatant and pack into a ssNMR MAS rotor for native membrane experiments.

5.4.2 Purification of specific membranes

If expression levels in native membrane samples are limiting, further purification of the membranes containing the MP of interest can improve the sensitivity of ssNMR experiments. Inner and outer *E. coli* membranes can be separated by the following protocol.

5.4.2.1 Protocol – Separation of inner and outer bacterial membranes

1) Harvest membranes as described above (steps 1-7)
2) Prepare a sucrose gradient in 50 mM Tris pH 8.0 buffer. For a 27 mL gradient, layer:
   a. 2 mL of 55 %(w/v) sucrose
   b. 8 mL of 51 %(w/v) sucrose

c. 8 mL of 45 %(w/v) sucrose
d. 5 mL of 36 %(w/v) sucrose
e. Membranes resuspended in 4 mL of 20 %(w/v) sucrose

3) Centrifuge overnight at 100,000 xg, preferably in a swinging bucket rotor (e.g. SW32-Ti (Beckman)). A fixed angle rotor will result in angled layers that will need to be kept flat.
4) Harvest with a syringe:
    a. outer membranes at the interface of the 55% and 51% sucrose steps.
    b. inner membranes at the interface of the 45% and 36% sucrose steps.
5) Wash membranes in ~ 20 mL 10 mM phosphate buffer (pH 7)
6) Pellet membranes at ~ 100 000 xg for 1 h and discard supernatant.
7) Repeat steps 5 and 6 to remove any sucrose and Tris from the samples.
8) Resuspend in 1 mL 10 mM phosphate buffer (pH 6.8)
9) Pellet membranes at ~ 125 000 xg for 2-3 h
10) Remove supernatant and pack into a ssNMR MAS rotor for native membrane experiments.

In our experience, after running sucrose gradients with membrane preparations from *E. coli* BL21-derived cell lines grown in M9, an additional band is observed at the interface between the 51% and 45% sucrose steps. This band seems to be composed of a mixture of inner and outer membranes, and reduces the amount of purified components that can be obtained. This band is not observed from membrane preparations of these cell lines grown in LB.

5.5 Isolation of cells and cellular membranes (eukaryotic cells)

In order to extend cellular solid-state NMR to eukaryotic systems, new labeling protocols have to be optimized for eukaryotic cells especially for cells that grow attached to petri plates. Here, we will discuss how to prepare [^{13}C, ^{15}N] media suitable for culturing human cancer cells, namely, A431 cell line, and how to isolate plasma membrane vesicles and whole cells suitable for ssNMR studies.

5.5.1 preparation of [^{13}C, ^{15}N] medium suitable for A431 cells

To perform ssNMR studies, ^{13}C and ^{15}N nuclei have to be incorporated into the sample. For bacterial systems, this is achieved by growing the cells in fully or specifically [^{13}C, ^{15}N] labeled M9 medium. However, for eukaryotic cells, especially cell lines which grow attached to the plates like A431, another medium has to be prepared that is enriched with ^{13}C and ^{15}N nuclei as discussed below.

5.5.1.1 fully [^{13}C, ^{15}N] labeled medium to label A431 cells

1) Use 1 liter of Dulbecco's Modified Eagle Medium (DMEM) without amino acids, without glutamine, with 1g/L glucose, with sodium pyruvate and 3.7 g $NaCHO_3$.

2) To the above-mentioned medium, add 1 g of [^{13}C, ^{15}N] labeled algae amino acid mixture. This mixture is devoid of cysteine, glutamine and tryptophan and contains the remaining amino acids in the following amounts:

ASX: 8.8%, THR: 3.2%, SER: 4.6%, GLU: 8.2%, PRO: 4.0%, GLY: 8.3%, ALA: 11.9%, VAL: 7.0%, MET: 2.0%, ILE: 6.0%, LEU: 12.2%, TYR: 3.9%, PHE: 5.4%, HIS: 1.2%, LYS: 5.8%, ARG: 5.9%.

3) Add 62 mg/L Cys, 16 mg/L Trp and Gln to 2mM end concentration (these can be added as unlabeled or labeled amino acids).

4) Shake the medium well before filter sterilizing it.

Note: At this stage it is better to stir the medium with a magnet for about 15-20 minutes to ensure that everything is well dissolved in the medium before filter sterilization.

5) Add 100 ml/L of dialyzed fetal calf serum to the medium.

6) Enrich the medium with 1.5 g/L of unlabeled glucose.

5.5.1.2 specifically [^{13}C, ^{15}N] labeled medium to label A431 cells

To make specifically [^{13}C, ^{15}N] labeled medium, then instead of adding the algae mixture in *5.5.1.1*, add the amino acids designated to be labeled as [^{13}C, ^{15}N] and all the remaining amino acids as unlabeled. The amino acids are added in the same amounts mentioned in *5.5.1.1* and then proceed with the same protocol of the fully labeled medium.

5.5.2 Preparation of whole cells and membrane vesicles from A431 cells for ssNMR studies.

In order to produce ssNMR sample of membrane vesicles enough to fill one 3.2 mm rotor, ~ 17-20 150 mm x 25 mm full plates are needed. However, to fill the 3.2 mm rotor with whole cells, ~ 5 150 mm x 25 mm plates are required.

1) Remove the DMEM medium from the plate and wash the cells with PBS.

2) Incubate the cells with PBS containing 2 mM EGTA for 15 minutes at 37° C to help to detach the cells.

Note: Do not use trypsin to detach the cells as this might digest EGFR also.

Note: after incubation with PBS containing EGTA, hit the plate for 3-4 times against the bench. This helps to detach the cells.

Note: from now on keep the cells always at 4° C.

3) Scrap the cells from the plate using a scraper. (Perform this step in the cold room).

4) Spin the cells down at 1000 g at 4° C for 5-10 minutes.

5) Wash the cells down with PBS to remove the EGTA.

6) Spin the cells down at 1000 g at 4° C for 5-10 minutes.

Note: Whole cells are now ready to be packed in to ssNMR rotor.

7) Resuspend the cells in homogenizing buffer (10 mM Tris-HCL pH 7.4, 250 mM sucrose, 1 mM EDTA).

Note: Add protease inhibitors and phosphatase inhibitors to the homogenizing buffer.

8) In order to lyse the cells, pass the cells through a syringe (21G x 1 ½ - 0.8 x 40 mm) for 10 times.

Note: at this stage you can put one drop of the lysed cells on a slide and examine it under the light microscope to check that the cells are lysed (the nuclei will be visible and small moving vesicles).

9) Spin the sample down at 1000 g for 10 minutes at 4° C. A big pellet will be seen which contains the nuclei and the unbroken cells and other cell debris.

Note: repeat step 9 until no pellet is seen any more. This can take 1- 1.5 hours in total.

10) Spin the supernatant of step 9 at 150,000 g for 30 minutes at 4° C.

11) Resuspend the pellet of step 10 with 10 mM HEPES PH 7 (with protease inhibitors and phosphatase inhibitors).

12) Keep the samples at -20° C.

5.6 Purification and reconstitution of MPs for ssNMR

Although unnecessary for native membrane samples, purification and reconstitution of MPs can provide valuable insight into the stability and environmental sensitivity of the MP of interest. As each protein requires a different treatment during purification, we only provide guidelines for parts of the process that are consistent for two broad classes of MPs: those that are produced folded in membranes, and those produced in inclusion bodies. Detailed discussions of chromatography of membrane proteins, for example, can be found elsewhere [25,26,27].

5.6.1 Detergent solubilization of folded membrane proteins

Once membranes have been harvested (see Protocol 4.4.1.1 above), they are solubilized with detergent to produce soluble protein-detergent complexes that can then be purified by traditional chromatography. In general, solubilization is done with high detergent concentrations, which can be an important cost factor, and it is worth optimizing solubilization conditions to improve protein yield and efficiency. Table 2 below summarizes conditions that can be varied during solubilization, and gives starting conditions to try for new target MPs. As for optimization of expression conditions, the best conditions are most easily determined by comparison of yields by SDS-PAGE or functional assays. Many of the parameters that affect solubilization will also affect the stability of the protein during purification; however, the same conditions might not be best for both steps. If protein stability is an issue, consider a second optimization for the subsequent purification process.

Table 2 Common parameters for optimization of detergent solubilization of folded MPs

	Starting condition	Possible variations
Type of detergent	Dodecyl maltoside	CHAPS, Triton, Digitonin, decyl maltoside
Detergent concentration	10 × critical micelle concentration (CMC)	5–20 × CMC
Solubilization time	2 h	30 min—overnight
Solubilization temperature	4 °C	4–25 °C
Salt concentrations	250 m*M* NaCl	0–500 m*M* NaCl, KCl, $MgSO_4$
Glycerol concentration	10% (v/v)	0–30% (v/v)

5.6.2 β-barrel purification and refolding

Overexpression as cytoplasmic inclusion bodies in *E. coli* can lead to high recombinant protein yields. Commonly, β-barrel proteins can be refolded from a denatured state into detergent micelles or preformed lipid bilayers. In general, inclusion bodies are solubilized in a denaturing agent and diluted into a refolding buffer containing detergent. However, buffer conditions and refolding protocols need to be established to obtain a sufficiently good yield of refolded protein. Table 3 shows common parameters to optimize during denaturation and refolding. In the choice of detergent, the critical micelle concentration (CMC) should be taken into account - the CMC should be sufficiently high to enable efficient removal of the detergent during the subsequent reconstitution in proteoliposomes (see Section 5.6.3). In addition, the amount of detergent required for optimal refolding should be cost-efficient. Varying the concentration of detergent in the refolding buffer, as well as the dilution factor of the unfolded protein into this buffer can help mitigate potential expenses.

β-barrel proteins have the additional advantage that folded and unfolded protein can be discriminated on (semi-native) SDS PAGE, a native gel that is run on ice in the presence of 0.2 % SDS in the sample buffer and 0.075 % SDS in the running buffer. The folded protein migrates faster than heat-denatured species [28,29,30,31], quantification of the band intensities leads to estimates of the refolding yield (**Fig. 3a**).

Table 3 Common parameters for optimization of β-barrel refolding

	Starting condition	Possible variations
Denaturant for inclusion body solubilization	8 *M* urea	6 *M* guanidine hydrochloride
Denatured protein concentration	100 μ*M*	20–500 μ*M*
Type of detergent	*N*-dodecyl-*N*,*N*-dimethylamine-*N*-oxide (LDAO)	Octyl glucoside, tetraethylene glycol monooctyl ether (C8E4), diheptanoyl phosphocholine (DHPC), dodecylphosphocholine (DPC), sulfobetaine 12 (SB12)
Detergent concentration	10 × critical micelle concentration (CMC)	5–20 × CMC
Dilution speed	Fast (rapid addition with stirring)	Slow (drop-wise addition with stirring)
Dilution factor (protein solution:refolding buffer)	1:10	1:2–1:100

5.6.3 Reconstitution of detergent solubilized MPs into lipid bilayers

Protein can be reconstituted in proteoliposomes by addition of lipids and removal of the detergent. For the latter, biobeads that absorb the detergent or dialysis are most commonly used. In some cases, biobeads do not result in the appearance of proteoliposomes that are sufficiently large for ssNMR sample preparation, possibly due to the quick removal of the detergent. Dialysis has the disadvantage that it takes longer to remove the detergent compared to biobeads, but depending on the MP this may be the preferential method.

5.6.3.1 Protocol – Reconstitution of detergent-solubilized MPs into lipid bilayers

1) Remove any aggregates by centrifugation for 20 m at 4000 xg and 4 °C
2) If necessary, concentrate the protein with a centrifuge concentrator (Millipore) and remove aggregates by centrifugation for 20 m at 4000 g and 4 °C
3) Dry the lipids (such as E.coli lipids) in chloroform under a stream of nitrogen, followed by > 1h vacuum drying

4) Take up the lipid film in 1 mL reconstitution buffer and incubate the lipids for 5 m at 37 °C (*necessary for certain lipids to reach the liquid crystalline phase*)
5) Resuspend the lipids thoroughly by vortexing to form multilamellar vesicles
6) Add the lipids to the detergent-solubilized protein (*the suspension should become clear as the detergent micelles solubilize the lipids*)
7) a) If removing detergent with dialysis, add reconstitution buffer to dilute below the CMC of the detergent and incubate 30 m at 37 °C

   b) If using biobeads, wash the correct amount (according to manufacturer's recommendations for the amount of detergent in the sample) of beads 3 times with water and then 3 times with reconstitution buffer and add the beads to the protein and lipid solution
8) a) Dialyze against reconstitution buffer (*optional if using biobeads*) at room temperature or 4 °C for 1-7 days, changing the dialysis buffer once or twice a day, until no more detergent (bubbles or foam) is observed when replacing the buffer. The protein-lipid solution has become turbid

   b) If using biobeads, separate the biobeads from the protein-lipid suspension on a gravity flow column
9) Harvest the proteoliposomes by ultracentrifugation for 1-2 h at 100,000 -125,000 xg and 4 °C and pack into a MAS rotor.

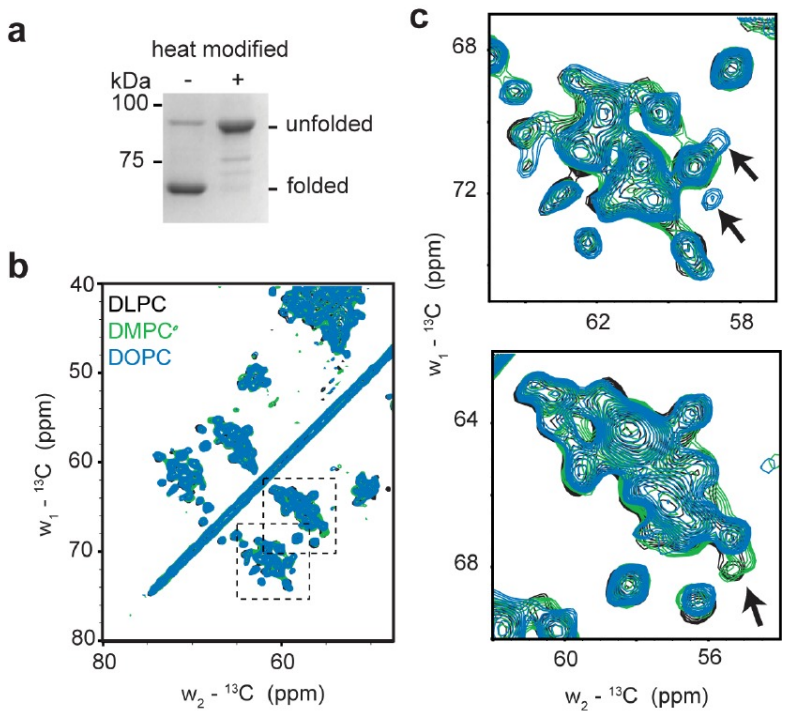


Fig. 3 - Proteins with β-barrel folds such as BamA (β-barrel assembly machinery A, UniProt ID: P0A940) from *E. coli* can be studied by ssNMR. a) Folding of BamA analyzed on semi-native SDS-PAGE. BamA reconstituted in DOPC vesicles at a molar lipid: protein ratio of 25:1 is shown. Native, folded BamA shows a characteristic shift of the electrophoretic mobility from 88 to 70 kDa, compared to heat-denatured protein. b and c) Comparison of BamA reconstituted in different lipid bilayers. Shown in b) is a section of a 2D $^{13}C,^{13}C$ correlation experiment with 30 ms PARIS mixing of BamA in DLPC (black), DMPC (green) and DOPC (blue) at a molar lipid: protein ratios of 25:1. c) Close-ups of the boxed regions in b) reveal small differences in certain regions of the spectra.

For reconstitution, it is useful to test a variety of pure lipids and lipid mixtures (see **Fig 3b,c**, such as *E. coli* polar lipid extract). The lipid-to-protein ratio (LPR), as well as the composition of the reconstitution buffer (pH, addition of salts) should also be screened to give optimal results. Reconstitution yields can be checked on SDS-PAGE by comparing the amount of protein in the proteoliposome pellet and that in the supernatant. For β-barrels, the amounts of folded protein can be estimated on (semi-native) SDS PAGE as described above. Ultimately however, ssNMR experiments such as a 2D $^{13}C,^{13}C$ correlation experiment are most suited (see section 5.7) to judge the success of sample preparation. It is not unusual for the best samples to yield ^{13}C line widths of 0.7-1 ppm. An example is shown in Figure 4 for the bacterial KcsA channel that was reconstituted in Asolectin (**Fig. 4**). The ($^{13}C,^{13}C$)

correlation spectrum was obtained using proton-driven spin diffusion with a mixing time of 30 ms. Figure 4 shows results of an NCA experiment (see Ref. [32] and references therein).

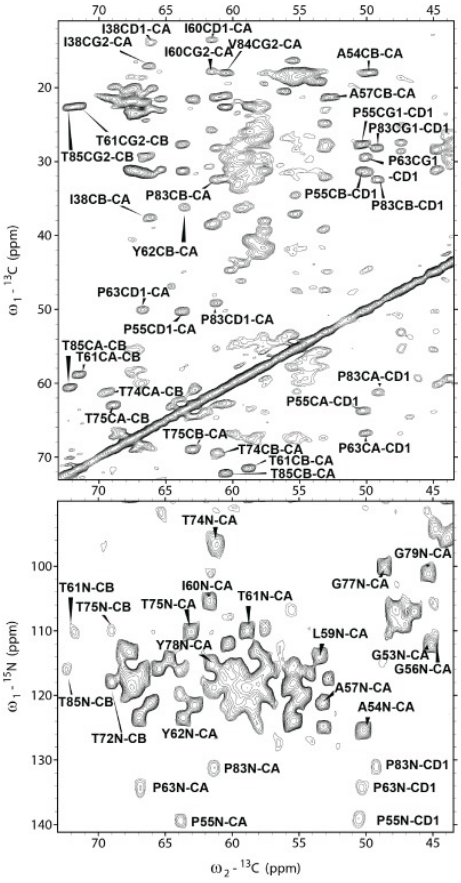


Fig. 4: High resolution spectra of U[^{13}C,^{15}N] labeled KcsA reconstituted in asolectin lipids. a) 2D ^{13}C,^{13}C correlation experiment recorded with 600 us CP time, 30 ms Proton-driven spin diffusion, and 10.921 KHz MAS. b) 2D NCA spectrum recorded with 600 us cp time, 3.2 ms SPECIFIC CP time. Both spectra were recorded at 700 MHz at 273 K.

5.7 Dedicated ssNMR experiments

In the last decade, a series of multidimensional ssNMR correlation experiments have been developed to obtain resonance assignments and structurally characterize solid-phase proteins. In principle, all these experiments are applicable to MPs embedded in synthetic membranes or when associated with cellular membranes/organelles or cells as described above.

To this end, most ssNMR –studies have used experiments that correlate N–C moieties or C–C spin networks in two, three or even four spectral dimensions. With such techniques several

3D structures have been obtained and functional aspects of MPs have been elucidated (see, e.g.,[1] for a recent review). More recently, proton detected experiments have been used for MPs embedded in synthetic bilayers [32,33], and already helped to understand structural aspects related to their mode of action .

With increasing protein size or the presence of endogenous proteinaceous or non-proteinaceous components, dedicated signal filtering and spectral simplification methods become an important aspect in ssNMR. As discussed in Section 5.3, signal overlap or unwanted molecular contributions may be suppressed by dedicated preparation procedures. Alternatively, the use of ^{15}N-edited dimensions greatly reduces the spectral components in non-proteinaceous cellular samples, for example leaving ^{13}C spectra largely containing protein signals or, depending on the experimental conditions and labeling procedures, containing labeled lipids[34] or nucleotides[35]. Another powerful means to deduce structural information in large integral membrane proteins or to zoom on specific molecular components in native preparations relates to solid-state NMR filtering approaches. Different motional scales often characterize signals from water-exposed and transmembrane protein segments. Accordingly, mobile protein loops and rigid transmembrane segments might be detected using through-bond and through-space transfer pathways, respectively[36]. Finally, spectral complexity can also be reduced by "divide-and-conquer" approaches. Here, NMR data are obtained on reference samples such as in synthetic bilayers[13] or in solution to aid the analysis of membrane-embedded MPs by ssNMR. We have used such methods in the case of membrane embedded histidine kinases[37], retinal complexes[38] or, more recently for the outer membrane protein BamA[39] .

On the other hand, correlating signals for the MP of interest to molecular interaction partners, or the surrounding cellular setting offers a unique opportunity to study the supramolecular structure of MPs or the functional roles. As we have described recently[40] there are a variety of experimental approaches that have been successfully applied to, mostly, synthetic bilayer preparations. Combining such experimental data with information from computational resources such as docking or molecular dynamics simulations further enhances the possibility to study MPs and their functional aspects[40].

5.8 Conclusions

The protocols and tips provided in this chapter are only a starting point for studying membrane proteins by ssNMR. One of the strongest advantages of solid-state NMR is its flexibility, and experiments can be designed to investigate highly specific questions unique to each biological system. In addition to this experimental flexibility, there is also the potential to compare a reconstituted system, where tight control over the molecular constituents is possible, to a native system such as a cellular membrane, where biological relevance is ensured but heterogeneity must be addressed. Although we restricted our discussion here to bacterial MP preparations, reports in the literature[41] as well as ongoing work in our own laboratory suggest that future ssNMR studies of MP will also be in many cases possible for eukaryotic systems.

As with other spectroscopic methods, sensitivity and spectral resolution are a critical factor. However, the development of hyperpolarization methods that were already demonstrated in the context of cellular ssNMR studies [35,42] and the further development of ultra-high field NMR systems are likely to further improve the prospects of studying complex MP systems by ssNMR under in-situ conditions.

References:

1. Baker, L. A., & Baldus, M. (2014). Characterization of membrane protein function by solid-state NMR spectroscopy. *Current Opinion in Structural Biology*, *27*, 48–55.

2. Rocchigiani, L., Ciancaleoni, G., Zuccaccia, C., & Macchioni, A. (2011). Low-Temperature Kinetic NMR Studies on the Insertion of a Single Olefin Molecule into a Zr-C Bond: Assessing the Counterion–Solvent Interplay. *Angewandte Chemie International Edition*, *50*(49), 11752–11755. doi:10.1002/anie.201105122.

3. Baker, L. a, Daniëls, M., van der Cruijsen, E. a W., Folkers, G. E. & Baldus, M. Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J. Biomol. NMR* **62,** 199–208 (2015).

4. Studier, F., Rosenberg, A., Dunn, J., & Dubendorff, J. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology*, *185*(1986), 60–89.

5. Acton, T. B., Xiao, R., Anderson, S., Aramini, J., Buchwald, W. A., Ciccosanti, C., … Montelione, G. T. (2011). Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods in Enzymology*, *493*, 21–60. doi:10.1016/B978-0-12-381274-2.00002-9.

6. Kubicek, J., Block, H., Maertens, B., Spriestersbach, A., & Labahn, J. (2014). Expression and purification of membrane proteins. *Methods in Enzymology*, *541*, 117–40. doi:10.1016/B978-0-12-420119-4.00010-0.

7. Berrow, N. S., Büssow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G. E., … Busso, D. (2006). Recombinant protein expression and solubility screening in Escherichia coli: a comparative study. *Acta Crystallographica. Section D, Biological Crystallography*, *62*(Pt 10), 1218–26. doi:10.1107/S0907444906031337.

8. Folkers, G. E., van Buuren, B. N. M., & Kaptein, R. (2004). Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. *Journal of Structural and Functional Genomics*, *5*(1-2), 119–31. doi:10.1023/B:JSFG.0000029200.66197.0c.

9. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification*, *41*(1), 207–34.

10. Studier, F. W. (2014). Stable expression clones and auto-induction for protein production in E. coli. *Methods in Molecular Biology (Clifton, N.J.)*, *1091*, 17–32. doi:10.1007/978-1-62703-691-7_2.

11. Miroux, B., & Walker, J. E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, *260*(3), 289–98. doi:10.1006/jmbi.1996.0399.

12. Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., … de Gier, J.-W. (2008). Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14371–6. doi:10.1073/pnas.0804090105.

13. Renault, M., Tommassen-van Boxtel, R., Bos, M. P., Post, J. A., Tommassen, J., & Baldus, M. (2012). Cellular solid-state nuclear magnetic resonance spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 4863–8.

14. Romanuka, J., van den Bulke, H., Kaptein, R., Boelens, R., & Folkers, G. E. (2009). Novel strategies to overcome expression problems encountered with toxic proteins: application to the production of Lac repressor proteins for NMR studies. *Protein Expression and Purification*, *67*(2), 104–12. doi:10.1016/j.pep.2009.05.008.

15. Filipp, F. V, Sinha, N., Jairam, L., Bradley, J., & Opella, S. J. (2009). Labeling strategies for 13C-detected aligned-sample solid-state NMR of proteins. *Journal of Magnetic Resonance (San Diego, Calif. : 1997)*, *201*(2), 121–30. doi:10.1016/j.jmr.2009.08.012.

16. Verardi, R., Traaseth, N. J., Masterson, L. R., Vostrikov, V. V, & Veglia, G. (2012). Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Advances in Experimental Medicine and Biology*, *992*, 35–62. doi:10.1007/978-94-007-4954-2_3.

17. Sinnige, T., Weingarth, M., Renault, M., Baker, L., Tommassen, J., & Baldus, M. (2014). Solid-State NMR Studies of Full-Length BamA in Lipid Bilayers Suggest Limited Overall POTRA Mobility. *Journal of Molecular Biology*, *426*(9), 2009–21. doi:10.1016/j.jmb.2014.02.007.

18. Di Mauro, E., Synder, L., Marino, P., Lamberti, A., Coppo, A., & Tocchini-Valentini, G. P. (1969). Rifampicin sensitivity of the components of DNA-dependent RNA polymerase. *Nature*, *222*(5193), 533–7.

19. Wehrli, W., Knüsel, F., Schmid, K., & Staehelin, M. (1968). Interaction of rifamycin with bacterial RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *61*(2), 667–73.

20. Suzuki, M., Zhang, J., Liu, M., Woychik, N. a, & Inouye, M. (2005). Single protein production in living cells facilitated by an mRNA interferase. *Molecular Cell*, *18*(2), 253–61. doi:10.1016/j.molcel.2005.03.011.

21. Almeida, F. C., Amorim, G. C., Moreau, V. H., Sousa, V. O., Creazola, a T., Américo, T. a, … Valente, a P. (2001). Selectively labeling the heterologous protein in Escherichia coli for NMR studies: a strategy to speed up NMR spectroscopy. *Journal of Magnetic Resonance*, *148*(1), 142–6.

22. Mao, L., Tang, Y., Vaiphei, S. T., Shimazu, T., Kim, S.-G., Mani, R., … Inouye, M. (2009). Production of membrane proteins for NMR studies using the condensed single protein (cSPP) production system. *Journal of Structural and Functional Genomics*, *10*(4), 281–9. doi:10.1007/s10969-009-9072-0.

23. Mao, L., Inoue, K., Tao, Y., Montelione, G. T., McDermott, A. E., & Inouye, M. (2011). Suppression of phospholipid biosynthesis by cerulenin in the condensed Single-Protein-Production (cSPP) system. *Journal of Biomolecular NMR*, *49*(2), 131–7.

24. Suzuki, M., Mao, L., & Inouye, M. (2007). Single protein production (SPP) system in Escherichia coli. *Nature Protocols*, *2*(7), 1802–10. doi:10.1038/nprot.2007.252.

25. Asenjo, J. A., & Andrews, B. A. (2009). Protein purification using chromatography: selection of type, modelling and optimization of operating conditions. *Journal of Molecular Recognition : JMR*, *22*(2), 65–76. doi:10.1002/jmr.898.

26. Crowe, J., Döbeli, H., Gentz, R., Hochuli, E., Stüber, D., & Henco, K. (1994). 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods in Molecular Biology (Clifton, N.J.)*, *31*, 371–87. doi:10.1385/0-89603-258-2:371.

27. Saraswat, M., Musante, L., Ravidá, A., Shortt, B., Byrne, B., & Holthofer, H. (2013). Preparative purification of recombinant proteins: current status and future trends. *BioMed Research International*, *2013*, 312709. doi:10.1155/2013/312709.

28. Dekker, N., Merck, K., Tommassen, J., & Verheij, H. M. (1995). In vitro folding of Escherichia coli outer-membrane phospholipase A. *European Journal of Biochemistry / FEBS*, *232*(1), 214–9.

29. Heller, K. B. (1978). Apparent molecular weights of a heat-modifiable protein from the outer membrane of Escherichia coli in gels with different acrylamide concentrations. *Journal of Bacteriology*, *134*(3), 1181–3.

30. Nakamura, K., & Mizushima, S. (1976). Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12. *Journal of Biochemistry*, *80*(6), 1411–22.

31. Robert, V., Volokhina, E. B., Senf, F., Bos, M. P., Van Gelder, P., & Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biology*, *4*(11), e377. doi:10.1371/journal.pbio.0040377.

32. Weingarth, M., Cruijsen, E. A. W. Van Der, Ostmeyer, J., Lievestro, S., Roux, B., & Baldus, M. (2014). Quantitative analysis of the water occupancy around the selectivity filter of a K + channel in different gating modes. *J Am Chem Soc*, *136, 2000-7.* doi: 10.1021/ja411450y.

33. Linser, R., Dasari, M., Hiller, M., Higman, V., Fink, U., Lopez del Amo, J.-M., … Reif, B. (2011). Proton-detected solid-state NMR spectroscopy of fibrillar and membrane proteins. *Angewandte Chemie (International Ed. in English)*, *50*(19), 4508–12. doi:10.1002/anie.201008244.

34. Cukkemane, A., & Baldus, M. (2013). Characterization of a cyclic nucleotide-activated K(+) channel and its lipid environment by using solid-state NMR spectroscopy. *Chembiochem : A European Journal of Chemical Biology*, *14*(14), 1789–98. doi:10.1002/cbic.201300182.

35. Renault, M., Pawsey, S., Bos, M. P., Koers, E. J., Nand, D., Tommassen-van Boxtel, R., … Baldus, M. (2012). Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angewandte Chemie (International Ed. in English)*, *51*(12), 2998–3001.

36. Andronesi, O. C., Becker, S., Seidel, K., Heise, H., Young, H. S., & Baldus, M. (2005). Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *J Am Chem Soc*, *127*, 12965–12974. doi:10.1021/ja0530164.

37. Etzkorn, M., Kneuper, H., Dünnwald, P., Vijayan, V., Krämer, J., Griesinger, C., … Baldus, M. (2008). Plasticity of the PAS domain and a potential role for signal transduction in the histidine kinase DcuS. *Nature Structural & Molecular Biology*, *15*(10), 1031–9. doi:10.1038/nsmb.1493.

38. Etzkorn, M., Seidel, K., Li, L., Martell, S., Geyer, M., Engelhard, M., & Baldus, M. (2010). Complex formation and light activation in membrane-embedded sensory rhodopsin II as seen by solid-state NMR spectroscopy. *Structure (London, England : 1993)*, *18*(3), 293–300. doi:10.1016/j.str.2010.01.011.

39. Sinnige, T. *et al.* Insight into the conformational stability of membrane-embedded BamA using a combined solution and solid-state NMR approach. *J. Biomol. NMR* **61,** 321–32 (2015).

40. Weingarth, M., & Baldus, M. (2013). Solid-state NMR-based approaches for supramolecular structure elucidation. *Accounts of Chemical Research*, *46*, 2037–46. doi:10.1021/ar300316e.

41. Goncalves, J., Eilers, M., South, K., Opefi, C. A., Laissue, P., Reeves, P. J., & Smith, S. O. (2013). Magic angle spinning nuclear magnetic resonance spectroscopy of G protein-coupled receptors. *Methods in Enzymology*, *522*, 365–89. doi:10.1016/B978-0-12-407865-9.00017-0.

42. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12,** 5–9 (2015).

CHAPTER SIX

Discussion and Future Perspectives

Until recently, structural biology, which aims at investigating biomolecules at atomic resolution, has mostly been successful in achieving its goal *in-vitro*, thereby neglecting the influence of the surrounding environment on the biomolecule under study. However, the final target of structural biology is to be able to scrutinize biomolecules at atomic-resolution in their native cellular environment. In recent years, solution-state Nuclear Magnetic resonance (NMR) has been able to investigate small soluble proteins in their cellular environment (a.k.a. in-cell NMR) (Ref.[1–5]). One limitation of solution-state NMR is that it is restricted to molecules that tumble quickly, which renders it unsuitable for membrane proteins. For a long time, solid-state NMR (ssNMR) has shown to be an ideal method to study membrane proteins at atomic resolution in synthetic lipid bilayers. Recently, ssNMR has been extended to investigate membrane proteins in their native environment including the bacterial cell envelope and whole bacterial cells (termed cellular ssNMR[6–8]). Traditionally, membrane proteins were investigated in synthetic lipid bilayers using ssNMR. Such experiments include expressing the protein in bacteria, protein purification, solubilization and, finally, reconstitution in the chosen lipid bilayers[9]. In addition to being a very cumbersome procedure, extracting the protein from its natural cellular environment may lead to the loss of important information about the structure and dynamics of the protein. Moreover, extracting the protein from its natural membrane environment during purification may cause artifacts and yield information that is irrelevant to the protein in its physiological condition (i.e., while being embedded in the cell environment). However, new challenges arise when trying to study membrane proteins in cell envelope or whole cells: Firstly, one must be able to differentiate the ssNMR signal of the investigated protein from the complex cellular environment in which the protein exists. In addition, the ssNMR sample volume inside an MAS rotor is occupied by cell envelope/or whole cells samples containing lipids, sugars and other endogenous molecules. Together these molecules reduce the sensitivity to detect the (membrane) protein of interest in cellular samples compared to the case of purified proteins. Finally, due to the heterogeneity of the cellular environment, the ssNMR spectra of these samples may be of lower spectral resolution compared to the in-vitro case.

In the beginning, cellular ssNMR was used to investigate small membrane proteins[10–13]. In this thesis, we extended this idea to large membrane complexes that span the whole cell envelope and further to include even eukaryotic membrane proteins. To achieve this, we combined ssNMR with sensitivity-enhancement methods, namely, Dynamic Nuclear

Polarization (DNP) to overcome the low sensitivity problem associated with cellular samples. Furthermore, we optimized a variety of aspects in the sample production protocol to decrease the background signals and the spectral crowding.

Firstly, we investigated the bacterial Type IV secretion system core complex (T4SScc), which is a one megadalton complex composed of three proteins (namely, B7, B9 and B10) and spans the whole bacterial cell envelope, in its natural cellular environment (i.e., the bacterial cell envelope). T4SScc is part of a bigger molecular machinery, the type IV secretion system (T4SS), which is involved in the transfer of many proteins, DNA, effector molecules between the bacteria and into the host cells leading to many diseases like peptic ulcer. A crystal structure is available for the upper part of T4SScc (the so-called outer layer complex), while the lower part (the inner layer complex) is still elusive for high-resolution studies. To investigate T4SScc in cell envelope using ssNMR, we had to decrease the background signals. In solution-state in-cell NMR, background signals are usually suppressed by producing a purified labeled version of the protein of interest and then injecting it into an unlabeled cellular environment where it can be studied provided the background signal does not dominate the NMR spectrum. However, this strategy is unsuitable for membrane proteins. Alternatively, blocking the bacterial RNA polymerase by adding rifampicin after induction, which does not interfere with T7 RNA polymerase diminishes the background signals drastically. Recently, it has been shown that the rifampicin method can be also applied successfully for membrane proteins samples[14]. Alternatively, we here decreased the background signals by using a special *E.coli* strain that lacks two naturally outer-membrane abundant proteins, namely, OmpA and OmpF (see ref.[6]). Moreover, to decrease spectral crowding, we produced specifically labeled samples, where certain amino acids are ^{15}N labeled and others are ^{13}C labeled, amenable for DNP studies. To prepare the specifically labeled proteins, amino acids were chosen so as to yield small number of sequential correlations in different secondary structure elements and distributed over the different parts of the complex. These labeled amino acids act as atomic probes in the different parts of the machinery, which yield local information about the structure and dynamics of T4SScc. Two specifically labeled T4SScc in *E.coli* cell envelopes were prepared, the first sample was ^{13}C, ^{15}N GlySerLeuVal (referred to as GSLV) labeled and the other sample was ^{15}N Val and ^{13}C, ^{15}N Thr (referred to as TV) labeled. The GSLV-T4SScc yielded 51 sequential correlations distributed over the whole complex. On the other hand, the TV labeled sample yielded only

seven sequential correlations distributed also over the whole machinery. Obtaining an enhancement factor of ~ 60 on the 400 MHz DNP machine for the specifically labeled samples allowed us to perform 3D experiments in short time. The spectra of GSLV-T4SScc were still crowded, nonetheless, we could confirm parts of the crystal fold of the outer layer complex and we could obtain new insights into the structure and dynamics of the inner layer complex in its native environment. However, due to the lacking of any specific atomic probes in the inner layer part of VirB10 (except one in the transmembrane domain in TV-T4SScc), we could not obtain specific information about this part of the protein. Therefore, a sample, which introduces specific probes in this part of the complex, might yield interesting information about this hitherto elusive part of the machinery. In addition, constructs that code for larger T4SS subcomplexes in addition to the T4SScc are also available and these would be certainly worth to investigate by labeling certain proteins only and then combine them to form the bigger machinery (for example by applying the so-called LEGO method of labeling[15]). The addition of a known substrate to this machinery while it is in the cell envelope and probing the interaction using ssNMR is one of the options for future studies regarding this project.

Concerning the cellular ssNMR on eukaryotic systems, we have here used ssNMR to investigate the epidermal growth factor receptor (EGFR) in its native environment. A major hurdle in applying cellular ssNMR to eukaryotic proteins is to produce [^{13}C, ^{15}N] labeled samples. In-cell NMR studies on soluble proteins often produce the protein of interest in a labeled medium, purify it and then introduce it in the mammalian cells. For example, Shirakawa and co-workers used a cell-penetrating peptide to introduce the protein of interest into the cell[16], while Shimada an co-workers used pore forming toxin (the streptolysin O)[17]. However, such methods are not suited for membrane proteins, as they have to be labeled in the mammalian cell directly. In this thesis, we have introduced a novel way to study the membrane protein EGFR in its cellular environment. To achieve this, human epidermoid carcinoma cells (A431), which endogenously express high levels of EGFR, were [^{13}C, ^{15}N] labeled by growing them on a medium enriched with [^{13}C, ^{15}N] algae mixture. Subsequently, membrane vesicles were isolated from the labeled cells, loaded into 3.2 mm MAS rotors and studied with ssNMR yielding new information about the dynamics of EGFR in its native environment in the presence and absence of its ligand EGF. To zoom in into certain regions of the receptor, we produced specifically labeled samples where Met and Phe are ^{13}C labeled,

and Thr and Leu are ^{15}N labeled (referred to as MFTL sample). This labeling scheme introduced probes into different domains of the protein, namely, domains I-III of the extracellular part, the kinase domain and the C-terminal tail. These local probes yielded information about the local dynamics of the protein before and after the binding to its ligand EGF.

The membrane vesicles isolated in these studies are not pure plasma membrane vesicles. For future studies, biochemical methods (like sucrose gradient) can be used to separate the plasma membrane vesicles from other organelles vesicles. Moreover, another specifically labeled samples may be produced to zoom-in into different regions of the protein, like domain IV, in which we did not have any atomic probes in our current studies.

Interestingly, many nanobodies, the so-called VHH's (heavy chain only antibodies) that specifically bind to EGFR are available. Some of these nanobodies bind to EGFR but do not activate it, while others (the so-called bi-parapotic nanobodies, which are composed of two nanobodies) induce clustering and internalization of the receptor without activating it. Solution-state NMR assignments for one of these nanobodies, namely 7D12, are already available in our group. Investigating the interaction of these nanobodies with EGFR and comparing that with EGF binding with ssNMR might yield new interesting information about this receptor and its activation profile.

In the recent years, remarkable developments in proton detection in ssNMR have taken place[18–22]. The ability to detect protons on cellular samples would potentially help to improve their sensitivity. Detecting protons on perdeuterated and fractionally deuterated proteins can yield spectra of very good quality, which would help to easily assign these proteins. For prokaryotic systems, we have shown in chapter 4 the possibility of producing perdeutarated and fractionally deuterated T4SScc in *E.coli* with high yield. This opens the doors to perform proton detection studies on T4SScc, preferably, on specifically labeled samples in a perdeuterated background (also called "proton clouds", see Ref.[23]). On the other hand, the application of proton detection on the eukaryotic samples remain very challenging at this stage as protocols for the deuteration of eukaryotic cells are to be optimized first which is far from being trivial. In a first step, ligands such the aforementioned nanobodies that can be

produced in bacteria may be suited for such studies by producing them in a perdeuterated form.

In our cellular ssNMR/DNP studies discussed in this thesis (for both prokaryotic and eukaryotic systems), we obtained comparatively low enhancement factors on the 800 MHz DNP machine compared to the 400 MHz DNP (on the 800 MHz DNP, the enhancement was ~ ¼ that obtained on the 400 MHz DNP field). Consequently, this restricted our studies on the 800 MHz DNP machine to 2D experiments. However, recent studies on our 800 MHz system (Mathies et al., Angewandte Chem. Int.ed in press) are suggesting new radicals which yield higher enhancement factors (through mechanisms other than the cross effect) at higher magnetic fields. Achieving higher enhancement factors at higher magnetic fields, would allow the performing of 3D experiments at 800 MHz DNP conditions in a shorter time. Combining this with the high-resolution associated with this magnetic field will help to make it possible to investigate more complex systems with cellular ssNMR/DNP.

The ability to study the life molecular machines at atomic resolution in their native environment has long been the goal of structural biology. Currently, some structural methods can investigate biomolecules in the cell at high resolution (like electron microscopy). However, cellular ssNMR has the advantage of being able to track the atomic structure and dynamics of the proteins in the cell. This can yield information that cannot be obtained by other techniques; neither can this information be obtained by *in-vitro* methods. The studies presented in this thesis are the first steps to make cellular ssNMR applicable to more biological systems that are known to be difficult to produce in prokaryotic cells such as microtubules. Moreover, the installation of the 950 MHz and the 1.2 GHz (in the near future) and combining that with other sensitivity enhancement methods like for example, non-uniformal sampling (NUS), with the associated improvement in resolution and sensitivity, would render cellular ssNMR/DNP capable of answering more challenging biological questions that cannot be tackled by any other technique.

References:

1. Selenko, P. & Wagner, G. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* **158,** 244–53 (2007).

2. Ito, Y. & Selenko, P. Cellular structural biology. *Curr. Opin. Struct. Biol.* **20,** 640–8 (2010).

3. Freedberg, D. I. & Selenko, P. Live cell NMR. *Annu. Rev. Biophys.* **43,** 171–92 (2014).

4. Hänsel, R., Luh, L. M., Corbeski, I., Trantirek, L. & Dötsch, V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angew. Chem. Int. Ed. Engl.* **53,** 10300–14 (2014).

5. Sakakibara, D. *et al.* Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458,** 102–5 (2009).

6. Renault, M. *et al.* Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 4863–8 (2012).

7. Renault, M. *et al.* Solid-state NMR spectroscopy on cellular preparations enhanced by dynamic nuclear polarization. *Angew. Chem. Int. Ed. Engl.* **51,** 2998–3001 (2012).

8. Kaplan, M. *et al.* Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR. *Nat. Methods* **12,** 5–9 (2015).

9. Baker, L. a *et al. Magic-angle-spinning solid-state NMR of membrane proteins. Methods Enzymol.* **557,** 307–28 (Elsevier Inc., 2015).

10. Etzkorn, M. *et al.* Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **46,** 459–62 (2007).

11. Fu, R. *et al.* In situ structural characterization of a recombinant protein in native Escherichia coli membranes with solid-state magic-angle-spinning NMR. *J. Am. Chem. Soc.* **133,** 12370–3 (2011).

12. Jacso, T. *et al.* Characterization of Membrane Proteins in Isolated Native Cellular Membranes by Dynamic Nuclear Polarization Solid-State NMR Spectroscopy without Purification and Reconstitution. *Angew. Chemie* **124,** 447–450 (2012).

13. Miao, Y. *et al.* M2 proton channel structural validation from full-length protein samples in synthetic bilayers and E. coli membranes. *Angew. Chem. Int. Ed. Engl.* **51,** 8383–6 (2012).

14. Baker, L. a, Daniëls, M., van der Cruijsen, E. a W., Folkers, G. E. & Baldus, M. Efficient cellular solid-state NMR of membrane proteins by targeted protein labeling. *J. Biomol. NMR* **62,** 199–208 (2015).

15. Mund, M., Overbeck, J. H., Ullmann, J. & Sprangers, R. LEGO-NMR spectroscopy: a method to visualize individual subunits in large heteromeric complexes. *Angew. Chem. Int. Ed. Engl.* **52,** 11401–5 (2013).

16. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458,** 106–109 (2009).

17. Ogino, S. *et al.* Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. *J. Am. Chem. Soc.* **131,** 10834–5 (2009).

18. Reif, B., Jaroniec, C. P., Rienstra, C. M., Hohwy, M. & Griffin, R. G. 1H-1H MAS correlation spectroscopy and distance measurements in a deuterated peptide. *J. Magn. Reson.* **151,** 320–7 (2001).

19. Chevelkov, V., Rehbein, K., Diehl, A. & Reif, B. Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew. Chem. Int. Ed. Engl.* **45,** 3878–81 (2006).

20. Ward, M. E. *et al.* Proton-detected solid-state NMR reveals intramembrane polar networks in a seven-helical transmembrane protein proteorhodopsin. *J. Am. Chem. Soc.* **133,** 17434–43 (2011).

21. Knight, M. J. *et al.* Structure and backbone dynamics of a microcrystalline metalloprotein by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 11095–100 (2012).

22. Weingarth, M., Cruijsen, E. A. W. Van Der, Ostmeyer, J., Lievestro, S. & Baldus, M. Quantitative Analysis of the Water Occupancy around the Selectivity Filter of a K + Channel in Di ff erent Gating Modes. (2014).

23. Sinnige, T., Daniëls, M., Baldus, M. & Weingarth, M. Proton clouds to measure long-range contacts between nonexchangeable side chain protons in solid-state NMR. *J. Am. Chem. Soc.* **136,** 4452–5 (2014).

Summary

Determining the structure of biomolecules, which is the main goal of structural biology, is pivotal to understand their function. However, this has mainly been achieved *in-vitro* by methods like Nuclear Magnetic Resonance (NMR) and X-ray crystallography. Due to the important role of the natural cellular environment on the proteins, extracting them from the native environment may yield information or structures that are irrelevant to the physiological conditions. For this reason, new methods are being developed to enable investigating biomolecules at high resolution in their native environment. Solution-state NMR has been successful in investigating small soluble proteins, at atomic resolution inside the cell (the so-called in-cell NMR). On the other hand, solid-state NMR (ssNMR) has long been the method of choice to scrutinize the dynamics and structure of membrane proteins at atomic resolution in synthetic lipid bilayers. More recently, ssNMR has made remarkable developments in studying small membrane proteins in their natural cellular environment (a.k.a. cellular ssNMR). In this thesis, we extend cellular ssNMR to more challenging prokaryotic and eukaryotic protein complexes.

In **chapter 2**, we explore the type IV secretion system core complex (T4SScc), a one-megadalton bacterial machinery, in the cell envelope of *E.coli* by using cellular ssNMR supported by Dynamic Nuclear Polarization (DNP) to enhance the sensitivity. Hitherto, T4SScc has been investigated using *in-vitro* based methods like X-ray crystallography and electron microscopy. However, some recent studies suggested that the complex adapts a conformation different to what was observed using the *in-vitro* based techniques. Moreover, high-resolution structural data are available for only half of T4SScc. For these reasons, we investigated T4SScc in its natural environment, this being the bacterial cell envelope. Our results confirmed parts of the previously existing crystal structure. Moreover, we could provide the first experimental evidence that the hitherto structurally elusive part of the complex is partly folded and inserts into the inner membrane of bacteria in the cellular settings.

In **chapter 3**, we extended the idea of cellular ssNMR to investigate, for the first time, eukaryotic membrane protein complexes, namely, the epidermal growth factor receptor (EGFR), in membrane vesicles isolated from A431 cancer cells. These cells are known to express high levels of EGFR on their surface. First, we optimized a protocol to label the cells

with NMR active nuclei (these being ^{13}C, ^{15}N) and then isolated the membrane vesicles (which have EGFR on their surface) and studied the structure and dynamics of EGFR using DNP-supported ssNMR. We investigated the effect of different temperatures on EGFR in the isolated vesicles in the presence and absence of its ligand, the epidermal growth factor (EGF). At low temperature (253 K), the addition of the ligand did not have influence on the spectra. However, at higher temperature (285 K), the addition of EGF to the isolated vesicles enhanced the sensitivity of the signals in the dipolar-based experiments, which probe the rigid parts of the sample. Further analysis of the spectra recorded at 285 K, in the presence and absence of EGF, showed that in the absence of the ligand, the only rigid part of the receptor is the kinase domain (with the transmembrane part) while the rest of the protein is dynamic. These results are in contrast to earlier *in-silico* based studies, which suggested that the extracellular part of the protein is also rigid due to its interaction with the membrane. Thereafter, we prepared A431 membrane vesicles which are specifically labeled, in other words, where only certain nuclei are visible in NMR while the remaining nuclei are the so-called NMR silent. Interestingly, by measuring these vesicles at DNP temperatures (100 K), we obtained broader peaks in our spectra without the ligand compared to the spectra recorded when EGF was present. The presence of broader line width in the absence of EGF indicates that the specific labeled sites in our sample have local dynamics that become limited when EGF is bound to the receptor. These local dynamics may hinder any interaction with the membrane, which was suggested by molecular dynamics simulations performed on EGFR neglecting the native environment. Our results presented here are the first high-resolution studies of the full length EGFR in its native environment.

In **chapter 4**, we combined the proton-detection methods that are being recently developed in our group with cellular ssNMR. Detecting protons has the advantage of increasing the sensitivity of NMR experiments (due to the high gyromagnetic ratio of protons). Here, we produced fractionally deuterated and perdeuterated *E.coli* cell envelops which express T4SScc amenable to proton detection studies. SDS-PAGE analysis of the isolated perdeuterated and fractionally deuterated *E.coli* cell envelops showed the possibility of expressing this complex under these conditions. Moreover, our preliminary data presented in this chapter illustrates the feasibility of detecting protons on cellular samples. By comparing the NH spectrum of fractionally deuterated *E.coli* cell envelop expressing T4SScc with NH spectrum of fractionally deuterated ubiquitin, we could conclude that the signal distribution

in our T4SScc spectra reflects different secondary structure elements which speaks in favor of a well-folded complex in the bacterial cell envelope.

In **chapter 5**, we present the protocols used to prepare membrane protein samples eligible for ssNMR experiments, both for conventional (purified) and cellular studies. A detailed description of how to express, purify and reconstitute membrane proteins (the potassium channel KcsA and the β-barrel folding machinery BAM) amenable for ssNMR investigations is given. Moreover, step-by step protocols are described for how to prepare cellular samples (prokaryotic and eukaryotic) eligible for ssNMR studies. Protocols of how to produce specifically labeled samples (purified and cellular) are also presented in this chapter.

In **chapter 6**, a discussion of the results of the previous chapters and possible future perspectives of cellular ssNMR are described in the light of the newly developed DNP radicals and the installation of the 950 MHz spectrometer in Utrecht.

In recent years, scientists are realizing more and more the importance of investigating biomolecules in their native cellular environment. NMR has shown to be a powerful method for studying proteins' structure and dynamics at atomic resolution. The work described in this thesis is a step to extend the realm of ssNMR to investigate large prokaryotic and eukaryotic membrane protein complexes in their native environment. This will help to achieve "cellular structural biology", that is studying biomolecules at atomic resolution inside the cell.

Samenvatting

Het bepalen van de structuur van biomoleculen, het voornaamste doel van de structuurbiologie, is van centraal belang om hun functie te begrijpen. Dit is echter hoofdzakelijk bereikt door in-vitromethodes zoals kernspinresonantie (NMR) en röntgendiffractie. Vanwege de belangrijke rol die de natuurlijke cellulaire omgeving speelt op eiwitten, kan de verwijdering van eiwitten uit hun natuurlijke omgeving informatie of structuren opleveren die irrelevant zijn voor de fysiologische omstandigheden. Dit heeft geleid tot enorme inspanningen in de laatste jaren om biomoleculen te onderzoeken op hoge resolutie in hun natuurlijke omgeving. Vloeistof NMR is succesvol in het bestuderen van kleine oplosbare eiwitten op hoge resolutie in de cell (het zogeheten in-cel NMR). Vaste-stof NMR (ssNMR) geniet daarentegen allang de voorkeur om de dynamica en structuur van membraaneiwitten onder de loep te nemen op atomaire resolutie in synthetische lipidendubbellagen. Recentelijk heeft ssNMR indrukwekkende ontwikkelingen ondergaan in het bestuderen van kleine membraaneiwitten in hun natuurlijke cellulaire omgeving (ook bekend als cellulaire ssNMR). In dit proefschrift passen we cellulaire ssNMR toe op uitdagendere prokaryotische en eukaryotische eiwitcomplexen.

In **hoofdstuk 2** onderzoeken we het type IV secretiesysteemkerncomplex (T4SScc), een 1 MDa bacteriële machinerij, in de celenvelop van E. coli door gebruik te maken van cellulaire ssNMR in combinatie met dynamische nucleaire polarisatie (DNP) om de gevoeligheid te vergroten. Tot nog toe is T4SScc alleen onderzocht met in-vitrogebaseerde methodes, zoals röntgendiffractie en electronenmicroscopie. Sommige studies hebben echter aangegeven dat het complex een andere conformatie aanneemt dan wat is waargenomen in in-vitrogebaseerde studies. Bovendien is er hoge-resolutie structuurdata beschikbaar voor enkel de helft van de T4SScc. Daarom hebben we T4SScc onderzocht in zijn natuurlijke omgeving, de bacteriële celenvelop. Onze resultaten bevestigde delen van de bestaande kristalstructuur. Daarnaast hebben we de eerste experimentele bewijzen geleverd dat het tot nog toe ongrijpbare deel van het complex een secundaire vouwing aanneemt en wordt ingevoegd in het binnenmembraan van bacteriën in de cellulaire setting.

In **hoofdstuk 3** breidden we het idee van cellulaire ssNMR verder uit om voor het eerst eukaryotische membraaneiwitcomplexen te onderzoeken, in dit geval de epidermaal-groeifactorreceptor (EGFR), in membraanvesikels geïsoleerd van A431 kankercellen. Deze

cellen staan erom bekend dat ze hoge levels van EGFR tot expressie brengen. Eerst optimaliseerden we het protocol om cellen te labelen met NMR-actieve nuclei (om ze zodoende zichtbaar te maken voor NMR) en om daarna de membraanvesikels te isoleren (welke EGFR op de oppervlakte hebben) en bestudeerden we de structuur en dynamica van EGFR d.m.v. met DNP-ondersteunde ssNMR. We onderzochten het effect van verschillende temperaturen op EGFR in de geïsoleerde vesikels in de aan- en afwezigheid van zijn ligand, de epidermaal-groeifactor (EFG). Bij lage temperatuur (258K) had de additie van het ligand geen invloed op het spectrum. Echter, bij hogere temperaturen (285K) bleek additie van EGF aan de geïsoleerde vesikels de gevoeligheid te verhogen van het signaal in de dipolair-gebaseerde experimenten, welke de rigide delen van het sample onderzoekt. Verdere analyse van het spectrum opgenomen bij 285K, zowel in de aan- en afwezigheid van EGF, toonde aan dat bij afwezigheid van het ligand het enige rigide gedeelte van de receptor het kinasedomein was (te samen met het transmembraan), terwijl het overige deel van het eiwit dynamisch is. Deze resultaten contrasteren met eerdere in-vitro gebaseerde studies, die aanduidden dat het extracellulaire gedeelte van het eiwit ook rigide is vanwege interacties met het membraan. Hierna bereidden we A431 membraanvesikels welke specifiek gelabeld waren, in andere woorden, waar nu andere nuclei zichtbaar waren in NMR, terwijl de overige nuclei de zogeheten onzichtbare nuclei zijn. Bij het meten van deze vesicles bij DNP temperaturen (100K) verkregen wij, interessant genoeg, bredere pieken in onze spectra in de absentie van het ligand vergeleken met de spectra opgenomen in het bijzijn van EGF. De aanwezigheid van bredere lijnbreedte in de absentie van EGF geeft een indicatie dat de specifiek gelabelde locaties in ons sample lokale dynamica ondervindt dat beperkt wordt wanneer EGF gebonden is aan de receptor. De presentie van deze lokale dynamica kan iedere interactie met het membraan hinderen, hetgeen ook was gesuggereerd door moleculaire-dynamicasimulaties uitgevoerd op het EGFR waarbij de native omgeving achterwege werd gelaten. Onze resultaten hier zijn de eerste hoge-resolutie studies van het volledige EGFR in zijn natuurlijke omgeving.

In **hoofdstuk 4** combineerden we proton-detectiemethodes die in onze groep recentelijk worden ontwikkeld met cellulaire ssNMR. Het detecteren van protonen heeft het voordeel dat het de gevoeligheid van NMR experimenten verhoogd (dankzij de hoge gyromagnetische ratio van protonen). Hier produceerden we gedeeltelijk gedeutereerde en geperdeutereerde E. coli celenveloppen, die T4SScc tot expressie brengen waarop proton-detectiestudies

toegepast kunnen worden. SDS-PAGE analyse van het geisoleerde geperdeutereerde en gedeeltelijk gedeutereerde E. coli celenveloppen toonde de mogelijkheid aan van het tot expressie brengen van T5SScc onder deze condities. Bovendien wijst onze vroegtijdige data in dit hoofdstuk aan de haalbaarheid van protondetectie op cellulaire samples. Door het vergelijken van het NH-spectrum van gedeeltelijke gedeutereerde E. coli celenveloppen welke T4SScc tot expressie brengen, met het NH-spectrum van gedeeltelijk gedeutereerd ubiquitine, kwamen we tot de conclusie dat de signaaldistributie in onze T4SScc spectra verschillende secundaire structuurelementen weerspiegelden, hetgeen duidt op een wel-gevouwen T4SScc in de bacteriële celenvelop.

In **hoofdstuk 5** tonen wij de protocollen die zijn gebruikt voor het bereiden van membraaneiwitsamples die gebruikt kunnen worden voor ssNMR experiment, zowel voor conventionele (gezuiverde) en cellulaire studies. We geven een gedetailleerde beschrijving van het tot expressie brengen, het zuiveren en reconstitueren van membraaneiwitten (het kaliumkanaal KcsA en de β-vatvouwingsmachinerij BAM) toegepast voor ssNMR onderzoeken. Verder beschrijven we stap-voor-stapprotocollen voor het prepareren van cellulaire samples (zowel prokaryotisch als eukaryotisch) voor ssNMR studies. Ook worden in dit hoofdstuk protocollen voor het produceren van specifiek gelabelde samples (gezuiverd en cellulair) gegeven.

In **hoofdstuk 6** beschouwen we de resultaten van de vorige hoofdstukken en bespreken we mogelijke toekomstige perspectieven van cellulair ssNMR in het licht van de huidige ontwikkelde DNP-radicalen en de installatie van de 950 MHz spectrometer in Utrecht.

De laatste jaren is de realisatie ontstaan dat het uiterst belangrijk is om biomoleculen te onderzoeken in hun native cellulaire omgeving. NMR is hierbij een krachtige methode om de structuur en dynamica van eiwitten te onderzoeken op atomaire resolutie. Het werk beschreven in dit proefschrift breidt de mogelijkheden van ssNMR verder uit om ook grote prokaryotische en eukaryotische membraaneiwitcomplexen te bestuderen in hun natuurlijke omgeving. Dit is een verdere stap voor het bereiken van “cellulaire structuurbiologie”, het bestuderen van biomoleculen op atomaire resolutie in hun native omgeving.

Acknowledgements

It is very difficult to predict how life will unfold in the future; here I am after four years writing the acknowledgements of my PhD thesis. It was a nice journey which would not have been possible without the help and support I got from many people.

First of all, I would like to express my profound gratitude and to thank my PhD supervisor, Prof. Marc Baldus for giving me the chance to join his group and to work with very nice and challenging projects. Also, for all the inspiration and supervision I received from him during the four years of my PhD.

I would like also to thank Prof. Rolf Boelens, Prof. Alexandre Bonvin and Prof. Rob Kaptein for their continuous help and fruitful discussions.

Many thanks go also to the solid state NMR group. Elwin, Eline and Tessa, thank you for your help when I started my PhD and for being very nice colleagues. I wish you all the best with the new challenges in your life. Sabine and Lindsay, thank you for the nice work environment. Abhi, thank you for many things (supervision, support, and delicious recipes). Deni, thank you for the nice time we spent at the DNP machine. Cecilia, I wish you all the best with your BAM project. Sid and Jon, it was a pleasure to have you first as students and now as colleagues. I wish all the best with your PhD.

Special thanks for you Markus, thank you for your help, supervision and support and for helping me to improve my NMR knowledge. I wish you all the best in your future carrier and in establishing your group.

Mark Daniels, thank you for the help in the wetlab and for all the nice discussions we had together.

Gert, I appreciate your help and support during my master and PhD time. I learned a lot from you. Thank you.

I would like to express my thanks to Klaartje for all the critical comments that helped to improve my work and for the supervision on the EGFR project. I wish all the best for you, Adrien and your nice children.

I would like also to thank Marie Renault for her help at the beginning of my PhD.

Special thanks for my roommate, Hans Wienk. Thank you for the nice four years in which we shared the same room and for all the nice discussions we had about almost everything in this life. I wish you all the best in your future carrier.

I would also like to show my gratitude for my colleagues from the solution NMR group. Maryam, thank you for being a very nice colleague and a friend. I wish you find soon a very nice job. Rama, I am sure you can do it! Ivan, thank you for all the nice discussions we always have at the coffee table. I wish you success in your PhD. Tobias, thanks for your supervision and help during my master project. I wish you all the best in Graz. Hugo, Lidjia and Henry, thank you for the nice environment at work.

The computational group. Gydo, thank you for being a nice colleague and a friend. Thanks for all the nice discussions we always have (especially about the philosophy of science) and thank you for being my paranymph (with Sid) and for your help in the Dutch summary. I hope to see you again in the US. Joao, I wish you all the best in Stanford and hope to meet you there also. Ezgi and Panos, it was always a pleasure to talk to you, I wish you all the best in Heidelberg. Li, Anna and Zeynep, thank you for the very nice atmosphere at work. Cunliang, I wish you all the best in your PhD (thank you for introducing me to the noodle world). Adrien, I am looking forward to see you again in February and I hope we can play football again. Many thanks go to Marc van Dijk, Gijs, Mikael and Christopher also.

I would like to thank Dr. Paul van Bergen en Henegouwen for all his help, support and supervision during my master and my PhD time. Also, many thanks to the group of Dr. Paul (Rachid, Sofia, Purvi, Bram, and Dusan), for their continuous help and support.

I would like to thank Dr. Lukas Kapitein and Dr. Eugene Katrukha for their help in the EGFR project.

Special thanks to Raimond Heukers, I was very lucky to have you as my supervisor during my master. I learned a lot from you and I wish you all the best in your future carrier.

I would like to show my gratitude to Johan for his help in the DNP measurements and for the nice discussions we always had together.

Very special thanks to Barbara for the continuous support at work and for the wonderful music concerts you invited me to. Also my gratitude to your kind family: Jils, Mees and Merel.
Outside my work, there were many people who gave me help and support during the tough moments of my PhD. Rayan, Ahmed and Tarik, thank you for being great friends. Firas, Sema, Abdullah and Maryam, thank you very much for being always ready to help me and for your regular invitations. Jan-kees Karels, Esther, Emma, Constantijn, Karel-jan and lieve Mary, thank you for the continuous help and support and for the great "musical" time we had together. Hasan Khalid thanks for the great time we spent together in Dallas. Peter, thanks for the nice visits.

I would also like to thank UAF (stichting voor vluchteling-studenten) for their support during my master studies.

I would like to thank Tom Rose for being a wonderful piano teacher.

Finally, my unlimited gratitude to my family. My parents, Mustafa and Nidhal, my sisters, Susan and Dalal (and their husbands Abdullah and Mohammed). Special thanks for Shahad for her help and support.

This thesis is dedicated to my parents.

Mohammed Kaplan.

List of Publications

1- **M. Kaplan et al.**, "Probing a cell-embedded megadalton protein complex by DNP-supported solid-state NMR.," *Nat. Methods*, 2015 Jul; 12(7): 649-52

2- L. Baker, ..., **M. Kaplan**,…and M. Baldus, "Magic-angle-spinning solid-state NMR of membrane proteins". *Methods Enzymol*, 2015; 557:307-28

3- M. Putker, …. **M. Kaplan**,…, B. M. T. Burgering, and T. B. Dansen, "Redox-dependent control of FOXO/DAF-16 by transportin-1.," *Mol. Cell*, 2013 Feb 21; 49(4): 730-42.

4- Deni Mance, …, **Mohammed Kaplan**, …, Marc Baldus, and Markus Weingarth "An efficient labeling approach to harness backbone and side chain protons in ^{1}H-detected solid-state NMR spectroscopy". accepted in *Angewandte chemie*.

5- **M. Kaplan et al.**, "Activation changes global and local dynamics of the full-length Epidermal Growth Factor Receptor in native cell membranes", *submitted*.

Manuscripts in preparation

1- **M. Kaplan et al.**, "Membrane protein complexes investigated by Nuclear Magnetic Resonance (NMR)", invited review for *Quarterly Reviews of Biophysics*.

Curriculum Vitae

Mohammed Kaplan was born on the 23th of October 1983 in Mosul, Iraq. In 2006, he obtained his Medicinae Baccalaureus, Baccalaureus Chirurgiae (M.B.Ch.B., MD degree) from the college of medicine, Mosul University. After working as a medical doctor for one year in Iraq he came to the Netherlands in 2008 and started the master Molecular and Cellular Life Sciences in 2009 at Utrecht University. During his master studies, he worked in the group of Dr. Paul van Bergen en Henegouwen on the internalization of EGFR, followed by a structural study on Wnt signaling pathway using solution-state NMR in the group of Prof. Rolf Boelens. In 2011, he started his PhD in the group of Prof. Marc Baldus at the Bijvoet center for Biomolecular research at Utrecht University, where he developed methods to study membrane proteins (prokaryotic and eukaryotic) in native environment. The work of his PhD is described in this thesis.