NMR Experiments to Characterize Cellular and Molecular Mechanisms: From Metabolomics to Protein Biogenesis

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Morgane Callon

aus Frankreich

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Sebastian Hiller

Prof. Dr. Henning Stahlberg

Basel, 15.09.2015

Prof. Dr. Jörg Schibler

The Dean of Faculty

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Schweiz
(CC BY-NC-ND 3.0 CH)

Sie dürfen: **Teilen** — den Inhalt kopieren, verbreiten und zugänglich machen

Unter den folgenden Bedingungen:

Namensnennung — Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Keine kommerzielle Nutzung — Sie dürfen diesen Inhalt nicht für kommerzielle Zwecke nutzen.

Keine Bearbeitung erlaubt — Sie dürfen diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.

Wobei gilt:

- **Verzichtserklärung —** Jede der vorgenannten Bedingungen kann **aufgehoben** werden, sofern Sie die ausdrückliche Einwilligung des Rechteinhabers dazu erhalten.
- **Public Domain (gemeinfreie oder nicht-schützbare Inhalte) —** Soweit das Werk, der Inhalt oder irgendein Teil davon zur Public Domain der jeweiligen Rechtsordnung gehört, wird dieser Status von der Lizenz in keiner Weise berührt.
- **Sonstige Rechte —** Die Lizenz hat keinerlei Einfluss auf die folgenden Rechte:
    - Die Rechte, die jedermann wegen der Schranken des Urheberrechts oder aufgrund gesetzlicher Erlaubnisse zustehen (in einigen Ländern als grundsätzliche Doktrin des **fair use** bekannt);
    - Die **Persönlichkeitsrechte** des Urhebers;
    - Rechte anderer Personen, entweder am Lizenzgegenstand selber oder bezüglich seiner Verwendung, zum Beispiel für **Werbung** oder Privatsphärenschutz.
- **Hinweis —** Bei jeder Nutzung oder Verbreitung müssen Sie anderen alle Lizenzbedingungen mitteilen, die für diesen Inhalt gelten. Am einfachsten ist es, an entsprechender Stelle einen Link auf diese Seite einzubinden.

Quelle: http://creativecommons.org/licenses/by-nc-nd/3.0/ch/ Datum: 12.11.2013

Summary

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique that allows non-invasive studies of biomolecules at atomic resolution. It provides information on structure and dynamics of biomolecules and is also broadly used in small molecule characterization. This thesis explores new possibilities of NMR spectroscopy to characterize molecular and cellular systems, the chaperone–protein interactions in the periplasm of *E. coli* and the metabolism of eukaryotic cells upon external modifications.

In the first part of this thesis, basic concepts of NMR spectroscopy are described, as well as the specific NMR experiments used in the experimental part. The second part of the thesis describes the application of NMR spectroscopy to characterize chaperone–protein complexes. Site-specific intermolecular short-range contacts were detected in a membrane-protein–chaperone complex. This was achieved by an orthogonal isotope-labeling scheme that permits the unambiguous detection of intermolecular NOEs between the well-folded Skp chaperone and the unfolded outer membrane protein A substrate ensemble. The residues involved in these contacts are part of the chaperone–substrate interface. Furthermore, the interaction between the periplasmic chaperone SurA and the BamA-POTRA domains was characterized by NMR spectroscopy. This interaction is supposed to induce the delivery of unfolded outer membrane protein substrates to the BAM complex for their insertion into the outer membrane of *E. coli*. The combination of sequence-specific assignment using triple-resonance experiments and chemical shift mapping upon interaction revealed the mechanism of SurA interaction with POTRA. A destabilization of SurA and the release of a hydrophobic surface on POTRA1 upon interaction presumably lead to the handover of the OMP precursor to the Bam complex.

The third part of the thesis describes studies of cellular metabolism by NMR spectroscopy by footprinting method and in living cells. 1D NMR experiments, combined with metabolite quantification methods characterize the metabolic changes in cells infected by *S. flexneri* and provide new insights into the infection mechanism of this highly virulent bacterium. Furthermore, the potential of dissolution dynamic nuclear polarization (DNP) NMR spectroscopy in the characterization of real time metabolic processes in living macrophages was successfully explored showing that dissolution-DNP NMR spectroscopy can be applied to a broad range of cell systems, and can become routinely applied for metabolic studies in the cell.

Content

Chapter 1

Nuclear Magnetic Resonance Spectroscopy for Molecular Biology

Introduction

Nuclear magnetic resonance (NMR) spectroscopy was first reported in bulk material by Bloch *et al*.[1] and Purcell *et al*.[2] in 1946, and was recognized by the 1952 Nobel Prize in Physics. Since then, NMR spectroscopy has become a major tool in analytical chemistry and molecular biology. Application of nuclear magnetic resonance (NMR) spectroscopy to biomolecules was strongly facilitated by the development of Fourier transform NMR spectroscopy by Ernst and Anderson[3] in 1966 and the conception of multidimensional NMR spectroscopy by Jeener in 1971[4]. Since the eighties, methods have become available to obtain complete sequence-specific resonance assignments of small proteins, DNA and RNA fragments. Since then, the number of structures solved by NMR spectroscopy grew exponentially to reach an approximate of 11,000 nowadays[5]. Together with X-ray crystallography and electron microscopy, NMR spectroscopy is one of three techniques capable of determining three-dimensional structures of biological macromolecules at atomic resolution and provides, readily, information on dynamic processes. Main obstacles to the development of NMR spectroscopy of biomolecules are its inherent low sensitivity and the complexity of the measured spectra. NMR spectroscopy of proteins in solution also suffers from a "size limitation" occurring from increased transverse relaxation. As a consequence, only 2 percent of protein structures solved by NMR spectroscopy comprise more than 200 residues. However, the development of stronger magnets and more sensitive NMR spectrometers push the sensitivity. Advances in sample preparation, through the selective use of ^{13}C and ^{15}N isotopes[6] combined with multidimensional heteronuclear NMR spectroscopy techniques[7] allowed to convert the spectral complexity into significant advantage providing crucial information in distances[8], connectivity, structures[9, 10]. The development of specific protein labeling patterns[11], the use of perdeuteration[12] as well as new spectroscopic techniques such as TROSY[13] and more recently the development of reverse micelles[14],[15] raise this limitation towards large systems, in favorable cases up to 1 MDa[16, 17]. Recently the development of dissolution dynamic nuclear polarization (DNP) NMR spectroscopy has increased the sensitivity of certain NMR experiment up to 10,000 fold[18]. DNP proves to be relevant in the study of various cell metabolites[19-23] and, thanks to its continuous development, could become a very powerful tool in molecular biology.

1.1 Basic Principles of Nuclear Magnetic Resonance Spectroscopy

Spin Theory

NMR spectroscopy exploits the fact that nuclei possess a spin angular momentum characterized by the nuclear spin quantum number I. The nuclear spin angular momentum is a vector quantity $\vec{I}$ that, because of the laws of quantum mechanics, can have 2I + 1 possible orientations, described by the magnetic quantum number m = (-I, -I + 1, …, I-1, I). The value of the nuclear spin quantum number I is the combination of the individual spins ½ of each proton and neutron in the nucleus (Table 1.1). Among them, only non-zero spins interact with the magnetic B-field and are observable by NMR spectroscopy. These nuclei possess a nuclear magnetic moment $\vec{\mu}$ defined by Eq. 1 where γ is the gyromagnetic ratio, a characteristic constant for a given nucleus (Table 1.1), determining its sensitivity in NMR spectroscopy.

$$\vec{\mu} = \gamma \vec{I} = \gamma \hbar m \qquad (1)$$

Liquid state NMR spectroscopy uses mainly nuclei possessing a spin ½, as the quadrupolar moment displayed by spins I > ½ leads to broad NMR signals. The nuclei, ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P are thus the most commonly used in biomolecular NMR spectroscopy.

Nucleus	^{1}H	^{2}H	^{13}C	^{14}N	^{15}N	^{19}F	^{31}P
$\boldsymbol{\gamma/2\pi}$ $(\mathbf{Hz \cdot T^{-1}})$	$42.6 \cdot 10^6$	$6.5 \cdot 10^6$	$10.7 \cdot 10^6$	$3.1 \cdot 10^6$	$-4.3 \cdot 10^6$	$40.1 \cdot 10^6$	$17.2 \cdot 10^6$
I	1/2	1	1/2	1	1/2	1/2	1/2

Table 1.1 Gyromagnetic ratios and nuclear spin quantum numbers for nuclei of particular importance in biological NMR spectroscopy.

In the absence of an external field, the quantum states corresponding to the 2I + 1 values of m have the same energy and the nuclear spin vector does not have preferred orientation. However, placed in an external magnetic field, the different spin states acquire different energies given by Eq. 2 in which $\vec{B}$ is the magnetic field vector.

$$E = \vec{\mu} \cdot \vec{B} \qquad (2)$$

In an NMR spectrometer, this static external magnetic field is directed along the z-axis, and the energies of the m spin states are obtained by multiplying the projections of the angular moments onto the z axis by the magnetic field shown in Eq. 3 where $\hbar$ is the Planck constant and B_0 the static magnetic field strength.

$$E_m = \mu_z B_z = -m\hbar\gamma B_0 \qquad (3)$$

The interaction results in 2I + 1 energy levels, equally spaced by energy gaps $\Delta E = \hbar\gamma B_0$, known as the Zeeman levels. In the case of a nucleus with a spin quantum number I = ½ in a static magnetic field B_0, the spin states can take two orientations, parallel (m = +1/2) denoted α and antiparallel (m = -1/2) denoted β at the energies $E_\alpha = \frac{1}{2}\hbar\gamma B_0$ and $E_\beta = -\frac{1}{2}\hbar\gamma B_0$ respectively (Fig. 1.1).

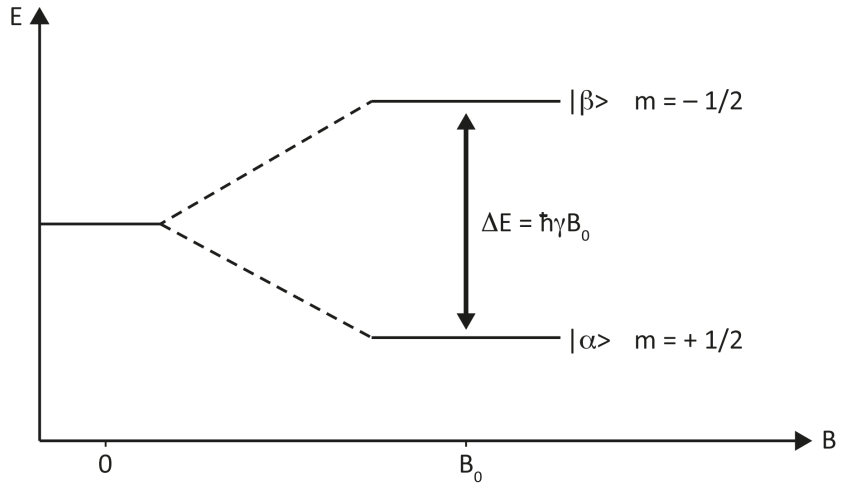


Figure 1.1 Energy levels $|\alpha\rangle$ and $|\beta\rangle$ of the two possible orientations of a nucleus possessing a spin quantum number I = ½ in a static magnetic field B_0 oriented along the z axis.

At thermal equilibrium, these different energy levels are unequally populated, as the lower energy orientation of the nuclear spins is more favorable according to Boltzmann equations. The difference of population between two consecutive states is called polarization (Eq. 4) and

determines the sensitivity of an NMR experiment. In the case of a ^{1}H nucleus, possessing a spin ½, the population difference between the α and β states in a static magnetic field of 800 MHz at temperature of 273 K it is equal to about 0.999872. This small population difference is responsible for the sensitivity of NMR spectroscopy.

$$\frac{n_\beta}{n_\alpha} = \exp\left(-\frac{\hbar\gamma B_0}{k_B T}\right) \qquad (4)$$

The One-Pulse NMR Experiment

An NMR experiment is performed by stimulating transitions between two consecutive Zeeman levels allowing net transitions of the excess of spins from the higher populated state to the consecutive one. Transitions are stimulated by applying an electromagnetic wave of energy corresponding to the difference between the levels $\Delta E = \hbar\gamma B_0$ to the system.

In a simple semiclassical vector model[1], the small population excess in the lower spin state gives rise to a net macroscopic magnetization $\vec{M}$ resulting from the superposition of the microscopic moments (Eq. 5).

$$\vec{M} = \sum_{sample}\left(n_\alpha - n_\beta\right)\vec{\mu} \qquad (5)$$

At thermal equilibrium, this bulk magnetization denoted $\overrightarrow{M_0}$ is parallel to the static magnetic field. An application of a radiofrequency (rf) pulse to the system with energy equal to the difference ΔE between two consecutive spin states (on-resonance pulse) will deviate the bulk magnetization vector from the z-axis (Fig. 1.2a). In the ideal one-pulse experiment, the on-resonance pulse B_1 achieves a rotation of the magnetization by 90° into the x,y-plane where the maximum magnetization is detected. Once in the x,y-plane, the magnetization starts to precess around the z-axis at its Larmor frequency (Eq. 6) during an acquisition period t, generating the signal recorded by the NMR spectrometer, the free induction decay (FID) (Fig 1.2b).

$$\omega_0 = \gamma B_0 = 2\pi\nu_0 \qquad (6)$$

Due to relaxation (detailed in section 1.2) the magnetization does not precess freely in the x,y-plane forever, but returns to the equilibrium state along the z-axis, leading to a decay of the FID. The time-domain FID is then Fourier transformed to produce a frequency-domain spectrum displaying a resonance peak at frequency ω_0 (Fig. 1.2c).

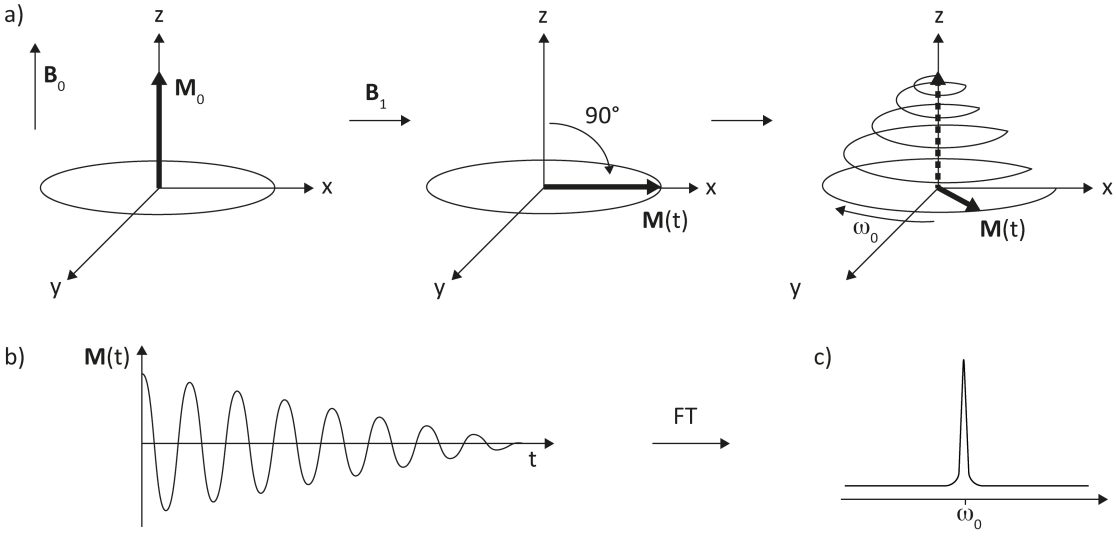


Figure 1.2 One-pulse NMR experiment. The vector quantities are highlighted in bold. a) Evolution of the magnetization during a one-pulse experiment. b) Free induction decay (FID) recorded in a one-pulse experiment and c) its Fourier transformed NMR spectrum displaying an NMR peak at the Larmor frequency ω_0.

Chemical Shift

Distinction between the different nuclei in an NMR spectrum is made possible as the resonance frequency of a given nucleus slightly differs from their Larmor frequency ω_0 (Eq. 7) due to its interaction with the local environment. This difference, called chemical shift, is caused by the motion of the electrons surrounding each spin, generating a second magnetic field acting as a chemical shield around the nuclei. Thus, the magnetic field experienced by nuclei is attenuated by a factor σB_0, where σ represents the shielding constant. This chemical shielding is very small in magnitude compare to the external magnetic field, with typical values of a few or several parts per million (ppm). Thus the resonance frequency observed for a nucleus is defined by Eq. 7 and variations in chemical shielding due to different electronic environments leads to different resonances frequencies of nuclei.

$$\omega = -\gamma(1-\sigma)B_0 \qquad (7)$$

For practical reasons, the chemical shift is defined relative to a resonance signal of a reference molecule (Eq. 8), where ω_{ref} is the resonance frequency of the reference substrate, defined as δ = 0 ppm. In biomolecular NMR of protons, the reference substrate is typically the methyl

resonance of 2.2-dimethyl-2-silapentane-5-sulfonic acid[24] (DSS), which has a lower resonance frequency than most protons in other molecules.

$$\delta = -\frac{\omega_{ref}-\omega}{\omega_{ref}} \cdot 10^6 ppm \quad (8)$$

The chemical shift of a nucleus depends on the electron distribution in its molecular orbitals and thus also on the local geometry (neighboring atoms, angles between bonds, bond lengths). As an example, the proton chemical shift strongly depends on the electronegativity of the bound nucleus, influencing the local electron density at the proton. In the case of proteins, secondary chemical shifts are also observed, dependent on dihedral angles and the type of secondary structure elements. Thus, the chemical shift contain contributions of the secondary and tertiary structure of proteins and its variation can be related to even small structural modifications. Therefore, chemical shift perturbation (CSP) studies allow to monitor changes in protein structure, such as unfolding[25], where the chemical shifts of the residues change towards their random-coil chemical shift values. CSP mapping is also a powerful technique to characterize protein-protein interactions[26]. There, changes in chemical shifts, for example of the amide moiety, are induced by the changes in local environment upon binding. The changes are quantified by Eq. 9, where $\Delta\delta(^1H)$ and $\Delta\delta(^{15}N)$ are the chemical shift changes of the ^1H and ^{15}N nuclei, respectively.

$$CSP = \sqrt{\left(\Delta\delta(^1H)\right)^2 + \left(\frac{1}{5}\Delta\delta(^{15}N)\right)^2} \quad (9)$$

Scalar Coupling

In an isolated spin system, the nucleus gives rise to one peak in the NMR spectra at its chemical shift position. However, interactions between covalently attached nuclei, mediated by the electrons forming the chemical bonds, cause the splitting of the resonance signals into multiplets[27]. This spin–spin coupling or scalar coupling is characterized by the scalar coupling constant ${}^nJ_{IS}$ (in Hertz), whose strength depends on the number of covalent bonds (n) separating the two nuclei I and S, as well as the nuclei type. For protons, detectable scalar coupling interaction typically have $n = 1$ to 4. In a two spin ½ system I and S, described by four states, $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$ and $\beta\beta$, corresponding to the two possible magnetic quantum numbers m $(\pm 1/2)$ (Fig. 1.3a), the allowed transitions occur between $|\alpha\alpha\rangle$ and $|\alpha\beta\rangle$ and between $|\beta\alpha\rangle$ and

|ββ> at the resonance frequency of the spin S ω_S, and between |αα> and |βα> and between |αβ> and |ββ> at the resonance frequency of the spin I, ω_I. If the spins are scalar coupled, the scalar coupling J_{IS} modifies the energy levels of the four spin states by a factor $\pm \frac{1}{2}\pi\hbar J_{IS}$ and thus, according to Eqs. 3 and 6, the resonance frequencies by a factor $\pm \frac{1}{2}\pi J_{IS}$. As a consequence, the NMR spectrum will display two peaks for each spin centered on the resonance frequencies ω_I and ω_S, and different by πJ_{IS} (Fig. 1.3b). The scalar coupling is used in multidimensional NMR experiments to transfer magnetization between spins.

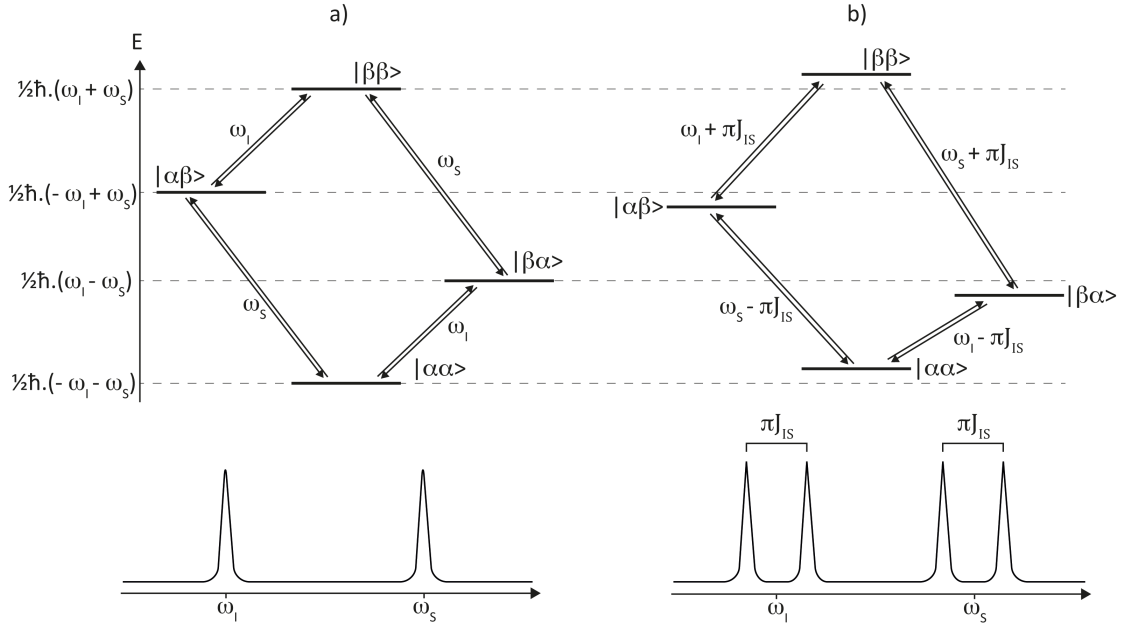


Figure 1.3 Two spin ½ system; adapted from Cavanagh, J. *et al.*[28]. a) Energy levels |αα>, |αβ>, |βα> and |ββ> of two non scalar coupled spins ½ and corresponding NMR spectrum where each non scalar coupled spin ½ gives rise to one peaks at its resonance frequency. b) Energy levels |αα>, |αβ>, |βα> and |ββ> of two scalar coupled spins ½ and corresponding NMR spectrum where each scalar coupled spin ½ gives rise to two peaks at its resonance frequency $\pm \frac{1}{2}\pi J_{IS}$.

Chemical Exchange

A modification of the chemical shift of a given nucleus can happen via perturbation of its environment. Such change can be caused by internal motion such as conformational exchange or by interaction with another molecule during complex formation. There, the chemical exchange rate constant k is defined as the rate of interconversion between these different

conformations[29]. For a nucleus exchanging between two conformations A and B, each conformation is giving rise to a distinct resonance with chemical shift values δ_A and δ_B, and with distinct frequencies ν_A and ν_B, distant of $\Delta\nu$. Possible exchange regimes between these conformations, equally populated, are displayed in Figure 1.4.

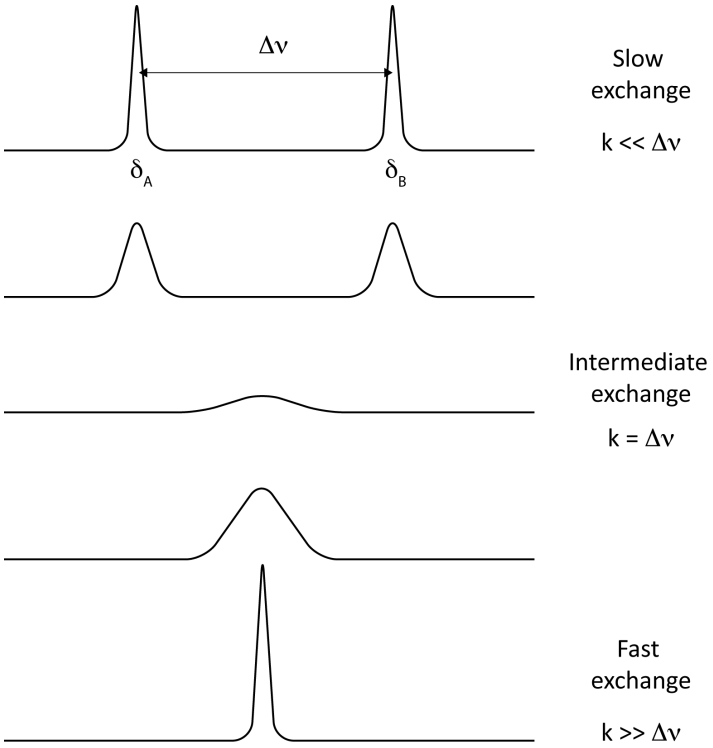


Figure 1.4. Possible chemical exchange regimes of a nucleus interconverting between two conformations A and B, with distinct chemical shift value δ_A and δ_B. k is the chemical exchange rate and $\Delta\nu$ the difference between the resonance frequencies of the nucleus in conformations A and B.

If the exchange rate is smaller than the difference in resonance frequencies $\Delta\nu$, the system is on slow exchange in the NMR time scale (millisecond regime) and the two different conformations of the nucleus are observed during the measurement time. Basically, while the NMR experiment is recorded, half of the molecules are in conformation A while the other half is in conformation B. As the rate of exchange increase, the molecule starts to swap between conformations A and B multiple times during the NMR time scale and the two original resonance frequencies of the nucleus therefore modulates each other. This causes a loss of coherence and thus a peak broadening, as well as a movement of the peaks, as the nuclei no longer have the chemical shifts δ_A and δ_B but $(n_A - n_B)\delta_A$ and $(n_B - n_A)\delta_B$ where n_A and n_B

are the proportion of time spent in conformations A and B respectively. The intermediate exchange or coalescence is reached when $\Delta\upsilon \approx k$, where the extreme broadening prevents the distinction of the peak in the NMR spectra. Finally, if the exchange rate is very large compared to the frequency difference between the two conformations, the system is in fast exchange on the NMR time scale and one single peak is observed at the average chemical shift of the two conformations. There, while the NMR experiment is recorded, the nucleus spends half of its time in each conformation.

1.2 Spin Relaxation

Introduction to Spin Relaxation

Relaxation is the process by which spins return to their equilibrium state where the population of the energy levels are those given by the Boltzmann distribution. For a spin ½, a transition from the α to the β states is induced by an electromagnetic wave, perturbing the spin population at a transition rate W modulated by the population of the α state. Then, relaxation is happening at the same transition rate W between the β to the α states modulated by the population of the β state (Fig. 1.5a). Thus, the rates of change of the population of the α and β states[30] can be described by Eq. 10 where $(n_\alpha - n_\alpha^0)$ and $(n_\beta - n_\beta^0)$ are the populations of the α and β states deviated form equilibrium ones n_α^0 and n_β^0.

$$\begin{aligned} \frac{dn_\alpha}{dt} &= -W(n_\alpha - n_\alpha^0) + W(n_\beta - n_\beta^0) \\ \frac{dn_\beta}{dt} &= -W(n_\beta - n_\beta^0) + W(n_\alpha - n_\alpha^0) \end{aligned} \qquad (10)$$

The net magnetization M_z along the z axis being equal to the population difference between the two states (Eq. 5), its evolution with time is defined by Eq. 11 where $M_0 = (n_\alpha^0 - n_\beta^0)$ is the equilibrium magnetization.

$$\frac{dM_z}{dt} = \frac{d(n_\alpha - n_\beta)}{dt} = -2W(M_z - M_0) \qquad (11)$$

$$M_z(t) = [M_z(0) - M_z^0]e^{-\frac{t}{T_1}} + M_z^0 \qquad (12)$$

Integration of Eq. 11 (Eq. 12) shows that the evolution of the z-magnetization M_z with time due to spin relaxation follows an exponential law (Fig. 1.5b, where $M_z(0)$ is the magnetization at time zero, $T_1 = \frac{1}{R_1} = \frac{1}{2W}$ is the longitudinal or spin-lattice relaxation time and R_1 the longitudinal relaxation rate. This longitudinal relaxation characterizes the process through which the magnetization goes back to its equilibrium position along the z axis through interaction with the environment, called lattice. Typical T_1 values for ^{1}H are between hundreds of milliseconds and tens of second while they are generally twice longer for ^{13}C nuclei.

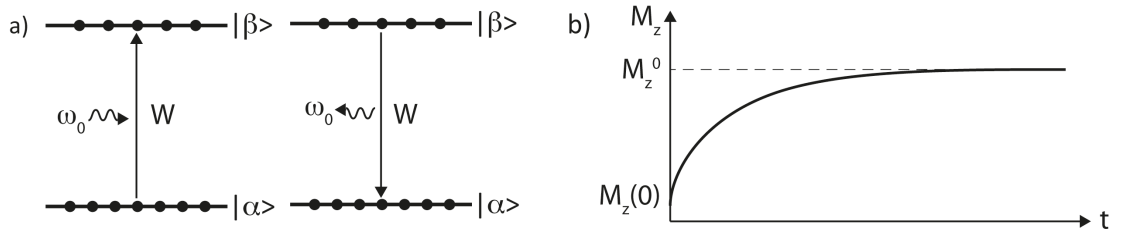


Figure 1.5. Relaxation in a spin ½ system; adapted from Keeler, J.[31]. a) Transitions between the α and β states in a spin ½ system where the transition from α and β decreases the population of the α state and a transition form β to α increase the population of the α state. b) Plot of the evolution of the z-magnetization (M_z) with time, following an exponential law, where M_z^0 is the z-magnetization at Boltzmann equilibrium and M_z (0) the magnetization at t = 0.

T_1 depends on the transition rate constant W influenced by the type of nuclei, the source of the relaxation mechanism and the magnetic field fluctuations happening in the sample. Indeed, spontaneous spin relaxation in the absence of external influences is essentially absent and for it to occur there must be magnetic field fluctuations in resonance with the energy difference between the spin states involved in the transition. Each nucleus has a local magnetic field whose direction is changing with molecular motion, giving rise to magnetic field fluctuations. These fluctuations create electromagnetic fields with wide ranges of frequencies due to the random nature of the molecular rotation. The spectral density $J(\omega)$ describes the amount of molecular motion at the frequency ω[32] (Eq. 13).

$$J(\omega) = \frac{2\tau_c}{1+\omega^2\tau_c^2} \propto W \quad (13)$$

J depends on τ_c, the molecular rotational correlation time of the molecule determined by different factors among which the molecular weight of the molecule is the most important. The larger is a molecule, the slower is its reorientation and the longer is its τ_c. The correlation between τ_c and $T_1(= 1/2W)$ is displayed in Figure 1.6, showing the influence of the molecular size on the relaxation time. In addition to molecular motions, the source of the local magnetic fields inducing relaxation influences the relaxation rate W of the nuclear spins.

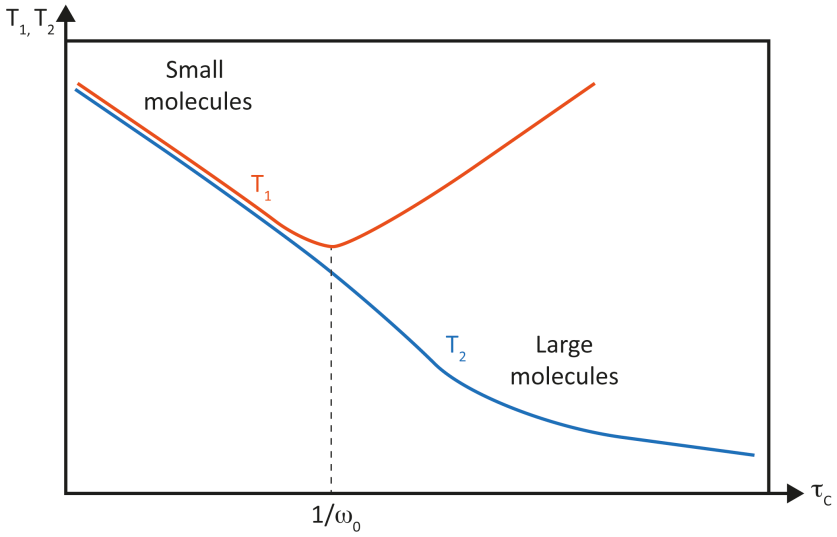


Figure 1.6. Longitudinal (T_1) and transverse (T_2) relaxation time function of the molecular rotational correlation time τ_c displayed in orange and blue respectively. Adapted from Reich, H. J.[33].

Sources of Relaxation

For spin ½ in proteins, the relevant sources of relaxation are the chemical shift anisotropy, the dipolar coupling and the paramagnetic effect. The chemical shift anisotropy (CSA) mechanism arises from the fact that the electronic environment around a nucleus is generally anisotropic. Thus, the magnitude and the direction of the additional magnetic field created by electrons surrounding the nucleus depends on the orientation of the molecule relative to the static magnetic field and modification of these fields by molecular motion induces relaxation. In NMR spectroscopy of biological molecules, while ^{13}C, ^{15}N and ^{31}P have significant CSA contributions to

relaxation, it is usually not a dominant factor for protons[28]. The dipolar relaxation mechanism is happening between pairs of dipolar-coupled spins I and S[34] where the fluctuations from the spin I will induce relaxation of the spin S and *vice versa*. This dipole–dipole relaxation is proportional to the square of the dipolar coupling constant D_{IS} (Eq. 14), implying strong dependence in the distance r_{IS} between the two nuclei. As well it depends on the gyromagnetic ratio where nuclei with high γ relax faster.

$$D_{IS} \sim \frac{\gamma_I \gamma_S \hbar}{4\pi^2 r_{IS}^3} \qquad (14)$$

In small molecules in solution, since the proton has the highest γ value of the common nuclei, it causes also the strongest dipole–dipole relaxation. Thus, dipolar relaxation is the principal relaxation pathway for protons as well as carbon and nitrogen directly attached to protons. In a dipolar coupled ^{1}H-X spin system, the dipolar relaxation rate of a nucleus X by its nearby protons is given by Eq. 15.

$$R_{1\,(DD)} = \frac{1}{T_{1\,(DD)}} = 2D_{HX}^2 \cdot J(\omega_{HX}) = 2 \cdot \gamma_H^2 \cdot \gamma_X^2 \cdot \hbar^2 \sum \frac{1}{r_{HX}^6} \cdot J(\omega_{HX}) \qquad (15)$$

As an example, the T_1 values of a ^{13}C nuclei directly bound to a proton vary typically between 0.1 and 10 seconds but longer values typically between 10 and 300 seconds are observed for quaternary carbons[35]. Thus, deuteration of molecules are commonly used to increase T_1 values[36] as the gyromagnetic ratio of deuteron is 7 times lower than the one of proton.
Finally, the paramagnetic effect[37] occurs between a spin and a paramagnetic center. As the presence of an unpaired electron is required for the effect, it is happening only in specific systems.

Nuclear Overhauser effect (NOE)

The nuclear Overhauser effect (NOE), named after its discoverer Albert Overhauser in 1953[38], consists in the transfer of magnetization between two dipolar coupled spins via cross relaxation. A NOE occurs in a two spin ½ system I and S by dipolar, and not scalar, coupling described by Figure 1.7. In such system, relaxation can happen between all energy levels and the transitions can be classified into three groups. Transition at rate constant W_1, between two consecutive energy levels, involves a spin flip of only one of the spins and corresponds to previously described T_1 relaxation. Transition at rate constant W_0 involves a simultaneous α to β flip for one

spin and β to α flip for the other spin, corresponding to a zero-quantum transition. Finally, the transition at rate constant W_2 involves a simultaneous flip of both spins in the same direction, corresponding to a net double quantum transition.

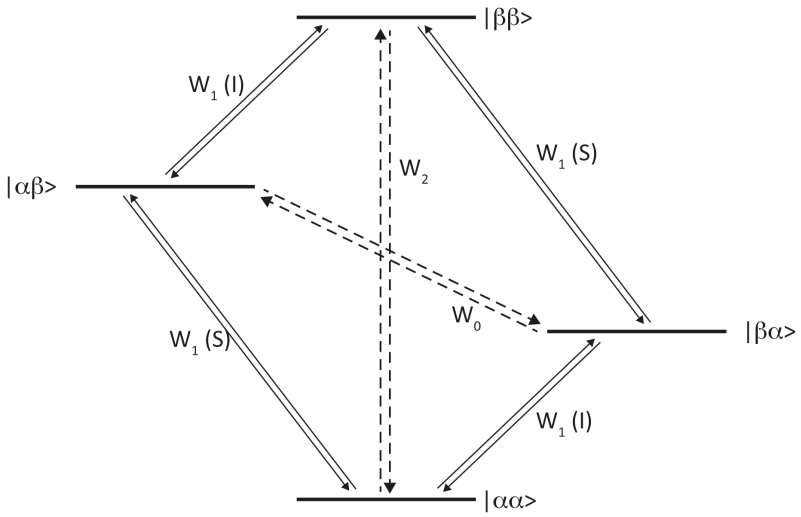


Figure 1.7. Energy diagram of a two spin ½ systems dipolar coupled in a magnetic field. The transitions displayed as solid lines correspond to the T_1 relaxation and the transitions displayed in dashed lines correspond to the cross-relaxation through dipolar coupling.

Thus, by analogy to Eq. 10, the rate of change of the populations of the different states can be described here for the|αα> state (Eq. 16).

$$\frac{dn_{|\alpha\alpha>}}{dt} = -W_1(S)n_{|\alpha\alpha>} - W_I(I)n_{|\alpha\alpha>} - W_2n_{|\alpha\alpha>} + W_1(S)n_{|\alpha\beta>} + W_I(I)n_{|\alpha\alpha>} + W_2n_{|\beta\beta>} \quad (16)$$

In this system, the net magnetizations I_z and S_z of the spins I an S along the z-axis are described as the population difference across the two I and S spins transition respectively, by Eq. 17. As well as a third combination of the population, $2I_zS_z$ is described.

$$\begin{aligned} I_z &= n_{|\alpha\alpha>} - n_{|\beta\alpha>} + n_{|\alpha\beta>} - n_{|\beta\beta>} \\ S_z &= n_{|\alpha\alpha>} - n_{|\alpha\beta>} + n_{|\beta\alpha>} - n_{|\beta\beta>} \\ 2I_zS_z &= n_{|\alpha\alpha>} - n_{|\beta\alpha>} - n_{|\alpha\beta>} + n_{|\beta\beta>} \end{aligned} \quad (17)$$

Thus, by analogy to Eq. 11, the evolutions of the spin magnetizations with time can be described, by combination of Eq. 16 and 17, by the Solomon equations[39] (Eq. 18) where the populations I_z and S_z, are the derivation from their equilibrium values I_z^0 and S_z^0.

$$\frac{dI_z}{dt} = -(2W_1(I) + W_2 + W_0)I_z - (W_2 - W_0)S_z - 2W_1(I)\ 2I_zS_z$$
$$\frac{dS_z}{dt} = -(W_2 - W_0)I_z - (2W_1(S) + W_2 + W_0)S_z - 2W_1(S)\ 2I_zS_z \qquad (18)$$
$$\frac{dS_zI_z}{dt} = -2W_1(I)I_z - 2W_1(S)S_z - (2W_1(I) + 2W_1(S))2I_zS_z$$

It shows that the rate of change of each spin magnetization depends on self-relaxation ($R_I = (2W_1(I) + W_2 + W_0)$ for spin I and $R_S = (2W_1(S) + W_2 + W_0)$ for spin S) equivalent to the above-described T_1 as well as on cross relaxation between the two spins described by $\sigma_{IS} = (W_2 - W_0)$, through double quantum and zero quantum transitions[40]. Thus, the NOE effect between two dipolar coupled nuclei I and S, depends on the cross-relaxation rate constant σ_{IS} that, using Eq. 15, can be written as

$$\sigma_{IS} = (W_2 - W_0) = 2.\gamma_I^2.\gamma_S^2.\hbar^2\frac{1}{r_{IS}^6}\left[J\left(\omega_{W_2}\right) - J\left(\omega_{W_0}\right)\right] \quad (19)$$

The frequency necessary to induce the transition at W_2 rate constant, ω_{W_2} is equal to the sum of the resonances frequencies for the two spins, $\omega_I + \omega_S$ and the frequency necessary to induce the transition at W_0 rate constant, ω_{W_0} is equal to the difference of the resonances frequencies for the two spins, $\omega_I - \omega_S$. Thus, in the case of two identical nuclei, with resonances frequencies ω_0 the cross-relaxation rate constant is

$$\sigma_{IS} \propto [6J(2\omega_0) - J(0)]\frac{1}{r_{IS}^6} = [6J(2\omega_0) - 2\tau_C]\frac{1}{r_{IS}^6} \quad (20)$$

It indicates that for $6J(2\omega_0) > J(0)$ or $2\tau_C$, the cross relaxation rate constant is positive. Therefore a short τ_C leads to a positive NOE effect. On the other hand, for long τ_C the NOE effect is negative. The crossing point is at $\omega_0\tau_C = \frac{\sqrt{5}}{2} \approx 1$, where the W_0 and W_2 effects cancel each other's and there is no NOE is observable (Fig 1.8). The NOE contains, via the cross relaxation rate constant information about molecular size as well as the distance between the two dipolar coupled nuclei through the r^{-6} dependence. This strong distance dependence prevents the detection of the NOE effect between spins far from each other. Basically, a [^{1}H,^{1}H] NOE typically occurs with distances below 6 Å in small proteins. Thus, the NOE effect is broadly used to measure intermolecular distances, where the intensity of the cross-relaxation rate constant is used as a constraint to calculate protein structures[41]. The NOE can also be used in a qualitative way. In protein assignment the NOE effect measured on the amide proton (^1H_N) allows to detect the surrounding protons, thus providing information on the amino-acid position within the protein and in unfolded ensembles, where the NOE effect contains the contribution

from different conformations of the structural ensemble (NOE measurement detailed in section 1.3).

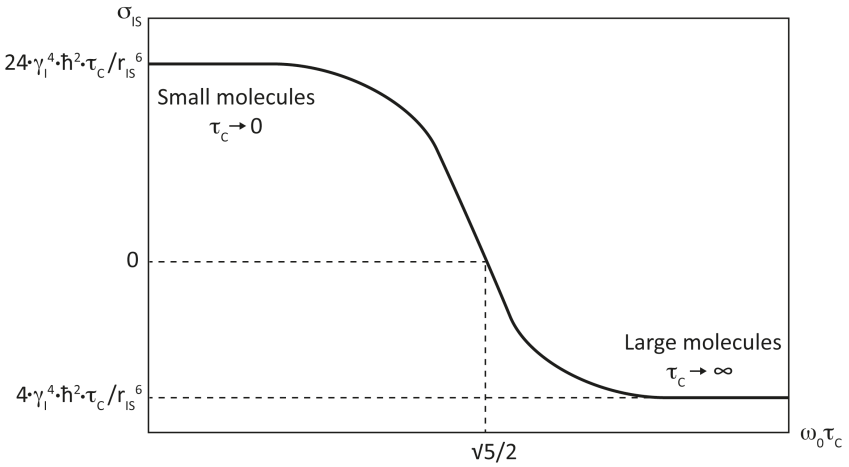


Figure 1.8. Variation of the sign of the NOE effect function of the molecular rotational correlation time τ_c, related to the size of the molecule and the resonance frequency ω_0, for two equivalent spin ½ dipolar coupled. Adapted from Gemmecker, G.[42].

Transverse Relaxation

In addition to the previously described spin-lattice relaxation, a decrease of NMR coherence over time also arises from the interaction of spins with their environment, including other spins. This effect is called transverse or spin-spin relaxation and determines the evolution of the transverse magnetization M_{xy} with time. By analogy to Eq. 12, transverse relaxation follows an exponential decay law described by Eq. 21 where $M_{xy}(0)$ is the magnetization at time zero, T_2 the transverse relaxation time constant and $R_2 = \frac{1}{T_2}$ the transverse relaxation rate constant.

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T_2}} \quad (21)$$

As for the longitudinal relaxation, this process occurs by random fluctuations of the magnetic field at the nucleus, triggering transitions of the spins between energy levels. The most relevant relaxation mechanisms in protein NMR arise from dipolar interaction and chemical shift anisotropy. Similar to the longitudinal relaxation, the transverse relaxation time constant T_2 depends on the spectral density function $J(\omega)$ (Eq. 22), where for biomolecules displaying a long τ_c ($> \omega_0$), the term $J(0)$ is dominant. Thus, in first order approximation, the relaxation time

constant T_2 for dipolar relaxation in a two-spin system is proportional to the molecular rotational correlation time τ_c and large molecules thus feature very rapid transverse relaxation (Fig. 1.6).

$$\frac{1}{T_2} \propto \left(J(0) + J(2\omega_0) + 2\,J(\omega_0)\right) = \left(2\tau_c + \frac{2\tau_c}{1+2{\omega_0}^2\tau_c^2} + \frac{4\tau_c}{1+{\omega_0}^2\tau_c^2}\right) \qquad (22)$$

The linewidth of an NMR experiment $\Delta\upsilon$, defined as the full width at half height of the resonance peak, is proportional to the relaxation time constant T_2 (Eq. 23). For an ideal Lorentzian line shape the linewidth is given by

$$\Delta\upsilon = \frac{1}{\pi T_2} \qquad (23)$$

Thus, the resonances become broader for large molecules, displaying long τ_c (Eq. 22). As relaxation is affecting both resolution and signal-to-noise ratio, an increase in molecular size leads to an overall decrease in spectral quality. This leads to a "size limitation" for solution state NMR spectroscopy. Several techniques allow to push this limit higher, such as perdeuteration of the protein[12, 36], specific labeling schemes[11] to minimize dipolar interactions, or optimal pulse sequences managing relaxation losses[13]. As a consequence, functional studies of proteins can be performed on large systems up to 1 MDa[16, 17].

1.3 Heteronuclear NMR Experiments Applied to Proteins

Amino acids, the constituent of proteins, are composed of proton, carbon, nitrogen, oxygen and sulfur, connected together intra–residually through atomic bonds, or inter–residually via secondary structure or macromolecular interactions. Thus, atomic resolution studies of proteins and protein complexes require the characterization of distances, connections and angles between these different nuclei. As discussed in section 1.1, the proton possess spin ½ and is thus well suited for liquid state NMR spectroscopy. ^{13}C and ^{15}N possess a spin ½ as well but their low natural abundance of 1.11 and 0.36 %, respectively, prevents their detection in NMR spectra. However, these heteronuclei can be specifically enriched towards 100 % ^{13}C and ^{15}N labeled by commonly used methods[6], making them suited for NMR spectroscopy. Thus, it is possible to

correlate heteronuclear resonances with protons using NMR experiments, transfering magnetization through bond or through space between these nuclei[43]. In heteronuclear experiments, the application of a radiofrequency pulse will rotate the spin magnetization to the x,y-plane followed by coherence transfer between the different coupled spins and then detection during an acquisition period. The heteronuclear NMR experiments generally used are heteronuclear multiple quantum coherence (HMQC)[44, 45] or heteronuclear single quantum coherence (HSQC)[7] mechanisms to transfer coherence between the spins. The sensitivity of heteronuclear experiments, calculated as the signal to noise ratio $\left(\frac{S}{N}\right)$, depends on the gyromagnetic ratios of the different nuclei as well as on the spin-lattice relaxation described in Eq. 24 where γ_{ex} and γ_{det} corresponds to the gyromagnetic ratios of the exited and the detected nucleus, T_C to the recycle time of the NMR experiment and $R_{1,ex}$ to the spin lattice relaxation rate of the exited nucleus. Thus, to increase sensitivity, indirect detection through proton is favored[46, 47].

$$\frac{S}{N} \propto \gamma_{ex}\gamma_{det}^{3/2}\left[1 - exp\left(R_{1,ex}T_C\right)\right] \quad (24)$$

Heteronuclear experiments are commonly used in protein studies to control sample quality, measure distances between nuclei, assign proteins, calculate structures and their large variety allows obtaining all necessary information for macromolecule characterization. In addition, they increase the NMR spectral dimensionality, thus provide better resolution, necessary for large macromolecules where the large number of resonances prevents their unambiguous detection in the highly overlapping one or multi–dimensional homonuclear NMR spectra.

2D [^{15}N,^{1}H]-HSQC, the Protein Fingerprint

The 2D [^{15}N,^{1}H]-HSQC pulse sequence displayed in Figure 1.9, correlates scalar coupled ^{15}N and ^{1}H nuclei and can be divided in four periods: the insensitive nuclei enhanced by polarization transfer (INEPT) element[48], transferring the polarization form the sensitive proton to the ^{15}N nuclei through one–bond scalar-coupling ($^{1}J_{NH}$), the evolution period, where the chemical shift of the ^{15}N nuclei is encoded, the reverse INEPT, transferring the polarization back to the proton and the acquisition, encoding the proton chemical shift.

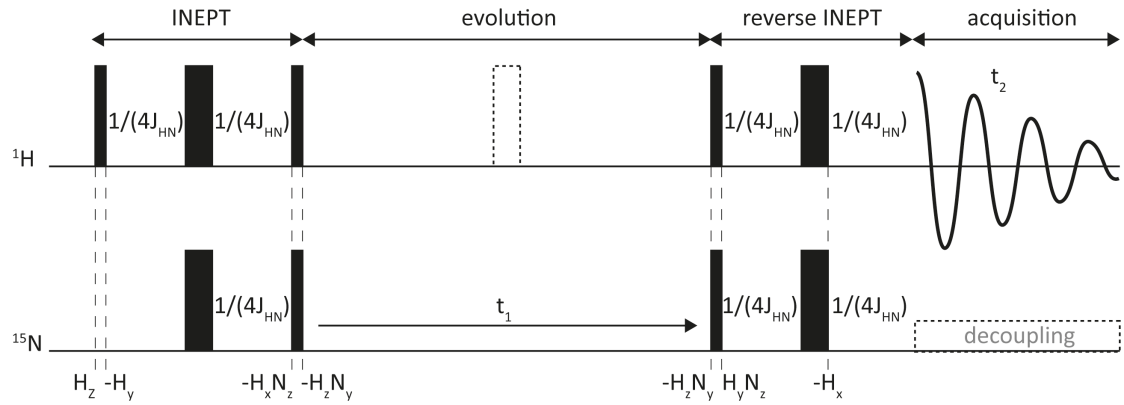


Figure 1.9. 2D [^{15}N,^{1}H]-HSQC pulse sequence where thin bars represents 90° pulses and thick bars 180° pulses. The dashed blocks represents the elements added to HSQC to decouple proton and nitrogen nuclei.

During the first INEPT block, the equilibrium H_z magnetization is converted to transverse magnetization $-H_y$ by a proton 90° pulse. The 180° pulse in the middle of the $1/(2J_{HN})$ time period allows evolution of the one–bond scalar coupling constant between the two nuclei while preventing proton chemical shift evolution. An antiphase ^{1}H magnetization with respect to ^{15}N is obtained ($-H_xN_z$). At the end of the INEPT block, the two simultaneous 90° pulses transfer the polarization from proton to nitrogen to obtain an antiphase ^{15}N magnetization with respect to ^{1}H ($-H_zN_y$). Then, the ^{15}N magnetization evolves during the t_1 time period, where the detection of the ^{15}N resonance frequencies is achieved by recording a series of experiments incrementing t_1. Finally, the magnetization is converted back into in-phase ^{1}H magnetization ($-H_x$) by a reverse INEPT block and the ^{1}H resonance frequencies are recorded during the acquisition time t_2. The proton and nitrogen resonance frequencies recorded during t_1 and t_2 evolution times give rise to four peaks in the 2D [^{15}N,^{1}H]-HSQC NMR spectrum for each ^{1}H-^{15}N moiety at frequencies $\omega_H \pm \pi J_{HN}/2$ and $\omega_N \pm \pi J_{HN}/2$, separated in each dimension by the one–bond scalar coupling constant $J_{HN} (\approx 92\ Hz)$ (Fig. 1.10a). In order to simplify the 2D [^{15}N,^{1}H]-HSQC spectrum, the heteronuclei can be decoupled, suppressing the evolution of the one–bond scalar coupling constant during t_1 and t_2. The decoupling of proton during ^{15}N evolution is achieved by application of a 180° proton pulse in the middle of the evolution period that refocuses the J_{HN} coupling, and the decoupling of ^{15}N during proton acquisition is achieved by applying a series of ^{15}N pulses continuously inverting the direction of the spin (Fig 1.9, dashed elements). Thus, the NMR spectrum displays a single resonance for each ^{1}H-^{15}N moiety at the resonance frequency of

the proton ω_H, in the direct dimension, and nitrogen ω_N, in the indirect dimension (Fig. 1.10b). In general, the HSQC is run in decoupled mode, unless measurement of J_{HN} is desired.

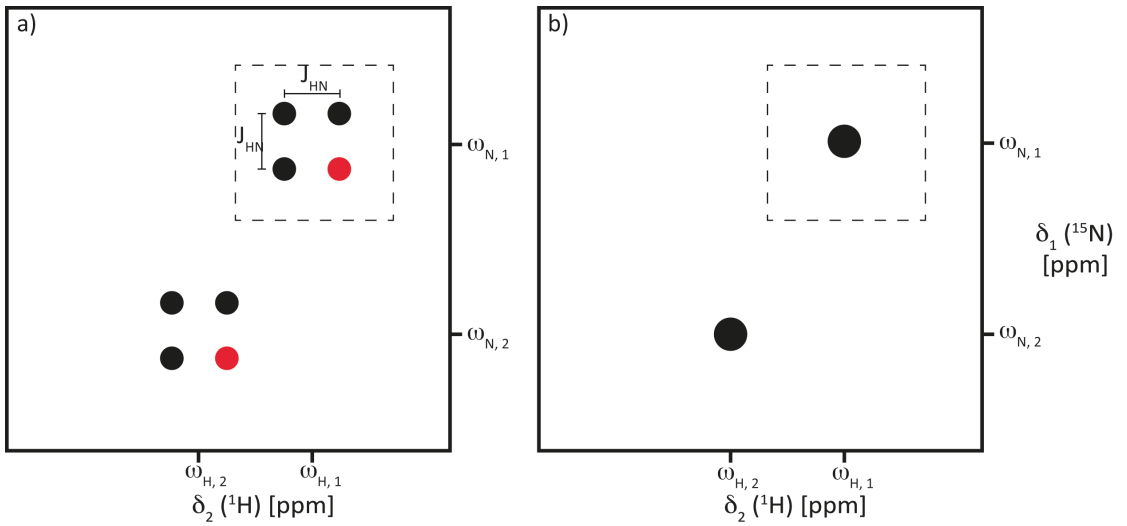


Figure 1.10. Schematic HSQC spectra; adapted from Pascal, S. M.[49] a) Zoom of a 2D [^{15}N,^{1}H]-HSQC spectrum of protein on a region showing peaks arising from two ^{1}H-^{15}N moieties of resonances frequencies $(\omega_{H,1}, \omega_{N,1})$ and $(\omega_{H,2}, \omega_{N,2})$. The dashed box contains peaks arising form one amino acid. The TROSY peak described in the following section is highlighted in red. b) 2D decoupled-[^{15}N,^{1}H]-HSQC spectrum on the same system than a).

In protein studies, 2D [^{15}N,^{1}H]-HSQC is one of the simplest and more useful two-dimensional pulse sequence. Each amino acid except proline contains an amide proton bound to a nitrogen. Thus, the 2D [^{15}N,^{1}H]-HSQC spectrum displays a peak for each amino acid, providing a fingerprint of the protein. In addition, it also contains peaks corresponding to the NH_2 groups of the side chains of asparagine and glutamine and the indole moieties of tryptophan, appearing in specific regions of the spectrum. Indeed, each ^{1}H-^{15}N moiety display a characteristic chemical shift, containing a random chemical shift part, specific for each amino acid[50], and a secondary chemical shift dependent on the protein structure. Thus, 2D [^{15}N,^{1}H]-HSQC spectra provide information on the protein secondary structure elements where amino acids involved in a β-sheet experience downfield shift (higher ppm value) compared to random proton chemical shift value[51], while amino acids involved in α-helix experience upfield shift[52]. Residues located in loops or unfolded protein regions display random coil chemical shift values, typically between 7.5 and 8.5 ppm. 2D [^{15}N,^{1}H]-HSQC experiments are broadly used to characterize protein folding or unfolding, following the proton chemical shift changes of the residues away from or towards

random chemical shift as well as protein interactions, following chemical shift changes upon titration with a binding partner.

Transverse Relaxation-Optimized Spectroscopy (TROSY)

HSQC spectra of large macromolecules (> 10 kDa) suffer from fast transversal relaxation due to the longer molecular rotational correlation time τ_c, leading to poor sensitivity (Eq. 22). At high magnetic field strengths, the major sources of relaxation in the amide moiety are chemical shift anisotropy (CSA) and dipolar interactions. Dipolar interaction to remote protons can be reduced by complete or partial deuteration of the proteins. To further reduce T_2 relaxation, transverse relaxation-optimized spectroscopy (TROSY) uses constructive interference between the N–H dipolar interactions and the ^{15}N CSA[13, 53]. In a sample, the direction of the z-component of the local magnetic field caused by dipolar interaction is either positive or negative if the two spins involved in the interaction are in a parallel or antiparallel orientation respectively. On the other hand, the direction of the z component of the local magnetic field caused by CSA is positive, irrespectively of the spins orientation. Thus, these two local magnetic fields partially cancel in an antiparallel orientation of the spins. In a non-decoupled 2D [^{15}N,^{1}H]-HSQC spectrum, this orientation corresponds to the peak located at the bottom right position of the ^{1}H–^{15}N quadruplet (highlighted in red in Fig. 1.10a). The TROSY experiment selects this TROSY peak to become the only one to be displayed in the NMR spectrum, extending the size limitation of proteins above 1 MDa in functional studies[17, 54-56].

3D NOESY-TROSY experiment

The above described HSQC or HMQC pulse sequences as well as their TROSY versions can be further combined with hetero- or homonuclear pulse sequences towards heteronuclear-edited NMR experiments[57, 58]. Such experiments yield additional information on proteins, such as relaxation parameters, distances or J-correlations between amino acids. The heteronuclear-edited 3D NOESY-HSQC[8] (or HMQC) experiment is highly useful to facilitate assignments of the [^{1}H–^{1}H] NOE cross peaks even in crowded spectral regions by adding an indirect dimension. The heteronuclear-edited 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY consists of a two-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) experiment[59, 60] combined with a TROSY

version of the HSQC pulse sequence. The NOESY experiment correlates protons that are close in space. Its pulse sequence (Fig. 1.11a) is composed of a first 90°-t_1-90° block that encodes the ^{1}H resonance frequencies and returns the magnetization to the z-axis. Then, the z-magnetization is transferred by the NOE effect to all dipolar coupled protons during a defined mixing time (τ_m). Finally, the magnetization is converted into a detectable in-phase ^{1}H magnetization by a 90° pulse and the ^{1}H resonance frequencies are recorded during the acquisition time t_2 (Fig. 1.11b).

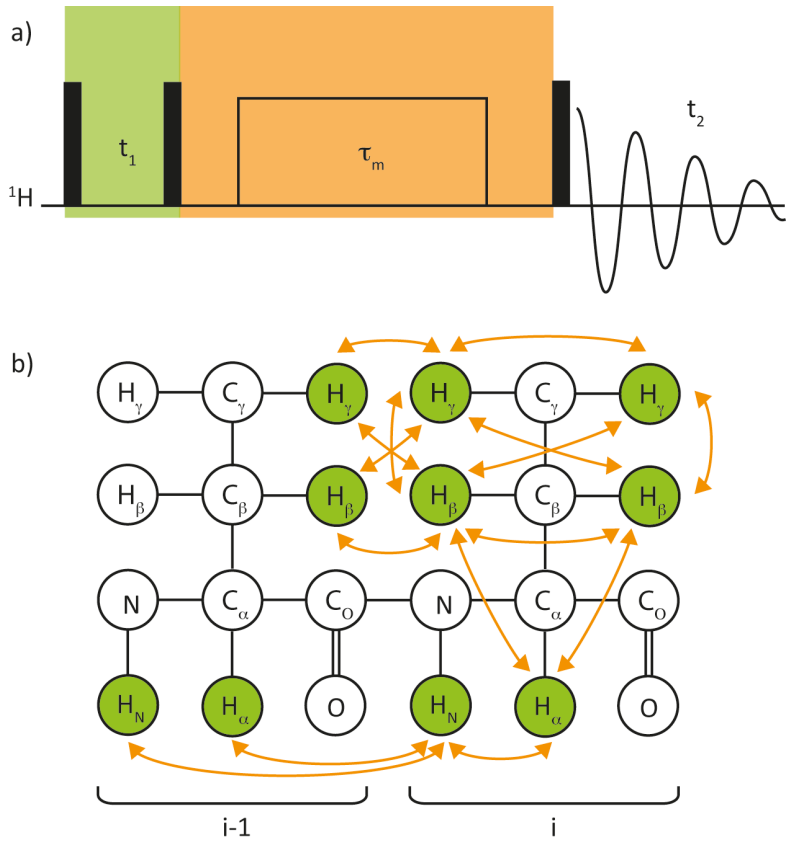


Figure 1.11. NOESY experiment; adapted from Higman, V. A.[61]. a) 2D [^{1}H,^{1}H]-NOESY pulse sequence where thin bars represents 90° pulses. b) Polarization transfer during a 2D [^{1}H,^{1}H]-NOESY experiment where the resonance frequencies of the protons, highlighted in green, are encoded during the 90°-t_1-90° and acquisition block (highlighted in green) and the magnetization transfer represented by orange arrows is happening during the mixing period highlighted in orange on the corresponding pulse sequence.

Thus for each proton H_i, the 2D [^{1}H,^{1}H]-NOESY spectrum contains a diagonal peak at frequency $(\omega_{H_i}, \omega_{H_i})$ and n cross peaks at frequencies $(\omega_{H_i}, \omega_{H_{jn}})$ arising from the interaction of H_i with its n dipolar coupled protons. The NOE cross peaks can be negative or positive according to the

molecular size (Fig. 1.8) and their intensities, which are proportional to the dipolar cross-correlation rate constants $\sigma_{H_iH_{jn}}$, contain distance information between the two coupled spins, $r_{H_iH_{jn}}$ (Eq. 20). Thus, intra- and inter-proton distances can be approximated, from a reference distance r_{ref} known by covalent geometry, using Eq. 25 where I_{ref} and $I_{H_iH_{jn}}$ are the integrated cross peak intensities.

$$r_{H_iH_{jn}} = r_{ref}\left(\frac{I_{ref}}{I_{H_iH_{jn}}}\right) \quad (25)$$

Importantly, the quantitative conversion of NOE cross peaks into interspin distances benefits from the use of short mixing times that avoid spin diffusion effects. Indeed, for long mixing times, the magnetization transferred between two spins is further relayed to other nuclei, modifying the intensity–distance correlation. Importantly, the conversion of NOE cross peaks into interspin distances can be performed correctly only for a single structural conformation. In a dynamic ensemble, the NOE cross peaks will contain contributions from different conformations, thus prohibiting their quantitative use.

In the 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY pulse sequence, the magnetization transferred by dipolar coupling between ^{1}H spins is then going through the HSQC pulse sequence (Fig 1.9), where it is transferred to ^{15}N-bound nuclei by an INEPT element through scalar coupling to encode the ^{15}N resonance frequency and brought back to the bound proton for detection. The 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY spectrum thus displays for each amide proton H_i a diagonal peak at frequency $\left(\omega_{H_i}, \omega_{N_i}, \omega_{H_i}\right)$, where ω_{N_i} is the resonance frequency of the $^{15}N_i$ nucleus of the ^1H_i-$^{15}N_i$ amide group, and n cross peaks at frequencies $\left(\omega_{H_{jn}}, \omega_{N_i}, \omega_{H_i}\right)$ corresponding to the interaction of the ^1H_i-$^{15}N_i$ amide group with its n dipolar coupled protons. In order to obtain information on protein side chains, the ^{15}N heteronucleus can be substituted by ^{13}C and the spectrum obtain will display cross peaks between protons and ^{13}C bound protons. Besides providing distance measurements between dipolar coupled protons, the 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY experiment is commonly used for protein assignment, where NOE cross-peaks are observed between amide protons (H_N) and neighboring protons with sufficiently short ^{1}H-^{1}H distance[41]. Indeed, cross peaks can often be observed between H_N of residues i and i±1. As well, cross peaks between H_α and H_N can often be observed between residues i and i+3 or i+4 in α-helical secondary structures, as well as between residues i and the facing residue in β-sheets and i and i+2 in tight turns. However, the increasing number of amide moieties in large molecules can lead to strong

chemical shift overlap preventing sequential assignment of protein backbone residues and alternative strategies have to be applied.

The HNCACB as an Example of Triple-resonance Experiments for Protein Assignment

As a method for sequential resonance assignment, a large number of three- and four-dimensional heteronuclear triple-resonance experiments correlating backbone $^{1}H_N$, ^{15}N, $^{1}H_{\alpha,\beta}$, $^{13}C_{\alpha,\beta}$ and $^{13}C'$ nuclei using scalar coupling have been developed[9, 10]. Some are based on ^{15}N–$^{13}C'$ transfer, connecting the amide group of residue i to $^{13}C'$ of residue i-1 as such as the HNCO[62] or involve ^{15}N–$^{13}C_{\alpha,\beta}$ transfer between the amide group and $^{13}C_{\alpha,\beta}$ of residue i and i-1 such as the HNCA or the HNCACB[8]. In standard nomenclature, the name of an experiment is formed by the nuclei involved in the coherence transfer in the order following the transfer pathway from the proton. Nuclei are given in brackets if their chemical shifts do not evolve.

The HNCACB experiment (Fig. 1.12), is a very useful triple-resonance experiment for protein assignment as it provides sequential information, connecting the ^{1}N–^{15}H moieties of residue i to $^{13}C_\alpha$ and $^{13}C_\beta$ of residues i and i-1. The HNCACB experiment consists of an INEPT block transferring the magnetization from $^{1}H_N$ to the scalar coupled nucleus ^{15}N via $^{1}J_{HN}$. Using another INEPT block, the magnetization is then transferred from ^{15}N to $^{13}C_\alpha$ via $^{2}J_{NC_\alpha}$. Then, part of the magnetization is transferred on $^{13}C_\beta$ using an INEPT block followed by an evolution period where the $^{13}C_{\alpha,\beta}$ resonance frequencies are encoded by t_1 increments. The magnetization is transferred back to $^{13}C_\alpha$ and then to ^{15}N using reverse INEPTs where the ^{15}N resonance frequency is recorded during the time period t_2. Finally, magnetization is transferred back to the scalar coupled amide proton for detection during the acquisition time t_3 (Fig. 1.12a). The resulting 3D HNCACB spectrum features for each ^{1}N–^{15}H amide moiety up to four cross peaks at positions $(\omega_{C_{A,i}}, \omega_{H_{N,i}}, \omega_{N_i})$, $(\omega_{C_{B,i}}, \omega_{H_{N,i}}, \omega_{N_i})$, $(\omega_{C_{A,i-1}}, \omega_{H_{N,i}}, \omega_{N_i})$ and $(\omega_{C_{B,i-1}}, \omega_{H_{N,i}}, \omega_{N_i})$ (Fig 1.12b). As each spectral strip displays the $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts of residues i and i-1, multiple strips can be linked together allowing sequential connection of amino acids (dashed lines on Fig. 1.12b). In addition, characteristic $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts of alanine, threonine, serine and glycine residues allow their identification form the HNCACB strips. This information coupled with the alignment of sequentially connected spin systems with the known amino acid sequence thus allows unambiguous sequential assignment.

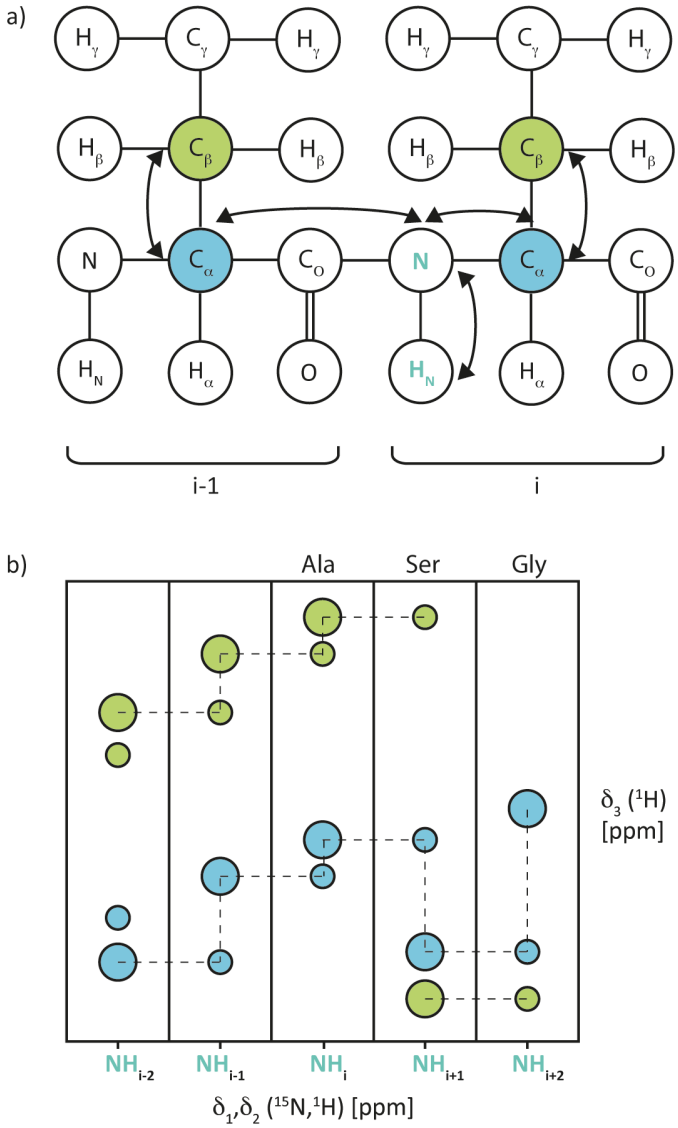


Figure 1.12. HNCACB experiment, adapted from Higman, V. A.[61]. a) Polarization transfer during a 3D HNCACB pulse sequence b) Schematic strips of ^{1}N-^{15}H amide moieties of a 3D HNCACB spectrum, where the cross peaks of the amide moiety with $^{13}C_{\alpha}$ are displayed in blue and $^{13}C_{\beta}$ in green. The less intense cross peaks with nuclei of residue i-1 are represented with smaller circles. The dashed lines represent the sequence connectivity between the stripes. The three last strips contain the typical $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts of alanine, serine and glycine.

1.4 Dissolution Dynamic Nuclear Polarization NMR Spectroscopy

Despite the powerful information provided by NMR spectroscopy, the technique suffers from a limited sensitivity, narrowing its application to samples with high concentration of active nuclei and requiring labeling strategies for biomolecular application. The thermal polarization of nuclei is very low and, even when using the highest available external magnetic fields, can not go greater than 0.01 percent. However, the sensitivity of NMR spectroscopy has dramatically increased since its discovery by Bloch *et al*.[1] and Purcell *et al*.[2] in 1946, through instrumental innovations as well as methodological developments such as Fourier transformation[3], improved polarization transfer schemes[48] and is now broadly used to study biomolecules. An additional approach to increase NMR sensitivity is the use of hyperpolarization, producing a spin population difference greater than the one obtained at thermal Boltzmann equilibrium. Several hyperpolarization techniques have been developed such as para-hydrogen induced polarization (PHIP)[63], optical pumping[64] and dynamic nuclear polarization (DNP)[65], each of which can increase the NMR sensitivity by several orders of magnitude. PHIP generates hyperpolarized states by particular chemical addition reactions, while optical pumping is limited to certain noble gases. DNP however can be applied to any sample containing molecules with unpaired electron, broadening its use to a large range of biological systems. DNP was first discovered by Overhauser in 1953[38] and was soon experimentally applied in conducting solid metals by Carver and Slichter[66], then later in solution[67]. DNP is achieved by applying continuous high-power microwave irradiation at or near the electron Larmor frequency to drive polarization transfer from the large electron spin polarization to the nuclear system[65]. DNP application to solid state NMR spectroscopy was developed in the eighties[68, 69] and expanded to a large range of biological systems by the development of improved instrumentation, such as high-power microwave sources, as well as by chemical advancements with the development of biradicals as electron source. Besides its application to solid systems, dissolution DNP was developed, for applications in solution first for *in vivo* MRI purposes[18, 70]. Dissolution DNP enables close to

complete polarization of organic molecules in a liquid solution allowing tracing of a hyperpolarized marker *in vivo* and its metabolic products. Dissolution DNP was successfully applied to systems *in vitro*, in combination with the use high-resolution NMR spectrometers, were the metabolisms of several cell systems, such as *E. coli* or *S. cerevisiae* were followed in real time[19-23].

Principle of Dynamic Nuclear Polarization

The sensitivity of an NMR experiment is proportional to the equilibrium polarization, the population difference between the basic energy states, described in Eq. 4 for a spin ½ in an external magnetic field. Proton spin polarization in a static magnetic field of 800 MHz at temperature near 273 K is about $1.28 \cdot 10^{-4}$, and this polarization can be significantly increased by lowering the temperature or increasing the magnetic field strength. Furthermore spins with high gyromagnetic ratio have increased polarization, such as the electron spin γ_e that features, compared to nuclear spin γ_I, a signal enhancement of the order of γ_e/γ_I corresponding to a factor 660 for ^{1}H and 2600 for ^{13}C. Furthermore, by decreasing temperature to 1 K, the electron spin virtually reaches its maximum polarization $P_e \sim 1$. The DNP principle consists of transferring this high electron spin polarization to the nuclear spins at low temperature. This polarization transfer can occur through two different mechanisms: the solid effect and the thermal mixing[65]. The solid effect[71] is happening in a system where electrons are diluted (< 1 mM) in an abundant nuclear spin matrix, leading to a one-to-one electron–nucleus dipole–dipole interaction. The transfer of the electron spin polarization to the dipolar coupled nucleus I is induced by microwave irradiation at frequency $\omega = \omega_e \pm \omega_I$ where ω_e is the electron spin resonance frequency and ω_I the nuclear spin resonance frequency (Fig. 1.13). These transitions, corresponding to the W_0 and W_2 transitions described for two dipolar coupled nuclear spins (Fig 1.7), are followed by electron relaxation T_{1e}, bringing back the electron spin to its equilibrium state, thus leading to high polarization P_I of the nucleus. The nuclear polarization decay, described by the longitudinal spin relaxation T_{1I} of the nucleus, typically happens on the order of an hour.

In a sample containing higher than ~ 1 mM radical concentration, the thermal mixing effect[72] is predominant. There, the ensemble of electrons and nuclear spins behave like energy reservoirs and can be described using a thermodynamic model[73]. The application of microwave irradiation

to the system cools down the electron reservoir to a temperature T_S. Thus, the nuclear spin reservoir tends to equilibrate his temperature by thermal contact to the electron reservoir temperature and the final nuclear polarization is given by Eq. 4 where the sample temperature is replaced by T_S.

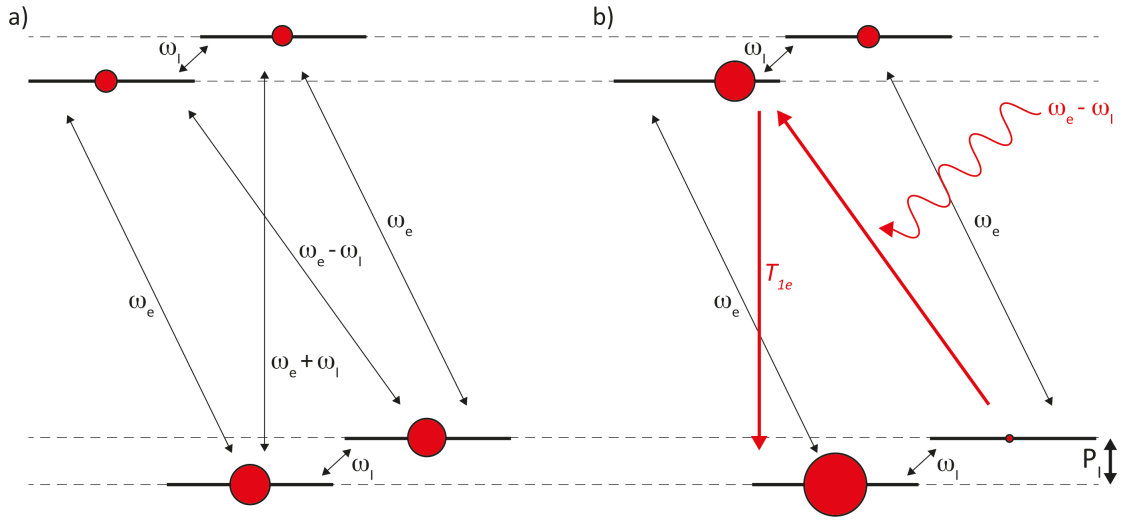


Figure 1.13. Solid effect in a system composed of one nuclear spin ½ and one electron spin; adapted from Bornet, A. et al.[74] a) Electron-spin dipolar coupled energy levels. The red balls represent the amount of spins in each level. b) Microwave of energy $\omega_e - \omega_I$ applied to electron-spin dipolar coupled energy levels, leading to the transfer of spin electron polarization to nuclear spin polarization (highlighted in red). The red balls represent the amount of spins in each level. P_I corresponds to the nuclear spin polarization obtained.

Dissolution dynamic nuclear polarization NMR spectroscopy

Application of DNP in solution, by transferring electron polarization to organic molecules in a liquid solution was first desribed by Arednkjaer-Larsen *et al.* in 2003[18]. Typically, the hyperpolarization is generated at temperatures between 1 and 4.2 K and at high static magnetic fields of e.g. 3.4 T[18, 75]. The hyperpolarized samples are glass-forming solutions containing the target molecule as well as an organic stable free radical. As dissolution DNP relies on the thermal mixing effect, radicals, such as TEMPO, are used to fulfill the DNP conditions. Subsequently, the hyperpolarized sample is rapidly dissolved by a burst of hot D_2O vapor to obtain a dilute solution at room temperature while preserving the enhanced nuclear spin polarization. This

hyperpolarized solution can be transferred to a high-resolution NMR spectrometer. The polarization enhancements of the dissolved samples can increase the sensitivity of certain NMR experiments up to 10,000 fold[18]. However, the dissolution DNP method is limited by the short lifetime of the hyperpolarization in solution, of the order of T_1. Thus, high sensitive protons with short T_1 values are difficult to detect and low-γ nuclei like ^{13}C are favored. In addition, the length of the measurement window is limited, allowing detection of fast processes only. However, the development of new polarization transfer schemes[76], progress in radical design as well as hyperpolarized molecules exhibiting longer lifetime[77] could partially overcome this issue. Hence, dissolution DNP, by increasing the sensitivity of molecular probes of several orders of magnitude[78-80], opens a broad range of *in vitro* applications such as metabolic cells studies[19-23].

References

[1] F. Bloch, *Phys. Rev.* **1946**, *70*, 460–474.

[2] E. M. Purcell, H. C. Torrey, R. V. Pound, *Phys. Rev.* **1946**, *63*, 37–38.

[3] R. R. Ernst, W. A. Anderson, *Rev. Sci. Instr.* **1966**, *37*, 93–102.

[4] J. Jeener, *Lecture, Ampère Summer School, Basko Polje, Yougoslavia*, **1971**.

[5] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.

[6] L. P. McIntosh, F. W. Dahlquist, *Q. Rev. Biophys.* **1990**, *23*, 1–38.

[7] T. J. Norwood, J. Boyd, J. E. Heritage, N. Soffe, I. D. Campbell, *J. Magn. Reson.* **1989**, *87*, 488–501.

[8] E. R. P. Zuiderweg, S. W. Fesik, *Biochemistry* **1989**, *28*, 2387–2391.

[9] M. Ikura, L. E. Kay, A. Bax, *Biochemistry* **1990**, *29*, 4659–4667.

[10] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, *J. Magn. Reson.* **1990**, *89*, 496–514.

[11] R. Kerfah, M. J. Plevin, R. Sounier, P. Gans, J. Boisbouvier, *Curr. Opin. Struct. Biol.* **2015**, *32*, 113–122.

[12] K. H. Gardner, L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357–406.

[13] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371.

[14] K. G. Valentine, G. Mathies, S. Bedard, N. V. Nucci, I. Dodevski, M. A. Stetz, T. V. Can, R. G. Griffin, A. J. Wand, *J. Am. Chem. Soc.* **2014**, *136*, 2800–2807.

[15] I. Dodevski, N. V. Nucci, K. G. Valentine, G. K. Sidhu, E. S. O'Brien, A. Pardi, A. J. Wand, *J. Am. Chem. Soc.* **2014**, *136*, 3465–3474.

[16] V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, *J. Am. Chem. Soc.* **2003**, *125*, 10420–10428.

[17] J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, *Nature* **2002**, *418*, 207–211.

[18] J. H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning, K. Golman, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10158–10163.

[19] T. Harris, H. Degani, L. Frydman, *NMR Biomed.* **2013**, *26*, 1831–1843.

[20] S. Meier, P. R. Jensen, J. O. Duus, *FEBS Lett.* **2011**, *585*, 3133–3138.

[21] S. Meier, P. R. Jensen, J. O. Duus, *ChemBioChem* **2012**, *13*, 308–310.

[22] S. Meier, M. Karlsson, P. R. Jensen, M. H. Lerche, J. O. Duus, *Mol. BioSyst.* **2011**, *7*, 2834–2836.

[23] M. E. Merritt, C. Harrison, C. Storey, F. M. Jeffrey, A. D. Sherry, C. R. Malloy, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19773–19777.

[24] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *Eur. J. Biochem.* **1998**, *256*, 1–15.

[25] V. Saudek, A. Pastore, M. A. Castiglione Morelli, R. Frank, H. Gausepohl, T. Gibson, F. Weih, P. Roesch, *Protein Eng.* **1990**, *4*, 3–10.

[26] P. Rajagopal, E. B. Waygood, J. Reizer, M. H. Saier, Jr., R. E. Klevit, *Protein Sci.* **1997**, *6*, 2624–2627.

[27] N. F. Ramsey, E. M. Purcell, *Phys. Rev.* **1952**, *85*, 143–144.

[28] J. Cavanagh, W. J. Fairbrother, P. A. G. Iii, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, CA, **1996**.

[29] A. D. Bain, *Prog. Nucl. Magn. Reson. Spectrosc.* **2003**, *43*.

[30] N. Bloembergen, E. M. Purcell, R. V. Pound, *Phys. Rev.* **1948**, *73*, 679–712.

[31] J. Keeler, *Understanding NMR Spectroscopy*, **2002**.

[32] P. Luginbühl, K. Wüthrich, *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 199–247.

[33] H. J. Reich, *Structure Determination Using NMR*, **2015**.

[34] A. Abragam, *Principles of Nuclear Magnetism*, Oxford University Press, **1961**.

[35] D. L. Pavia, G. M. Lampman, G. S. Kriz, J. A. Vyvyan, *Introduction to Spectroscopy, 5th Edition*, Cengage Learning, **2015**.

[36] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy, 2nd edition*, Elsevier Academic Press, **2007**.

[37] N. Bloembergen, L. O. Morgan, *J. Chem. Phys.* **1960**, *34*, 842–850.

[38] A. W. Overhauser, *Phys. Rev.* **1953**, *92*, 411.

[39] I. Solomon, *Phys. Rev.* **1955**, *99*, 559–565.

[40] M. Goldman, *J. Magn. Reson.* **1984**, *60*, 437–452.

[41] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Widley-Interscience, U.S.A., **1986**.

[42] G. Gemmecker, *Advanced NMR Spectroscopy*, **1999**.

[43] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **1976**, *64*, 2229–2246.

[44] L. Müller, *J. Am. Chem. Soc.* **1979**, *101*, 4481–4484.

[45] A. Bax, M. Ikura, L. E. Kay, D. A. Torchia, R. Tschudin, *J. Magn. Reson.* **1989**, *86*, 304–318.

[46] R. R. Ernst, A. A. Maudsley, *Chem. Phys. Lett.* **1977**, *50*, 368–372.

[47] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185–189.

[48] G. A. Morris, R. Freeman, *J. Am. Chem. Soc.* **1978**, *101*, 760–762.

[49] S. M. Pascal, *NMR primer: an HSQC-based approach*, IM Publications LLP, **2008**.

[50] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2008**, *36*, D402–408.

[51] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *5*, 67–81.

[52] D. S. Wishart, B. D. Sykes, D. R. Richards, *Biochemistry* **1992**, *31*, 1647–1651.

[53] R. Riek, K. Pervushin, K. Wüthrich, *Trends Biochem. Sci.* **2000**, *25*, 462–468.

[54] C. Fernandez, G. Wider, *Curr. Opin. Struct. Biol.* **2003**, *13*, 570–580.

[55] R. Riek, G. Wider, K. Pervushin, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4918–4923.

[56] R. Riek, J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, *J. Am. Chem. Soc.* **2002**, *124*, 12144–12153.

[57] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13585–13590.

[58] J. P. Loria, M. Rance, A. G. Palmer, 3rd, *J. Magn. Reson.* **1999**, *141*, 180–184.

[59] A. Kumar, R. R. Ernst, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.

[60] J. Jenner, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **1976**, *71*, 4546–4553.

[61] V. A. Higman, *Protein NMR: A Practical Guide*, **2012**.

[62] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, *J. Am. Chem. Soc.* **1999**, *121*, 844–848.

[63] K. Golman, O. Axelsson, H. Johannesson, S. Mansson, C. Olofsson, J. S. Petersson, *Magn. Reson. Med.* **2001**, *46*, 1–5.

[64] B. M. Goodson, *J. Magn. Reson.* **2002**, *155*, 157–216.

[65] A. Abragam, M. Goldman, *Rep. Prog. Phys.* **1978**, *41*, 395–467.

[66] T. R. Carver, C. P. Slichter, *Phys. Rev.* **1956**, *102*, 975.

[67] L. H. Bennett, H. C. Torrey, *Phys. Rev.* **1957**, *108*, 499.

[68] D. J. Singel, J. H. Seidel, R. D. Kendrick, C. S. Yannoni, *J. Magn. Reson.* **1989**, *81*, 145–161.

[69] M. Afeworki, R. A. McKay, J. Schaefer, *Macromolecules* **1992**, *25*, 4084–4091.

[70] J. Kurhanewicz, D. B. Vigneron, K. Brindle, E. Y. Chekmenev, A. Comment, C. H. Cunningham, R. J. Deberardinis, G. G. Green, M. O. Leach, S. S. Rajan, R. R. Rizi, B. D. Ross, W. S. Warren, C. R. Malloy, *Neoplasia* **2011**, *13*, 81–97.

[71] A. Abragam, W. G. Proctor, *Compt. Rend.* **1958**, *246*, 2253–2256.

[72] A. Goldman, *Spin temperature and nuclear magnetic resonance in solids*, Clarendon Press; 1st edition, **1970**.

[73] W. T. Wenckebach, *Nucl. Instrum. Meth. A* **1995**, *256*, 1–4.

[74] A. Bornet, J. Milani, S. Wang, D. Mammoli, R. Buratto, N. Salvi, T. F. Segaw, V. Vitzthum, P. Mieville, S. Chinthalapalli, A. J. Perez-Linde, D. Carnevale, S. Jannin, M. Caporinia, S. Ulzega, M. Rey, G. Bodenhausen, *Chimia* **2012**, *66*, 734–740.

[75] A. Comment, B. Van den Brandt, K. Uffmann, F. Kurdzesau, S. Jannin, J. A. Konter, P. Hautle, W. T. Wenckebach, R. Gruetter, J. J. van der Klink, *Concept. Magn. Reson. B* **2007**, *31B*, 255–269.

[76] B. Vuichoud, J. Milani, A. Bornet, R. Melzi, S. Jannin, G. Bodenhausen, *J. Phys. Chem. B* **2014**, *118*, 1411–1415.

[77] M. Carravetta, O. G. Johannessen, M. H. Levitt, *Phys. Rev. Lett.* **2004**, *92*, 153003.

[78] R. B. van Heeswijk, K. Uffmann, A. Comment, F. Kurdzesau, C. Perazzolo, C. Cudalbu, S. Jannin, J. A. Konter, P. Hautle, B. van den Brandt, G. Navon, J. J. van der Klink, R. Gruetter, *Magn. Reson. Med.* **2009**, *61*, 1489–1493.

[79] R. Sarkar, A. Comment, P. R. Vasos, S. Jannin, R. Gruetter, G. Bodenhausen, H. Hall, D. Kirik, V. P. Denisov, *J. Am. Chem. Soc.* **2009**, *131*, 16014–16015.

[80] P. Mieville, S. Jannin, L. Helm, G. Bodenhausen, *J. Am. Chem. Soc.* **2010**, *132*, 5006–5007.

Chapter 2

NMR Spectroscopy as a Tool to Study Molecular Chaperones and their Complexes with Substrate

2.1 Introduction

A molecular chaperone is a protein necessary for the correct fold and assembly of a polypeptide substrate, but it is not part of the final stable and functional fold[1]. Molecular chaperones are found in all cell compartments where they interact with unfolded polypeptide chains, ensuring their correct folding and transport[2]. They assist a large variety of protein maturation processes in an ATP-dependent or independent manner and can be classified in different functional groups[3]. Chaperones assist translocation of unfolded polypeptide across cell membranes (translocases), transport of substrates to prevent misfolding (holdases), insertion of proteins into a membrane in a functional state (insertases), and support the folding of soluble proteins (foldases). Despite their common functionality, chaperones share little sequence specificity but often act on a broad range of substrates. These substrate interactions happen in a highly dynamic manner due to the inherent disordered nature of the unfolded and partially folded polypeptide substrates. Although X-ray elucidated a number of chaperone structures in their apo form, the flexibility of the substrate usually prevents its use to obtain information on their bound state. NMR spectroscopy is a powerful technique to provide structural information of transient and dynamic protein–protein interactions at atomic resolution[4, 5]. Different NMR techniques have successfully been employed for characterizing structures of chaperones and chaperone-protein complexes. Chemical shift perturbation (CSP) was successfully employed to characterize substrates bound to the chaperone GroEL and GroEL/ES complex[6-8], Hsp90[9, 10] and Hsp70[11], where the interaction surface was mapped for a classical protein–substrate interaction where the chaperone recognizes a certain segment of the substrate and binds it in a single defined conformation. The nuclear Overhauser effect (NOE), providing information on atomic distances, was successfully applied to characterize intra- and inter-molecular contacts in complexes and allowed solving chaperone–substrate structures, among which Trigger Factor[12]. Paramagnetic relaxation enhancement (PRE) was used to measure intermolecular distances in classical chaperone–substrate complexes as Hsp90[9, 13] as well as in a dynamic conformation

ensemble as the Skp-Omp complex[14]. Furthermore, several other NMR techniques such as hydrogen/deuterium exchange[15, 16] or spin relaxation[14] brought essential information on a broad range of chaperone–substrate complexes.

Among the large number of chaperone–substrate complexes in the cell, this work focuses on further investigation of the periplasmic chaperones of *E. coli*, the outer membrane protein (OMP) holdases Skp and SurA, for which the molecular basis and the functional details of their mechanism at the atomic level remains only partly understood. Previous NMR interaction studies of Skp with its OMP substrate revealed that the unfolded polypeptide binds the chaperone in a compact ensemble of rapidly interconverting conformations devoid of regular secondary structure elements, the "fluid globule"[14], providing for the first time structural and dynamic information at the atomic level of the chaperone:substrate interplay. Furthermore, NMR studies of substrates bound to the SurA chaperone indicate that the outer membrane protein A (OmpA) interacts in a similar fluidic conformation state ensemble[14], pointing to a general fluid globule-like state for the OMP within the chaperone. The understanding of this energy-independent chaperone-assisted transport and delivery mechanism at the atomic level is fundamental as they play an essential role in the cell, preventing aggregation of the unfolded OMP during transport from the inner membrane (IM) to the outer membrane (OM) of Gram-negative bacteria. In *E. coli*, OMPs, located in the lipidic OM, display β-barrel structure consisting of antiparallel amphipathic β-strands[17]. They are synthesized in the cytoplasm by the ribosome, as precursors with N-terminal signal sequence, required for translocation across the inner membrane. These precursors are initially bound by the chaperone trigger factor and subsequently handed over to the chaperone secretory-translocase B (SecB), which prevent their aggregation during their transport through the cytoplasm[18]. The N-terminal signal sequence target the OMP precursor to the Sec machinery, embedded in the inner membrane, for its translocation across the membrane into the periplasm where the signal sequence is cleaved off[18, 19]. There, the nascent OMPs are accessible to the periplasmic chaperones Skp (seventeen kilodalton protein), DegP and SurA (survival factor A), ensuring their transport to the outer membrane and their delivery to the β-barrel assembly machinery (Bam complex). The main component of the Bam complex, BamA is composed of a sixteen-stranded β-barrel and five periplasmic polypeptide transport associated (POTRA) domains. The Bam complex ensures the insertion of the OMPs into the OM (Fig. 2.1)[20]. The exact roles of the periplasmic chaperones are not yet clear, as individual knockouts of these proteins do not show lethal phenotypes. Lethality was however observed in cells lacking both Skp and SurA as well as DegP and SurA, but

not DegP and Skp[21]. These data suggest a parallel pathway where SurA may act as the primary periplasmic chaperone and the Skp/DegP path as an alternative rescue for off–pathway OMP polypeptides[22].

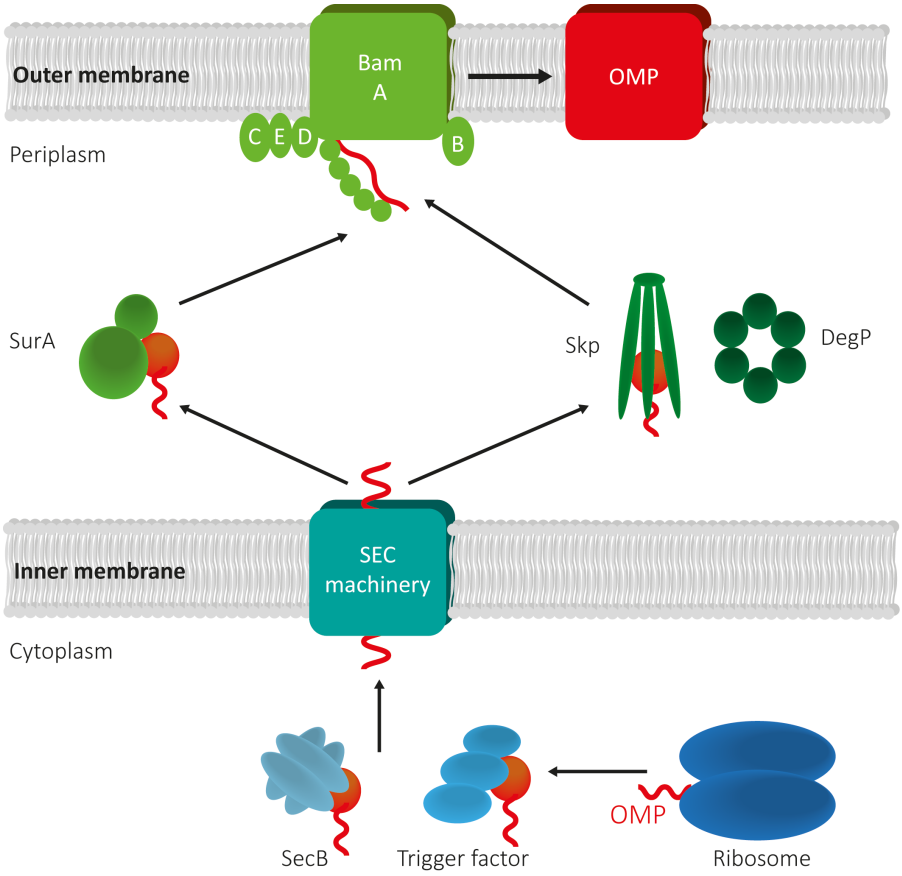


Figure 2.1. Schematic representation of the OMP biogenesis in *E. coli* [23] where the OMP is depicted in red.

In this chapter, the interaction between the Skp chaperone and its OMP substrate characterized at atomic resolution (section 2.2) as well as the investigation of the interaction between SurA and POTRA domains (section 2.3) allowed to take a step forward in the understanding of this essential chaperone mechanism.

References

[1] R. J. Ellis, S. M. van der Vies, *Annu. Rev. Biochem.* **1991**, *60*, 321–347.

[2] S. M. Doyle, O. Genest, S. Wickner, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 617–629.

[3] B. M. Burmann, S. Hiller, *Prog. NMR Spect.* **2015**, *86-87C*, 41–64.

[4] A. M. Bonvin, R. Boelens, R. Kaptein, *Curr. Opin. Chem. Biol.* **2005**, *9*, 501–508.

[5] J. Y. Suh, M. Cai, D. C. Williams, Jr., G. M. Clore, *J. Biol. Chem.* **2006**, *281*, 8939–8949.

[6] J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, *Nature* **2002**, *418*, 207–211.

[7] R. Horst, E. B. Bertelsen, J. Fiaux, G. Wider, A. L. Horwich, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12748–12753.

[8] M. Yagi-Utsumi, T. Kunihara, T. Nakamura, Y. Uekusa, K. Makabe, K. Kuwajima, K. Kato, *FEBS Lett.* **2013**, *587*, 1605–1609.

[9] F. Hagn, S. Lagleder, M. Retzlaff, J. Rohrberg, O. Demmer, K. Richter, J. Buchner, H. Kessler, *Nat. Struct. Mol. Biol.* **2011**, *18*, 1086–1093.

[10] G. E. Karagöz, A. M. Duarte, E. Akoury, H. Ippel, J. Biernat, T. Moran Luengo, M. Radli, T. Didenko, B. A. Nordhues, D. B. Veprintsev, C. A. Dickey, E. Mandelkow, M. Zweckstetter, R. Boelens, T. Madl, S. G. Rüdiger, *Cell* **2014**, *156*, 963–974.

[11] S. Y. Stevens, S. Cai, M. Pellecchia, E. R. Zuiderweg, *Protein Sci.* **2003**, *12*, 2588–2596.

[12] T. Saio, X. Guan, P. Rossi, A. Economou, C. G. Kalodimos, *Science* **2014**, *344*, 1250494.

[13] O. R. Lorenz, L. Freiburger, D. A. Rutz, M. Krause, B. K. Zierer, S. Alvira, J. Cuellar, J. M. Valpuesta, T. Madl, M. Sattler, J. Buchner, *Mol. Cell* **2014**, *53*, 941–953.

[14] B. M. Burmann, C. Wang, S. Hiller, *Nat. Struct. Mol. Biol.* **2013**, *20*, 1265–1272.

[15] S. E. Nieba-Axmann, M. Ottiger, K. Wüthrich, A. Plückthun, *J. Mol. Biol.* **1997**, *271*, 803–818.

[16] S. J. Park, B. N. Borin, M. A. Martinez-Yamout, H. J. Dyson, *Nat. Struct. Mol. Biol.* **2011**, *18*, 537–542.

[17] C. Goemans, K. Denoncin, J. F. Collet, *Biochim. Biophys. Acta* **2013**, *1843*, 1517–1528.

[18] J. de Keyzer, C. van der Does, A. J. Driessen, *Cell. Mol. Life. Sci.* **2003**, *60*, 2034–2052.

[19] E. Papanikou, S. Karamanou, A. Economou, *Nat. Rev. Microbiol.* **2007**, *5*, 839–851.

[20] C. L. Hagan, T. J. Silhavy, D. Kahne, *Annu. Rev. Biochem.* **2011**, *80*, 189–210.

[21] A. E. Rizzitello, J. R. Harper, T. J. Silhavy, *J. Bacteriol.* **2001**, *183*, 6794–6800.

[22] J. G. Sklar, T. Wu, D. Kahne, T. J. Silhavy, *Genes Dev.* **2007**, *21*, 2473–2484.

[23] B. M. Burmann, S. Hiller, *Chimia* **2012**, *66*, 759–763.

2.2 Structural Mapping of a Chaperone–Substrate Interaction Surface

Original publication

Morgane Callon, Björn M. Burmann, and Sebastian Hiller

Structural Mapping of a Chaperone–Substrate Interaction Surface

Angew. Chem. Int. Ed. **2014**, *53*, 5096–5072

Protein NMR Spectroscopy

DOI: 10.1002/anie.201310963

Structural Mapping of a Chaperone–Substrate Interaction Surface**

*Morgane Callon, Björn M. Burmann, and Sebastian Hiller**

Abstract: *NMR spectroscopy is used to detect site-specific intermolecular short-range contacts in a membrane–protein–chaperone complex. This is achieved by an "orthogonal" isotope-labeling scheme that permits the unambiguous detection of intermolecular NOEs between the well-folded chaperone and the unfolded substrate ensemble. The residues involved in these contacts are part of the chaperone–substrate contact interface. The approach is demonstrated for the 70 kDa bacterial Skp-tOmpA complex.*

Molecular chaperones play an essential role in the biogenesis of proteins in all kingdoms of life.[1,2] As a general mode of function, they interact with unfolded protein substrates, prohibiting their progression on non-native aggregation pathways.[3] Still, the molecular basis of chaperone–substrate interactions is not well understood at the atomic level.

NMR spectroscopy is a powerful technique to provide structural information of transient and dynamic protein–protein interactions at atomic resolution.[4,5] It has been successfully employed for the direct observation of substrates bound to the chaperones GroEL,[6–10] Hsp90,[11–13] Hsp70,[14] and Skp.[15,16] Furthermore, long-range spatial correlations between substrates and chaperones have been detected for Skp,[16] Hsp90,[11] and GroEL[10] by the use of paramagnetic relaxation enhancement (PRE). However, for a full description of the molecular interactions, the identification of direct contact interfaces in the unperturbed chaperone–substrate complex is necessary. Thereby, chemical shift perturbations do not necessarily yield the true interaction surface, because they are ensemble averages of the dynamic systems and can contain significant contributions of allosteric effects. Additional experimental approaches are thus required.

Herein, we use the nuclear Overhauser effect (NOE) to detect site-specific intermolecular short-range contacts in a membrane-protein–chaperone complex. We measure these contacts in the 70 kDa complex of the chaperone Skp with its bound substrate tOmpA. Skp is a homotrimeric, 51 kDa chaperone from the Gram-negative bacterium *Escherichia coli (E. coli)* that prevents outer membrane proteins from aggregation during their transport across the periplasm.[17,18] Its apo-form features three long α-helical "arms" forming a central cavity.[19] The 19 kDa transmembrane domain of the outer membrane protein A, tOmpA, is a natural substrate of Skp.[20] Previous solution NMR studies have shown that tOmpA and the alternative substrate OmpX, when bound to Skp, adopt the conformational state of a "fluid globule", a compacted ensemble of rapidly interconverting conformations devoid of regular secondary structure elements, but with non-random backbone dynamics.[16]

The interpretation of intermolecular cross-relaxation events between the folded Skp and the unfolded tOmpA has to consider that the ^{1}H–^{1}H cross-relaxation rate constant σ^{jk} between a proton j on the chaperone and a proton k on the substrate is an ensemble average (Supporting Information, Figure 1), similar to intramolecular NOEs in unfolded protein ensembles.[21–24] The amplitude of an NOE cross-peak between spin j and k thus contains contributions from different conformations of the structural ensemble, and importantly, owing to the r^{-6} dependence of σ^{jk}, it is biased by conformations with short interspin distances r. Furthermore, as the dynamic ensemble features a narrow chemical shift dispersion, the observed cross-peaks in 3D NOESY spectra are typically a superimposition of multiple substrate spins with similar chemical shifts. These features generally prohibit a quantitative conversion of NOESY cross-peak amplitudes into interspin distances.

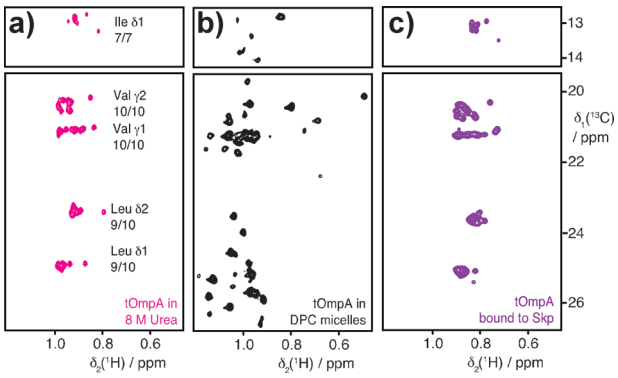


Figure 1. Random-coil-like averaging of tOmpA side chain conformations in the Skp-bound state. a) 2D [^{13}C,^{1}H]-HSQC spectrum of 200 μM [*U*-^{2}H,^{15}N, Ile-δ1-^{13}CH$_3$, Leu,Val-^{13}CH$_3$]-tOmpA in NMR buffer with 8 M urea at 37 °C. The number of detected and expected resonances for each methyl moiety type are indicated. b) 2D [^{13}C,^{1}H]-HMQC spectrum of 550 μM [*U*-^{2}H,^{15}N, Ile-δ1-^{13}CH$_3$, Leu,Val-^{13}CH$_3$]-tOmpA in NMR buffer with 150 mM DPC at 37 °C. c) 2D [^{13}C,^{1}H]-HMQC spectrum of 560 μM [*U*-^{2}H,^{15}N, Ile-δ1-^{13}CH$_3$, Leu,Val-^{13}CH$_3$]-tOmpA bound to [*U*-^{2}H]-Skp in NMR buffer at 37 °C.

However, measurements of the NOE can be used to identify those residues of the chaperone that feature a significantly populated amount of conformations in close spatial

[*] M. Callon, Dr. B. M. Burmann, Prof. Dr. S. Hiller
Biozentrum, University of Basel
Klingelbergstrasse 70, 4056 Basel (Switzerland)
E-mail: sebastian.hiller@unibas.ch

[**] This work was supported by grants from the Swiss National Science Foundation (Grant PP00P3_128419) and the European Research Council (FP7 contract MOMP 281764) to S.H., and by a personal fellowship from the Werner-Siemens Foundation to M.C.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201310963.

contact with spin probes in the substrate. These residues are part of the chaperone–substrate contact surface. For mapping of these residues, we use an "orthogonal" side-chain labeling scheme, that is, a combination of amino acid side-chain moieties with non-overlapping ^{1}H chemical shift dispersion. These are the methyl groups of isoleucine, leucine, and valine for the unfolded substrate (ILV-tOmpA) and the methyl groups of alanine residues for the chaperone (A-Skp), which were introduced in $^{13}CH_3$-labeled form on a deuterated background using established techniques.[25,26] Formation of the complex was accomplished by rapid dilution of purified, denatured tOmpA into separately produced Skp solution and subsequent buffer exchange. The integrity of the complex, its homogeneous 3:1 Skp:tOmpA stoichiometry, and the folding competence of tOmpA from the complex was assessed as described previously.[16]

A set of 2D [^{13}C,^{1}H]-correlation spectra of ILV-labeled tOmpA in three different preparational states serves as an initial characterization of the substrate side chains (Figure 1). In 8 M urea aqueous solution, tOmpA adopts a fast exchanging random-coil ensemble and accordingly, the ILV side chain moieties feature narrow ^{1}H and ^{13}C chemical shift dispersions of 0.22 ppm and 0.2–0.6 ppm, respectively (Figure 1 a). In the folded state of tOmpA in DPC micelles, the protein populates an 8-stranded β-barrel structure as the single, stable conformation. Therein, the ILV side chains are involved in specific contacts, resulting in ^{1}H and ^{13}C chemical dispersions about four times larger than in denaturant (Figure 1 b). Bound to the chaperone Skp, the methyl moieties of ILV-tOmpA feature again narrow chemical shift dispersions, closely resembling the random-coil spectrum and thus indicating that the tOmpA side-chains in Skp lack a well-defined structure (Figure 1 c). The observation of a single set of resonances confirms the presence of a conformational ensemble in the fast exchange regime, that is, with individual lifetimes below 1 ms, validating the use of an ensemble model as the basis of the data interpretation.

The alanine CH_3 resonances in Skp feature a large dispersion, both in the apo and the holo form, in accordance with the adoption of a well-defined secondary and tertiary structure (Figure 2 a). Intermolecular cross-peak amplitudes were measured by 3D ^{13}C-resolved [^{1}H,^{1}H]-NOESY experiments. Importantly, the orthogonal side-chain labeling pattern ensures the absence of spectral overlap between the intermolecular cross-peaks and the strong diagonal peaks, as well as between inter- and intramolecular cross-peaks (Figure 2 b). The observation of a cross-peak in the isoleucine, leucine, and valine ^{1}H chemical shift range 0.55–0.9 ppm in the holo- relative to the apo-spectrum can thus be directly assigned to intermolecular NOE contacts between a specific site on Skp and the ILV methyl spins in the tOmpA ensemble (Figure 2 c,d). The integration of alanine NOE cross-peaks shows non-zero cross-peak amplitudes for residues A34, A75, A78, and A86, all of which have their side-chains pointing into the central part the Skp cavity (Supporting Information, Table S1). Alanine residues located at the trimer interface, as well as at the bottom of Skp, did not show a detectable NOE to the spin probes in the substrate ensemble. The contact surface mapping thus shows that tOmpA is bound in the

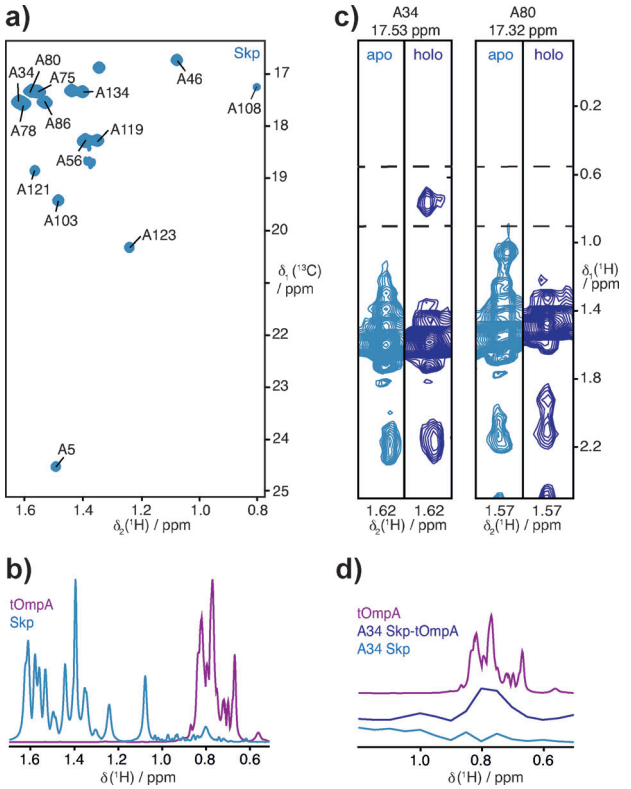


Figure 2. Measurement of intermolecular methyl–methyl NOEs between Skp and tOmpA. a) 2D [^{13}C,^{1}H]-HMQC spectrum of 700 μM [*U*-^{2}H,^{15}N, Met-$^{13}CH_3$, Ala-$^{13}CH_3$]-Skp (A-Skp; light blue) in NMR buffer at 37 °C. The sequence-specific resonance assignments are indicated. b) Overlay of ^{1}H projections of a 2D [^{13}C,^{1}H]-HMQC spectrum of 560 μM [*U*-^{2}H,^{15}N, Ile-δ1-$^{13}CH_3$, Leu,Val-$^{13}CH_3$]-tOmpA (ILV-tOmpA) bound to [*U*-^{2}H]-Skp in NMR buffer at 37 °C (purple) and of a 2D [^{13}C,^{1}H]-HMQC spectra of 700 μM A-Skp in NMR buffer at 37 °C (blue). c) 2D strips from 3D ^{13}C-edited-[^{13}C,^{1}H]-NOESY spectra taken at the positions of Ala 34 and Ala 80 of 700 μM A-Skp in its apo form (light blue) and 250 μM A-Skp with bound ILV-tOmpA (holo form, dark blue) in NMR buffer at 37 °C. Spectra were recorded with a NOESY mixing time of 200 ms. The dashed horizontal lines indicate the ^{1}H chemical shift dispersion of ILV residues in tOmpA bound to Skp (0.55–0.9 ppm). d) Overlay of a ^{1}H projection of a 2D [^{13}C,^{1}H]-HMQC spectrum of 560 μM ILV-tOmpA bound to Skp (purple) and 1D cross sections from the 3D ^{13}C-edited-[^{13}C,^{1}H]-NOESY spectra shown in (c), taken at the position of Ala 34.

central part of the Skp cavity (Figure 3). Notably, these intermolecular cross-peaks have the same sign as the diagonal peaks, indicating an overall negative NOE for the ensemble average and thus the existence of the slow motion limit for the chaperone–substrate contacts.[27,28] The local lifetimes of segments of the substrate bound to the chaperone thus have a lower limit of $\tau_i \gg \omega_0^{-1} \approx 1$ ns. A comparison to cross-peaks with known intramolecular distances within Skp in the same spectrum shows that the ensemble-averaged intensities correspond to fully populated distances in the range 5–6 Å (Supporting Information, Table S2). Since the intermolecular NOE peaks are a superposition of individual signals from up to 45 spin probes in tOmpA, we estimate that, on average, conformations with a given Skp–tOmpA ^{1}H–^{1}H spin pair in close contact are populated by about 0.5–2 %.

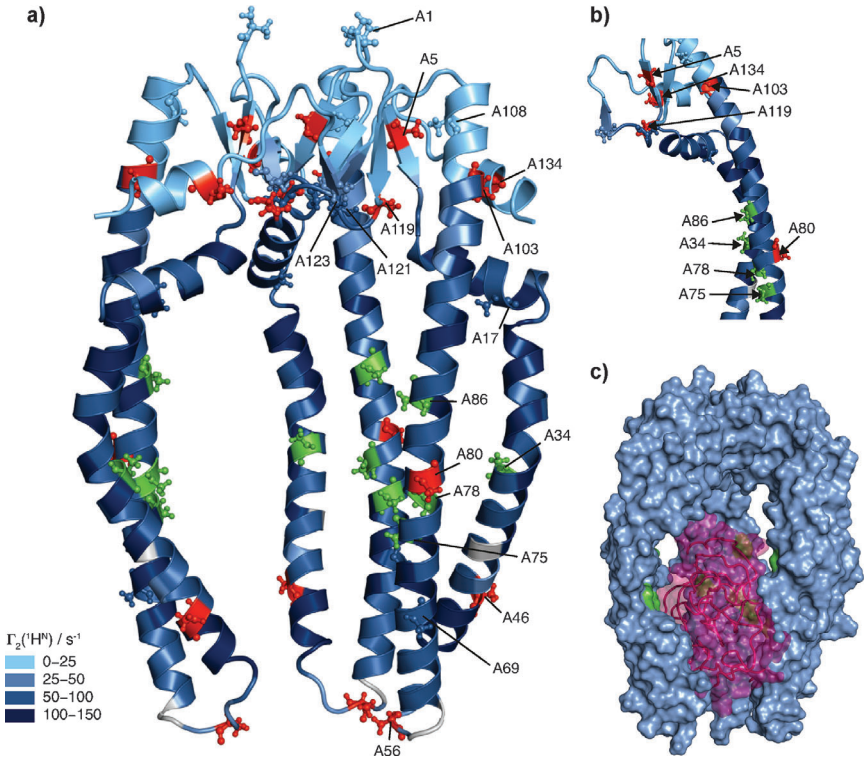


Figure 3. The contact interface of tOmpA and Skp. a) Ribbon representation of the Skp crystal structure with the alanine residues shown as ball and sticks (PDB 1SG2).[19] Alanines shown in green and red correspond to residues with and without a significant intermolecular cross-peak in the 3D ^{13}C-edited-[^{13}C,^{1}H]-NOESY spectrum of Skp bound to tOmpA, respectively (Supporting Information, Table S1). Intermolecular PREs on Skp from a paramagnetic spin label attached to OmpX are indicated by a blue gradient for Γ_2-values between 0 and 150 s^{-1}.[16] b) Zoom on one Skp monomer. c) Structural model of a single conformation from the dynamic Skp-tOmpA ensemble. Skp is shown in surface representation (blue). A randomly generated tOmpA polypeptide conformation is shown in purple.[16] Alanine residues with significant NOE cross-peak amplitude, as in (a), are highlighted in green.

Finally, we compare the result from the NOE-based mapping with previous measurements of intermolecular PREs from a paramagnetic spin label attached to OmpX and detected on the individual amide moieties of Skp.[16] Overall, the positions of the Omp substrate in the center of the Skp cavity determined by the two different spin-interaction methods PRE and NOE are in excellent agreement (Figure 3a). Importantly, however, the NOE mapping distinguishes residues A46 and A80, which are located in the Skp arms at the mid-height of the cavity, but with their side-chains pointing outside, from residues A34 and A78, which have both close contacts to tOmpA. The PRE, in turn, is similar for all four residues with Γ_2-values in the range 70–130 s^{-1}. The NOE-based contact interface mapping thus shows that the helical arms of Skp do not twist from their canonical position in the crystal structure and that they have a distinct substrate interaction surface on their inside.

In summary, measurements of intermolecular NOE in combination with a suitable orthogonal methyl group labeling strategy enabled the structural mapping of the Skp–tOmpA contact interface at the atomic level. Structurally meaningful intermolecular NOEs in the slow motion limit were detected in a folded–unfolded protein–protein complex. In particular, and compared to the long-ranged PREs, the short-ranged NOEs allow the distinction of residues according to their side-chain orientation. The approach can be tailored to a system of interest by suitable orthogonal combinations of methyl-group side chains. For example, methionine CH_3 moieties with proton signals from 1.95 to 2.25 ppm could be readily integrated in the present setup. Since methyl groups on a deuterated background are highly sensitive and since sequence-specific resonance assignments are known for molecular sizes larger than 150 kDa,[29,30] we estimate that with 70 kDa, the size limitation of the approach has not been reached in the present study. The approach thus opens new perspectives for the investigation of molecular interactions in protein biogenesis.

Experimental Section

Details on protein biochemistry, isotope labeling and NMR spectroscopy are given in the Supporting Information. Sequence-specific side-chain resonance assignments of apo- and holo-Skp have been deposited to the BMRB data base with accession codes 19733 and 19730, respectively.

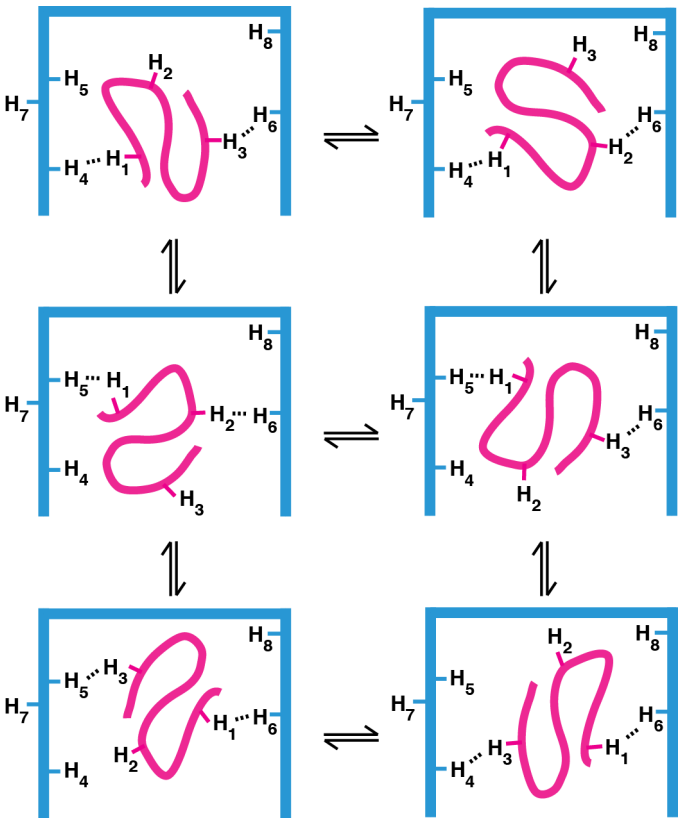
Received: December 18, 2013
Revised: February 17, 2014
Published online: April 2, 2014

Keywords: chaperone proteins · membrane proteins · NMR spectroscopy · protein ensembles · protein–protein interactions

[1] R. J. Ellis, S. M. van der Vies, *Annu. Rev. Biochem.* **1991**, *60*, 321–347.
[2] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. U. Hartl, *Annu. Rev. Biochem.* **2013**, *82*, 323–355.
[3] S. M. Doyle, O. Genest, S. Wickner, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 617–629.
[4] A. M. Bonvin, R. Boelens, R. Kaptein, *Curr. Opin. Chem. Biol.* **2005**, *9*, 501–508.
[5] J. Y. Suh, M. Cai, D. C. Williams, Jr., G. M. Clore, *J. Biol. Chem.* **2006**, *281*, 8939–8949.
[6] R. Horst, E. B. Bertelsen, J. Fiaux, G. Wider, A. L. Horwich, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12748–12753.
[7] E. Koculi, R. Horst, A. L. Horwich, K. Wüthrich, *Protein Sci.* **2011**, *20*, 1380–1386.
[8] Z. Wang, H. Feng, S. J. Landry, J. Maxwell, L. M. Gierasch, *Biochemistry* **1999**, *38*, 12537–12546.

[9] M. Yagi-Utsumi, T. Kunihara, T. Nakamura, Y. Uekusa, K. Makabe, K. Kuwajima, K. Kato, *FEBS Lett.* **2013**, *587*, 1605 – 1609.
[10] D. S. Libich, N. L. Fawzi, J. Ying, G. M. Clore, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11361 – 11366.
[11] F. Hagn, S. Lagleder, M. Retzlaff, J. Rohrberg, O. Demmer, K. Richter, J. Buchner, H. Kessler, *Nat. Struct. Mol. Biol.* **2011**, *18*, 1086 – 1093.
[12] S. J. Park, B. N. Borin, M. A. Martinez-Yamout, H. J. Dyson, *Nat. Struct. Mol. Biol.* **2011**, *18*, 537 – 542.
[13] S. Rüdiger, S. M. V. Freund, D. B. Veprintsev, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11085 – 11090.
[14] S. Y. Stevens, S. Cai, M. Pellecchia, E. R. Zuiderweg, *Protein Sci.* **2003**, *12*, 2588 – 2596.
[15] T. A. Walton, C. M. Sandoval, C. A. Fowler, A. Pardi, M. C. Sousa, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1772 – 1777.
[16] B. M. Burmann, C. Wang, S. Hiller, *Nat. Struct. Mol. Biol.* **2013**, *20*, 1265 – 1272.
[17] N. Ruiz, D. Kahne, T. J. Silhavy, *Nat. Rev. Microbiol.* **2006**, *4*, 57 – 66.
[18] J. G. Sklar, T. Wu, D. Kahne, T. J. Silhavy, *Genes Dev.* **2007**, *21*, 2473 – 2484.
[19] I. P. Korndörfer, M. K. Dommel, A. Skerra, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1015 – 1020.
[20] S. Jarchow, C. Lück, A. Görg, A. Skerra, *Proteomics* **2008**, *8*, 4987 – 4994.
[21] K. Wüthrich, *Curr. Opin. Struct. Biol.* **1994**, *4*, 93 – 99.
[22] K. M. Fiebig, H. Schwalbe, M. Buck, L. J. Smith, C. M. Dobson, *J. Phys. Chem.* **1996**, *100*, 2661 – 2666.
[23] E. Barbar, *Biopolymers* **1999**, *51*, 191 – 207.
[24] K. A. Crowhurst, J. D. Forman-Kay, *Biochemistry* **2003**, *42*, 8687 – 8695.
[25] N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis, L. E. Kay, *J. Biomol. NMR* **1999**, *13*, 369 – 374.
[26] S. R. Tzeng, M. T. Pai, C. G. Kalodimos, *Methods Mol. Biol.* **2012**, *831*, 133 – 140.
[27] R. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon, Oxford, **1987**.
[28] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[29] I. Gelis, A. M. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou, C. G. Kalodimos, *Cell* **2007**, *131*, 756 – 769.
[30] R. Rosenzweig, S. Moradi, A. Zarrine-Afsar, J. R. Glover, L. E. Kay, *Science* **2013**, *339*, 1080 – 1083.

Supplementary Figure



Supplementary Figure 1. Conformational averaging of intermolecular NOEs in a chaperone–substrate complex. Schematic model of six conformations of a chaperone protein (blue) in complex with its unfolded substrate (purple) in fast chemical equilibrium. Eight protons H_1–H_8 are highlighted. For each substrate conformation, pairs of protons involved in short-range contacts, corresponding to strong intermolecular NOEs, are connected with dashed lines.

Supplementary Tables

Supplementary Table 1. Integrals of intermolecular NOE cross-peak volumes between alanine methyl groups of Skp and ILV methyl groups in tOmpA

Skp Residue	I_{apo} (AU)	I_{holo} (AU)	I_{holo}–I_{apo} (AU)	$(I_{holo}–I_{apo})/I_{apo}$
A1	no assignment			
A5	46'100	45'000	-1100	-0.02
A17	no assignment			
A34	17'000	43'900	26'900	1.58
A46	41'300	53'000	11'700	0.28
A56	12'500	2'000	-10'500	-0.84
A69	no assignment			
A75	8'000	39'400	31'400	3.93
A78	2'900	21'000	18'100	6.24
A80	14'200	4'300	-9'900	-0.70
A86	1'400	26'300	24'900	17.79
A103	44'000	37'000	-7'000	-0.16
A108	*	*	*	*
A119	21'700	19'000	-2'700	-0.12
A121	4'300	no assignment		
A123	30'000	52'400	22'400	0.75
A134	17'500	12'100	-5'400	-0.31

*: δ_H (A108) = 0.793 ppm. This resonance overlaps with the ILV methyl groups of tOmpA. Intermolecular cross-peaks between A108 and ILV-tOmpA are thus masked by the strong diagonal cross-peak of A108.

Supplementary Table 2. Integrals of intramolecular NOE cross-peak volumes between protons with known distance in Skp.

Proton pair	I_{holo}–I_{apo} (AU)	Distance in crystal structure (Å)
H^N(K98) – H^β(A96)	56’200	5.2

Supplementary Methods

Protein biochemistry: Skp and tOmpA were expressed and purified as described.[1] In brief, Skp lacking its signal sequence and with cleavable N-terminal His_6-tag was expressed in *E.coli* BL21 (λ DE3) cells and purified by Ni^{2+}-affinity twice, in aqueous solution and in a 6M Gdm/HCl solution. The transmembrane domain of OmpA (residues 1–177; tOmpA) lacking its signal sequence was expressed in *E.coli* BL21 (λ DE3) cells. The inclusion body pellet was washed twice, solubilized in 6 M Gdm/HCl and rebuffered into 8 M urea solution, applied to a HiTrap Q HP column and rebuffered into 6 M Gdm/HCl solution. The Skp–tOmpA complex was assembled by adding an excess of denatured tOmpA to Skp in NMR buffer (25 mM MES, 150 mM NaCl, pH 6.5) in a dropwise fashion under continuous stirring at room temperature. After centrifugation for 20 min at 4000xg, the supernatant was concentrated by ultrafiltration.

Isotope labeling: [*U*-^{2}H,^{15}N, Ile-δ1-$^{13}CH_3$, Leu,Val-$^{13}CH_3$]-tOmpA (ILV-tOmpA) was obtained by growing *E. coli* in D_2O-based M9-minimal medium containing ($^{15}NH_4$)Cl and [*U*-^{2}H,^{12}C]-glucose as the sole nitrogen and carbon sources, respectively. 100 mg of 3-[^{2}H], [$^{13}CH_3$]-ketoisovalerate and 75 mg of [^{2}H], 3,3-[$^{13}CH_3$]-ketobutyrate were added to the medium 1 h prior to induction.[2] [*U*-^{2}H,^{15}N, Met-$^{13}CH_3$, Ala-$^{13}CH_3$]-Skp (MA-Skp) was obtained by growing *E. coli* in M9 minimal medium containing ($^{15}NH_4$)Cl, [*U*-^{2}H,^{12}C]-glucose, 2% of bioexpress rich medium (Cambridge Isotope Laboratories) and D_2O. 100 mg of 2-[^{2}H], 3-[^{13}C] L-alanine and 250 mg of [$^{13}CH_3$] methionine were added to the medium 1 h prior to induction.[3] Besides bioexpress, all isotopes were purchased from Sigma Aldrich.

NMR spectroscopy: NMR spectra were recorded at 37°C for the Skp apo and holo forms on a Bruker AscendII-700 spectrometer equipped with a cryogenic triple-resonance probe. The proton chemical shifts were referenced to internal DSS[4] and those for carbon-13 were indirectly referenced. The 2D [^{13}C,^{1}H]-HSQC of urea-denatured ILV-tOmpA was recorded in a total experiment time of 1 h 30 min. The ^{1}H carrier was centered on the water resonance, the ^{13}C carrier at 16.5 ppm. The interscan delay was set to 1 s. In the direct dimension, 2048 complex points were recorded in an acquisition time of 182 ms, multiplied with a 75°-shifted sine bell, zero-filled to 4096 points and Fourier transformed. In the indirect dimension, 1024 complex points were measured with a maximal evolution time of 363 ms, multiplied with a 75°-shifted sine bell, zero-filled to 2048 points and Fourier transformed. The 2D [^{13}C,^{1}H]-HMQC were recorded in a total experiment time of 13 h for ILV-tOmpA in detergent micelles, in 20 h for ILV-tOmpA in Skp and in 6 h 30 min for Skp in its apo form. The ^{1}H carrier was

centered on the water resonance, the ^{13}C carrier at 20 ppm. The interscan delay was set to 1 s. In the direct dimension, 2048 complex points were recorded in an acquisition time of 182 ms, multiplied with a 75°-shifted sine bell, zero-filled to 4096 points and Fourier transformed. In the indirect dimension, 300 complex points were measured with a maximal evolution time of 49 ms, multiplied with a 75°-shifted sine bell, zero-filled to 1024 points and Fourier transformed. The 3D [^{13}C,^{1}H]-NOESY were recorded in a total experiment time of 102 h for Skp in its apo form and in 87 h for Skp-tOmpA. The ^{1}H carrier was centered on the water resonance, the ^{13}C carrier at 17.5 ppm. The interscan delay was set to 0.97 s. In the direct dimension, 1024 complex points were recorded in an acquisition time of 91 ms, multiplied with a 75°-shifted sine bell, zero-filled to 2048 points and Fourier transformed. In the carbon indirect dimension, 130 complex points were measured with a maximal evolution time of 43 ms for Skp in its apo form (122 points and 36 ms for Skp-tOmpA), multiplied with a 75°-shifted sine bell, zero-filled to 256 points and Fourier transformed. In the proton indirect dimension, 136 complex points were measured with a maximal evolution time of 16 ms for Skp in its apo form (124 points and 14 ms for Skp-tOmpA), multiplied with a 75°-shifted sine bell, zero-filled to 512 points for Skp in its apo form (256 points for Skp-tOmpA) and Fourier transformed. For all spectra polynomial baseline correction was applied in all dimensions.

Data analysis: NMR data were processed using PROSA[5] and analyzed with CARA (cara.nmr.ch) and XEASY.[6] Alanine residues of Skp were assigned sequence-specifically from previously established $^{13}C_{\beta}$ assignments,[1] in combination with an analysis of 3D ^{15}N-resolved-[^{1}H,^{1}H]-NOESY and 3D ^{13}C-resolved-[^{1}H,^{1}H]-NOESY spectra. NOESY spectra were integrated using XEASY.[6] Spectral intensities in 3D ^{13}C-resolved-[^{1}H,^{1}H]-NOESY spectra of apo and holo Skp were measured by integration at the ^{1}H chemical shift position of ILV residues (0.55–0.9 ppm), using a cuboid integration volume in three dimensions, and resulting in the integrals I_{apo} and I_{holo}, respectively. Intermolecular NOE peaks with significant signal-to-noise ratio are detected, if $(I_{holo}-I_{apo})/I_{apo} > 1$.

Supplementary References

[1] B. M. Burmann, C. Wang, S. Hiller, *Nat. Struct. Mol. Biol.* **2013**, *20*, 1265–1272.

[2] N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis, L. E. Kay, *J. Biomol. NMR* **1999**, *13*, 369–374.

[3] S. R. Tzeng, M. T. Pai, C. G. Kalodimos, *Methods Mol. Biol.* **2012**, *831*, 133–140.

[4] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *J. Mol. Biol.* **1998**, *280*, 933–952.

[5] P. Güntert, V. Dötsch, G. Wider, K. Wüthrich, *J. Biomol. NMR* **1992**, *2*, 619–629.

[6] C. Bartels, T. H. Xia, M. Billeter, P. Güntert, K. Wüthrich, *J. Biomol. NMR* **1995**, *6*, 1–10.

2.3 Atomic Details of the SurA–POTRA Interaction: Insights into the Delivery Mechanism of unfolded Outer Membrane Protein to the Bam Complex

Abstract

In *E. coli*, the periplasmic chaperone SurA interacts with the soluble BamA-POTRA domains. This interaction plays a role in delivering the unfolded OMP to the Bam complex for its insertion in the OM, as part of OMP biogenesis. So far, no structural details are known about this interaction. Here, we used NMR spectroscopy to determine the binding site between SurA and POTRA *in vitro*. Different constructs were tested to define the minimal constructs exhibiting the SurA–POTRA interaction for further structural characterization. Using a chemical shift perturbation approach, the binding site was mapped on the first domain of POTRA. On SurA, partial sequence-specific resonance assignments of the protein were established to map the interaction site. The assignments indicate an interaction with POTRA on the anti-parallel β-sheet formed by β-strands located at the beginning and at the end of the N- and C-terminal domains, respectively. The possible role of arginine 64, located within the α2 helix of POTRA1, was investigated revealing a potential OMP delivery mechanism, by inducing the formation of a possible OMP–specific receiving surface on POTRA1. Based on our findings, a mechanism of SurA interaction with POTRA inducing the OMP substrate delivery is suggested, where the destabilization induced on the SurA chaperone and hydrophobic interactions on POTRA are the critical features of the process.

Introduction

The *E. coli* SurA gene was first discovered in a screen designed to identify genes required for survival of *E. coli* during stationary phase[1-3]. Subsequently, it was shown to be located in the periplasm and assist the folding of OMP polypeptides[4]. SurA is a 47 kDa protein belonging to the class of transport chaperones composed of four domains: the N-terminal domain (N), the peptidyl-prolyl isomerase domain 1 (PPIase1) and the C-terminal domain (C) form a core module while the peptidyl-prolyl isomerase domain 2 (PPIase2) is a satellite domain linked to the core of the protein by a long flexible linker (Fig. 2.2a)[5]. The chaperone activity of SurA is carried out by the N- and C-terminal domains alone, the two PPIase domains being dispensable for SurA function[6]. The function of these two PPIase domains is still unclear, as only PPIase2 domain was shown to exhibit peptidyl-prolyl isomerase (PPIase) activity. It was suggested that the first domain can be a selectivity filter for unfolded OMPs[7], while the second domain could act as a trap to close the substrate inside the core domain, be a second peptide-binding site or exhibit PPIase activity[5]. The N- and C-terminal domains bind OMP polypeptides carrying preferentially an aromatic-polar-aromatic-nonpolar-proline (Ar-Π-Ar-nonΠ-P) motif. The Ar-X-Ar (where X can be any residue) motif is found with high frequency in OMP sequences suggesting that SurA selectively recognize these specific substrates[8]. This selection could also involve the side chain orientation along two faces of the peptide, consistent with the β-strand secondary structure of the OMP[9]. Atomic details on the interaction of these selected substrates with SurA is not known but the crystal structure suggests an interaction of the OMP in a crevice formed by the N- and C-terminal domains[5, 7].

The delivery of the OMPs to the Bam complex, for their folding in the OM is happening through the polypeptide transport associated (POTRA) domains[10]. POTRA domains are N-terminal periplasmic soluble domains of BamA, followed by the β-barrel protein at the C-terminus, assisting the refolding of the unfolded OMP inside the OM. They are present in all Omp85 family proteins essential for biogenesis and assist polypeptide transport[11]. The crystal structure of the soluble BamA-POTRAs (Fig. 2.2b)[12] shows a fishhook shape with successive individual POTRAs rotated in a right handed direction, each domain sharing low sequence similarities but a similar fold of 2 helices and 3 β-strands ordered as β–α–α–β–β. In 2010, Topalova *et al.* showed that POTRA domains could also adopt an extended conformation by SAXS, that may bridge the IM and the OM across the periplasm, suggesting an additional transport pathway for the unfolded OMP[13]. NMR studies of POTRA-OMP interaction indicate a weak binding in fast exchange on the NMR timescale (> 1ms) located on the first β-strand of POTRA1 and the second β-strand of

POTRA2[14]. Based on this observation it was proposed that the transport of the OMP to the barrel is achieved by β-augmentation where the unfolded peptide will use the β-strands of POTRA as a folding template[13]. To date, no structural data could confirm or infirm this hypothesis. Furthermore, no functional details are known on the interaction between SurA and POTRA besides mutagenesis experiments, showing that the interaction is taking place in the vicinity of Arginine 64 (R64) located on the second helix of POTRA1 and that an alteration of this helix leads to an Omp biogenesis defect as well as a reduced interaction with SurA[15, 16]. Here, we used NMR spectroscopy, combined with complementary biophysical tool to characterize the interaction between POTRA and SurA at atomic resolution, indicating a possible OMP delivery and transport mechanism along POTRA domains.

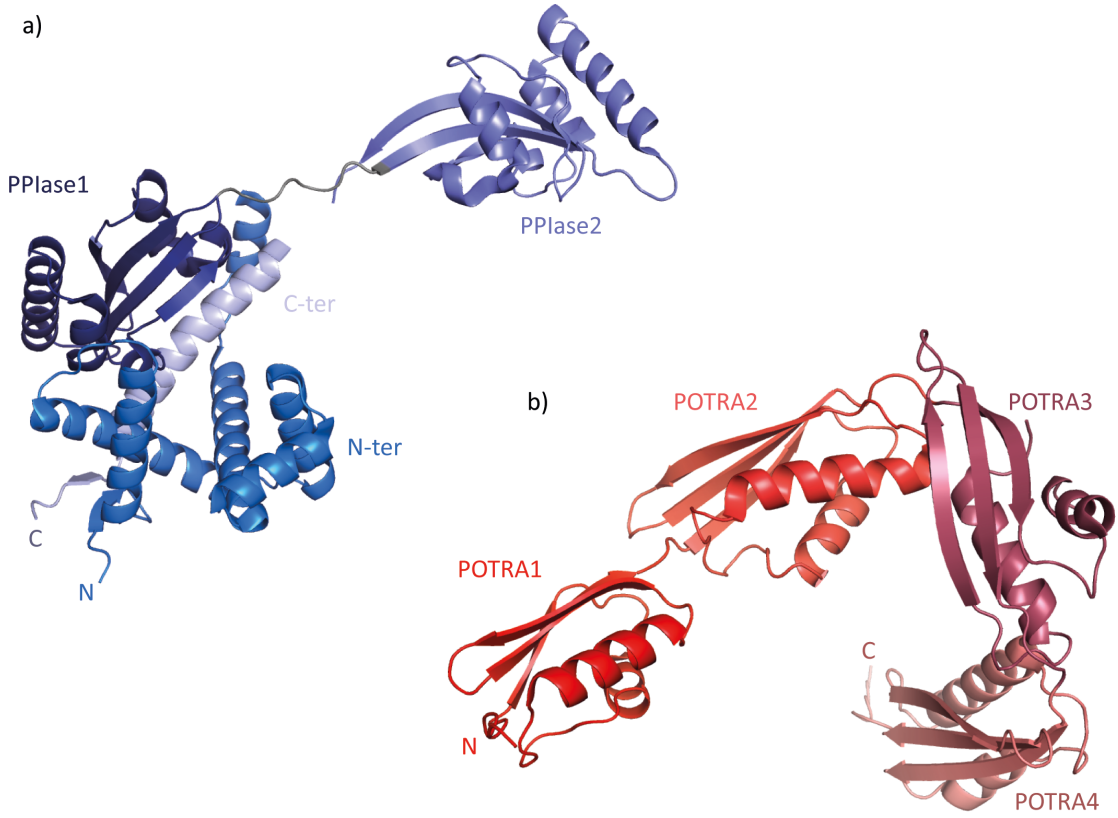


Figure 2.2. Crystal structures of periplasmic chaperones in *E. Coli*. a) Crystal structure of the chaperone protein SurA where the different domains are depicted with a blue shade (PDB: 1M5Y)[5]. b) Crystal structure of the first four POTRA domains of BamA where the different domains are depicted with a red shade (PDB: 2QDF)[12].

Results

Characterization of the periplasmic chaperone SurA *in vitro*

To assess the state and suitability of SurA for *in vitro* NMR studies, 2D [^{15}N, ^{1}H]-TROSY-HSQC spectra of the protein in two different buffers and at four different temperature conditions were recorded. All spectra featured a wide chemical shift dispersion as well as a single and coherent set of the amide resonances, characteristic of a folded protein in a single conformation. Among these conditions, the protein gave the best spectral characteristics at a temperature of 37°C and in a buffer containing 20 mM KPi, 100 mM KCl, pH 6.5 and 0.5 mM EDTA, chosen as NMR conditions for all subsequent experiments (Fig 2.3a).

One of the most important aspect for the mechanism by which chaperone proteins maintain their OMP substrate in an unfolded folding competent state during transport through cell compartments, was proposed to be the formation of a cavity by chaperone oligomerization, therefore providing a protective environment for the aggregation-prone substrate[6]. In the case of SurA, Xu *et al*. proposed the formation of the cavity through dimerization of the core domain, upon substrate binding on the basis of co-crystallization studies with model peptides[7]. To test whether the dimerization is also happening in the absence of substrate, 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N,^{2}H]-SurA at increasing concentrations of 300 μM, 600 μM and 1 mM were recorded (data not shown). The comparison of the normalized spectra showed a decrease of intensity of the resonances with increasing concentration. Interestingly, the resonances located in the spectral region typical for β-strand secondary structure (8.5 < δ(^{1}H) < 10 ppm) undergo larger intensity decrease than the peaks in the random-coil region of the spectrum (7.5 < δ (^{1}H) < 8.5 ppm), corresponding to amide moieties in loops or unfolded regions. This indicates an oligomerization of the protein, where the increase in size of the molecule causes a general increase of the width, and decrease of intensity, of the resonances with a weaker effect on the loops or unfolded regions due to their flexibility. To characterize this aspect further, SurA oligomerization in a concentration range from 15.7 to 500 μM was quantified by analytical ultracentrifugation. The protein sedimentation coefficient c(s), directly related to particles size in solution was fitted, showing an increase of size from the SurA monomer, at 4.6 S, towards its dimer, at 6.2 S (Fig 2.3b). Thus, to determine SurA dimerization constant (K_D), the weight averaged s-values (s_W) calculated form the AUC curves were fitted to a monomer-dimer model (Eq. 6, Material and Methods section) (Fig 2.3c) giving a K_D value of 1.2 mM for the full length SurA.

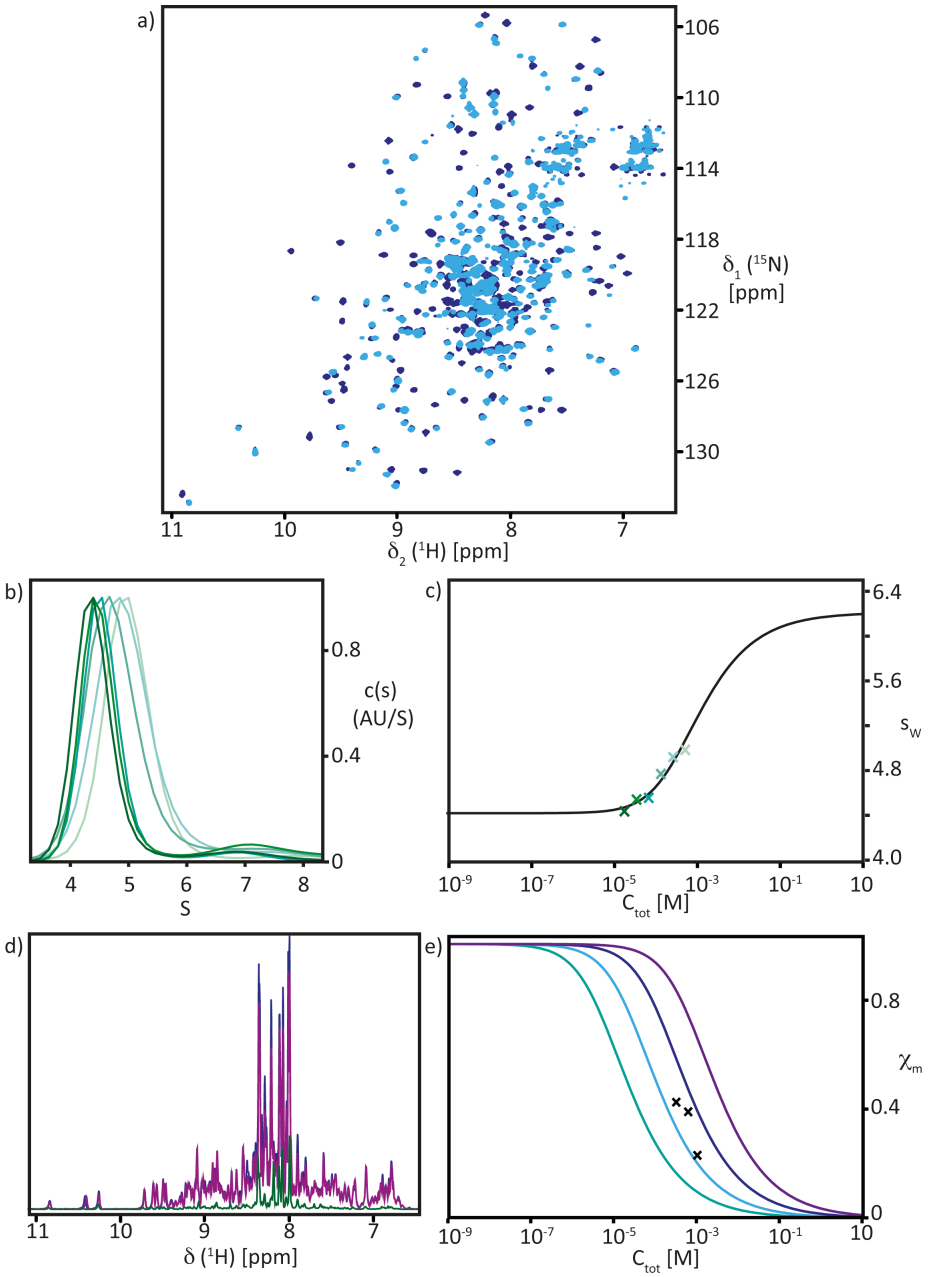


Figure 2.3. SurA characterization. a) Overlay of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 300 μM [*U*-^{15}N,^{2}H]-SurAfl in NMR buffer at 37°C (dark blue) and 300 μM [*U*-^{15}N, ^{2}H]-SurAΔP2 in NMR buffer at 37°C (light blue). b) AUC titration curve of SurAfl at 15.7, 31.3, 62.5, 125, 250 and 500 μM (green color gradient) where the x axis represents the sedimentation coefficient S and the y axis the normalized sedimentation coefficient distribution c(s). c) Weight averaged s-values (s_W) calculated from the c(s) distribution for each concentration (b) (crosses) fitted to monomer-dimer model (Eq. 6, Material and Methods section) shown as solid line. d) Overlay of ^{1}H projections, of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 300μM (dark blue), 600 μM (pink) and 1 mM (green) [U-^{15}N,^{2}H]-SurAΔP2 in NMR buffer at 37°C. The projections are normalized to the lowest concentration using the following formula: $[Sample] = [Ref] \cdot SF \cdot p_{1\,sample}/p_{1\,ref}$. e) Ratio between the integrals at 8.5 to 10 and at 7.5 to 8.5 ppm of the ^{1}H projections at 300, 600 μM and 1 mM (d) *vs.* concentration (crosses). The data were compared to fits of a monomer-dimer model using Eq. 2 (Material and Methods section) using K_D of 20, 100, 500 μM and 2.5 mM (left to right) shown as solid lines.

To determine whether the dimer interface is located in the core module of the protein, similar 2D [^{15}N,^{1}H]-TROSY-HSQC experiments were recorded on a SurA construct lacking its PPIase 2 domain (SurAΔP2). As for the full-length, the spectra displayed wide chemical shift dispersion and a single and coherent set of resonances (Fig. 2.3a). Furthermore, a similar oligomerization behavior was observed with concentration (Fig 2.3d). Measurements of the ratio between the integrals at 8.5 to 10 and at 7.5 to 8.5 ppm of the ^{1}H projections give an estimate of the monomeric fraction (χ_m) of SurAΔP2 at each concentration. This rough assumption was made assuming a ratio of 1 for the monomer only ($\chi_m = 1$) and 0 for the dimer only ($\chi_m = 0$) as the 2D [^{15}N,^{1}H]-TROSY-HSQC spectra could not be directly compared due to different acquisition parameters (number of complex points and carrier position in both dimensions, acquisition and evolution times). Plot of these ratio *vs.* the total concentration were compared to fits of a monomer-dimer model using Eq. 2 (Material and Methods section) with K_D of 20, 100, 500 μM and 2.5 mM (Fig 2.3e). This approach estimates the dimerization constant for SurAΔP2 between 100 and 500 μM, smaller than for the full-length protein. However, the K_D calculated from AUC data for SurAfl could be overestimated due to the possible increase in frictional ratio at high concentration used to measure the sedimentation coefficient distribution. Thus, both AUC and NMR measurements have to be repeated in order to better estimate the K_D and determine whereas the binding interface is located in the core module or in the PPIase2 domain of SurA.

POTRA interacts with the N- and C-terminal domains of SurA

During OMP biogenesis, SurA chaperone activity is carried out by its N- and C-terminal domains, while the PPIase domains are dispensable for the interaction[4]. To further investigate which domains of SurA are involved in the interaction with POTRA, the two PPIase domains of SurA were expressed independently (SurA PPIase1 and SurA PPIase2) and 2D [^{15}N,^{1}H]-TROSY-HSQC spectra were recorded (Fig. 2.4). The 2D [^{15}N,^{1}H]-TROSY-HSQC spectrum of SurA PPIase2 showed well-dispersed resonances overlaying with the peaks of the SurAfl spectrum, indicating a similar fold of this domain alone and within the protein (Fig. 2.4a). Conversely, while the 2D [^{15}N,^{1}H]-TROSY-HSQC spectrum of SurA PPIase1 features wide chemical shift dispersion, characteristic of a structured domain, a large number of resonances did not overlay with the peaks of the SurAfl spectrum (Fig. 2.4b). This is consistent with previous observations showing that within the core module, a part of the PPIase1 domain is occluded by the N and C-terminal domains and displays

a displaced β-strand compared to the typical PPIase structures[5].

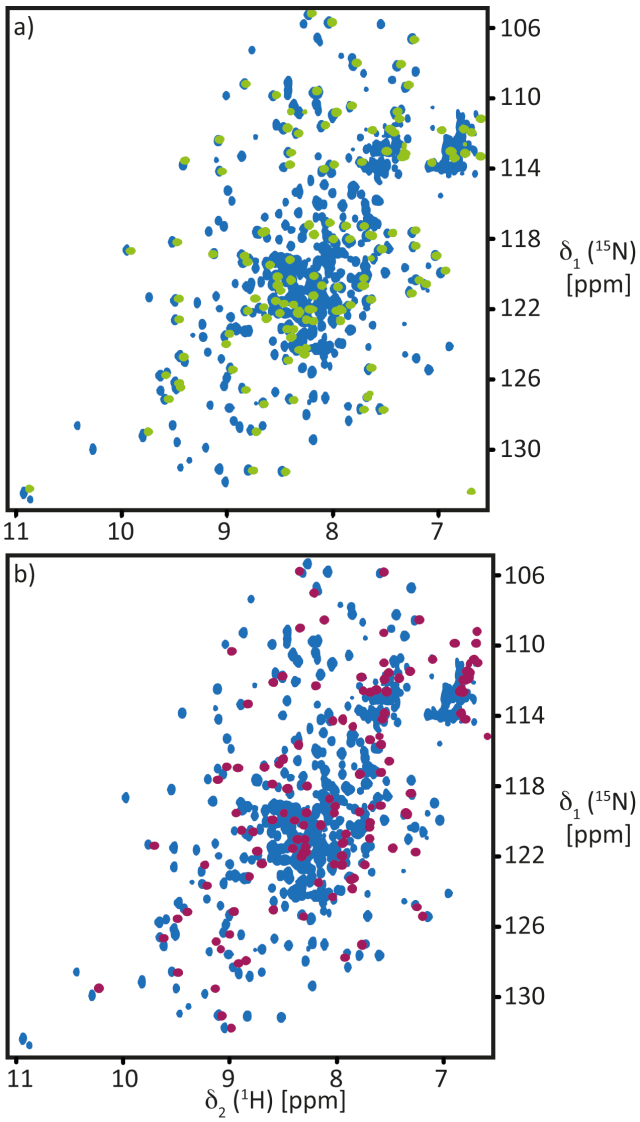


Figure 2.4. Individual domains of SurA. a) Overlay of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 300 μM [*U*-^{15}N,^{2}H]-SurAfl (blue) and of 300 μM [*U*-^{15}N,^{2}H]-SurA PPIase2 (green) in NMR buffer. b) Overlay of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 300 μM [*U*-^{15}N]-SurAfl (blue) and of 400 μM [*U*-^{15}N]-SurA PPIase1 (pink) in NMR buffer.

Interaction of these singles domains with POTRA was further analyzed, using a construct containing the two first domains only (POTRA12) as we show below that they are sufficient for the interaction with SurA. 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of apo-[*U*-^{15}N]–POTRA12 and in presence of the individual PPIase domains at a 1:0.5 and 1:1.5 POTRA12:SurAPPIase2 and 1:1 POTRA12:SurAPPIase1 ratio were recorded (Fig 2.5a and b respectively). No chemical shifts changes were observed upon SurAPPIase1 and SurAPPIase2 titration, indicating that these individual domains do not have a detectable affinity for POTRA.

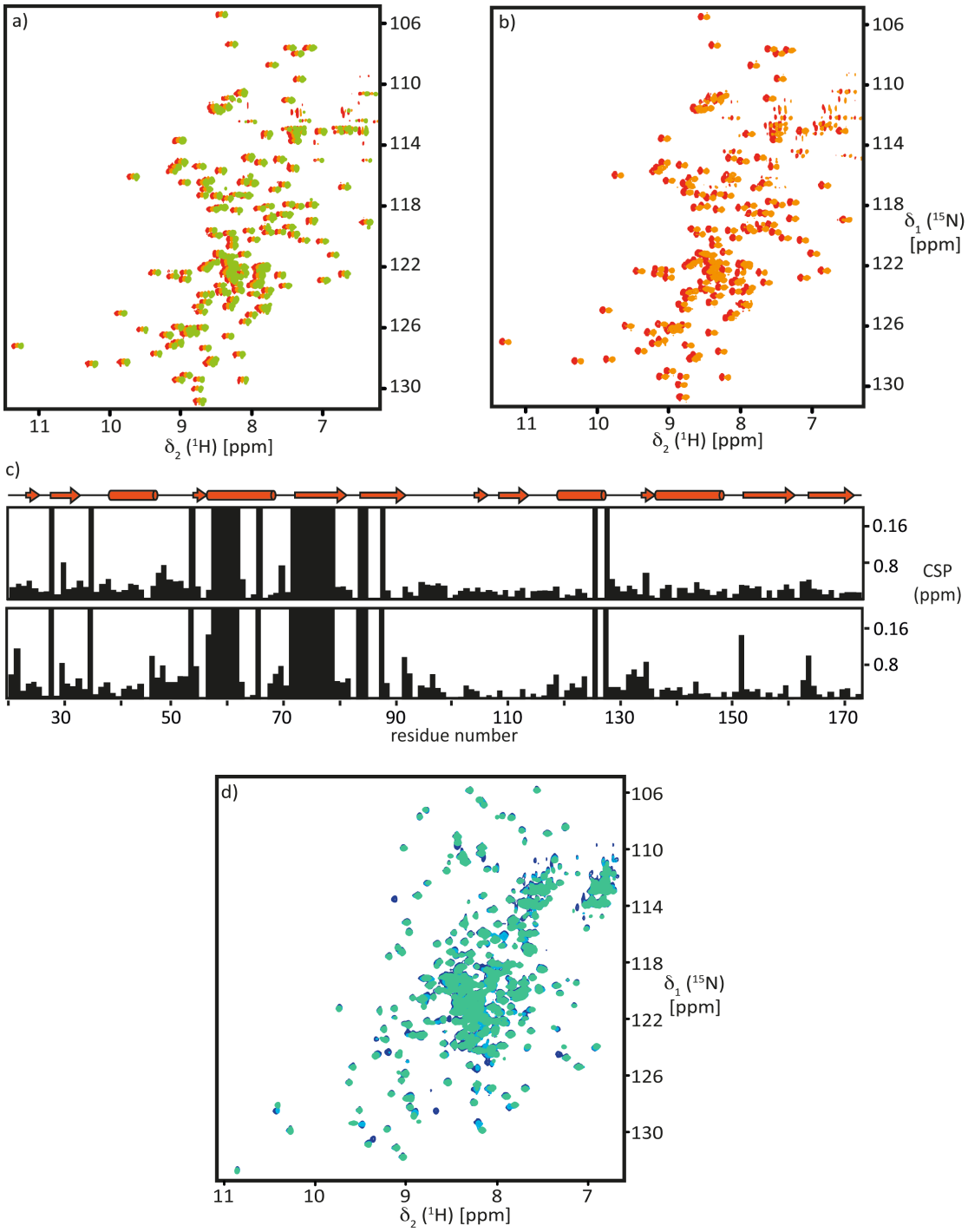


Figure 2.5. Identification of the SurA binding interface for POTRA. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 220 μM [*U* -^{15}N]-POTRA12 in NMR buffer (red) upon addition of 0.5 molar (orange) and 1.5 molar (green) equivalent of SurAPPIase2. For a clear visualization the resonances of the titrated POTRA are arbitrarily shifted. b) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 300 μM [*U*-^{15}N]-POTRA12 in NMR buffer (red) upon addition of 1 molar equivalent of SurAPPIase1 (orange). For a clear visualization the resonances of the titrated POTRA are arbitrarily shifted. c) Chemical shift perturbations of POTRA12 residues, in the presence of SurAfl (top) and SurAΔP2 (bottom) plotted *vs.* the amino acid residue number of POTRA12. The secondary structure of the POTRA domains is indicated above. d) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 400 μM [*U*-^{15}N,^{2}H]-SurAΔP2 in NMR buffer (dark blue) upon addition of 1 molar (blue), and 2 molar (green) equivalent of POTRA12.

These finding were confirmed by analyzing 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N]-POTRA12 (Fig 2.5c), displaying similar changes in chemical shifts upon SurAfl and SurAΔP2 titration. As well, similar resonances were affected on the 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of ^{15}N-SurAfl and ^{15}N-SurAΔP2, displayed in Figure 2.5d, upon POTRA12 titration.

To further confirm that the N- and C-terminal domains alone are required for the interaction with POTRA, a SurA construct lacking the two PPIase domains (SurANC) was designed, but studies on this construct were not pursued due to its low expression yield and precipitation. As well, the expression of the N- and C-terminal domains independently was not tested as it had previously been shown that these domains alone are not stable and functional[4].

The N- and C-terminal β-strands of SurA are involved in the OMP delivery mechanism

To characterize the interaction surface on SurA upon POTRA binding, sequence-specific resonance assignments had to be obtained for the NMR spectra of the SurAΔP2 construct. NMR sequence-specific backbone assignment was carried out recording 3D TROSY-HNCA, 3D TROSY-HNCACB[17], 3D TROSY-HNCO[18], 3D TROSY-HNCACO[19] and 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY[20] on a 600 μM [*U*-^{15}N,^{2}H]-SurAΔP2 sample. With these data, about 40 % of the backbone resonances of the protein could be assigned in a sequence-specific manner (Fig 2.6). The chemical shift perturbations (CSP) between the 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N,^{2}H]-SurAΔP2 apo and in complex with POTRA12 at a 1:1 ratio were measured (Fig 2.7a) and, for the assigned residues, transferred on SurA crystal structure (PDB: 1M5Y) (Fig 2.7b and c). As observed previously, SurAPPIase 1 domain is not playing a role in the interaction with POTRA as no chemical shift changes could be observed for its residues. The largest CSP were measures on the amino acids constituting the N- and C-terminal β-strands of the protein, forming an anti-parallel β-sheet, as well as on α1, α2 and α3 helices of the N-terminal domain. The anti-parallel β-sheet formed by β-strands located at the beginning and at the end of the N-terminal and C-terminal domains respectively was shown to act as a structural support of the protein[21]. Observation of major chemical shift changes on these residues can indicate that (1) they are part of the binding interface with POTRA or that (2) the interaction of POTRA at an allosteric site induce an opening of the anti-parallel β-sheet leading to a partial destabilization of the protein structure which could be the release mechanism of the bound OMP substrate, in absence of external energy.

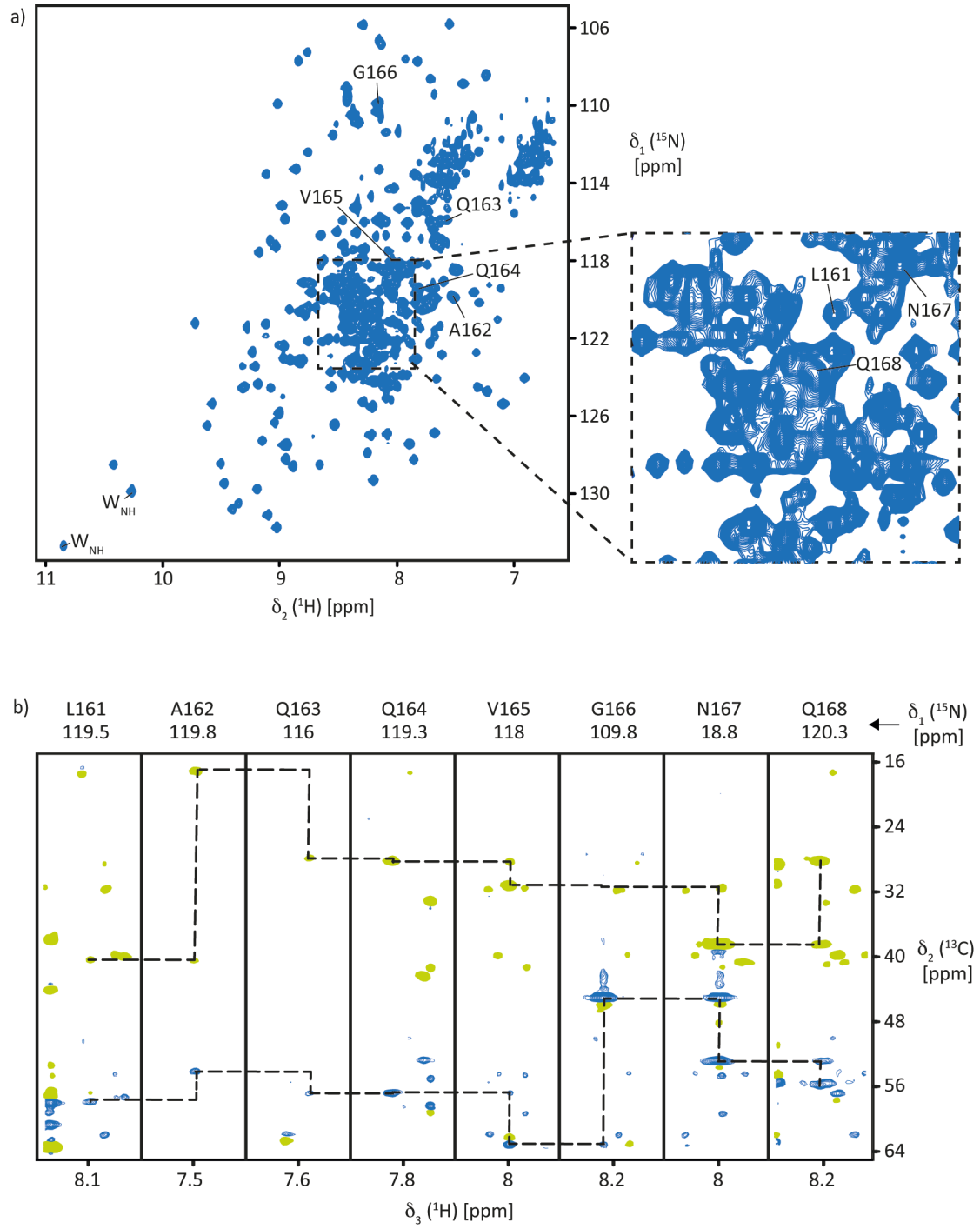


Figure 2.6. Sequence-specific resonance assignment of SurAΔP2. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 400 μM [*U*-^{15}N,^{2}H]–SurAΔP2 in NMR buffer and zoom of the central part of the spectrum indicated with a dashed box. Selected sequence-specific resonance assignments are indicated. b) Strips from 3D TROSY-HNCACB spectrum of 700 μM [*U*-^{15}N,^{2}H]–SurAΔP2 in NMR buffer. Resonances belonging to C_α atoms are displayed in blue and Resonances belonging to C_β atoms are displayed in yellow.

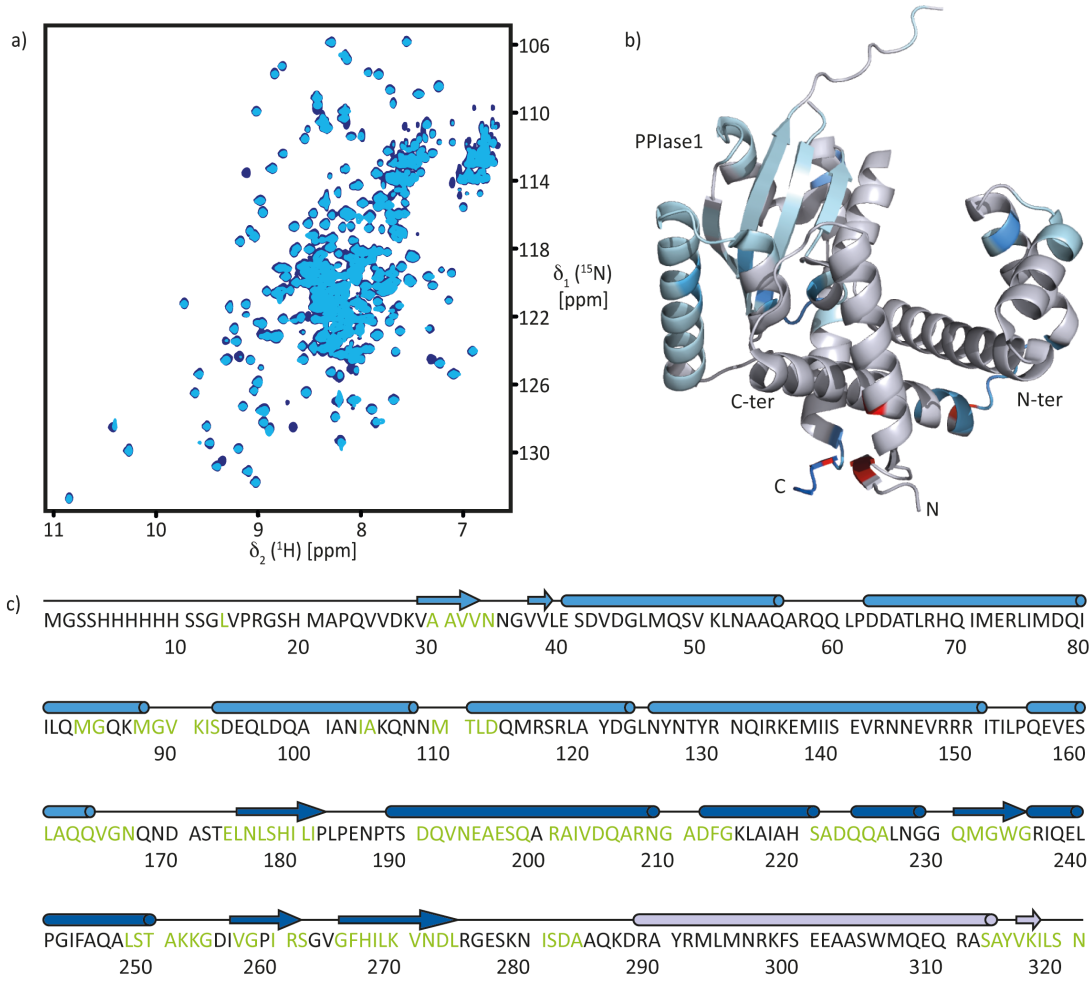


Figure 2.7. SurAΔP2 binding interface with POTRA12. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 400 μM [*U*-^{15}N,^{2}H]–SurAΔP2 in NMR buffer (dark blue) upon addition of 1 molar equivalent of POTRA12 (blue). b) Ribbon representation of SurAΔP2 (PDB: 1M5Y), where changes upon POTRA12 interaction are indicated by a cyan gradient for CSP values between 0 and 0.13. Amino acid resonances disappearing upon interaction with POTRA12 are indicated in red. c) Amino acid sequence of SurAΔP2 where the assigned residues are highlighted in green. The secondary structure elements of the N-terminal domain (blue), PPIase1 domain (dark blue) and C-terminal domain (light blue) are indicated above.

Interaction between SurA and the first two domains of POTRA *in vitro*

In the cell, the interaction between the BamA-POTRA domains and SurA is supposed to result in the handover of SurA-bound OMPs to the POTRA domains for subsequent assembly and insertion into the outer membrane. *In vitro*, the complex of full-length SurA (SurAfl) with [*U*-^{15}N,^{2}H]–labeled POTRA containing the 5 domains (POTRA15) was assembled at a 1:1 ratio by mixing the two proteins in NMR buffer and a 2D [^{15}N,^{1}H]-TROSY-HSQC spectrum was recorded

for comparison with the spectrum of the apo form of POTRA15. The apo state appeared to be consistent with previous NMR studies of POTRA and assignments were achieved (B. Burmann, unpublished data).

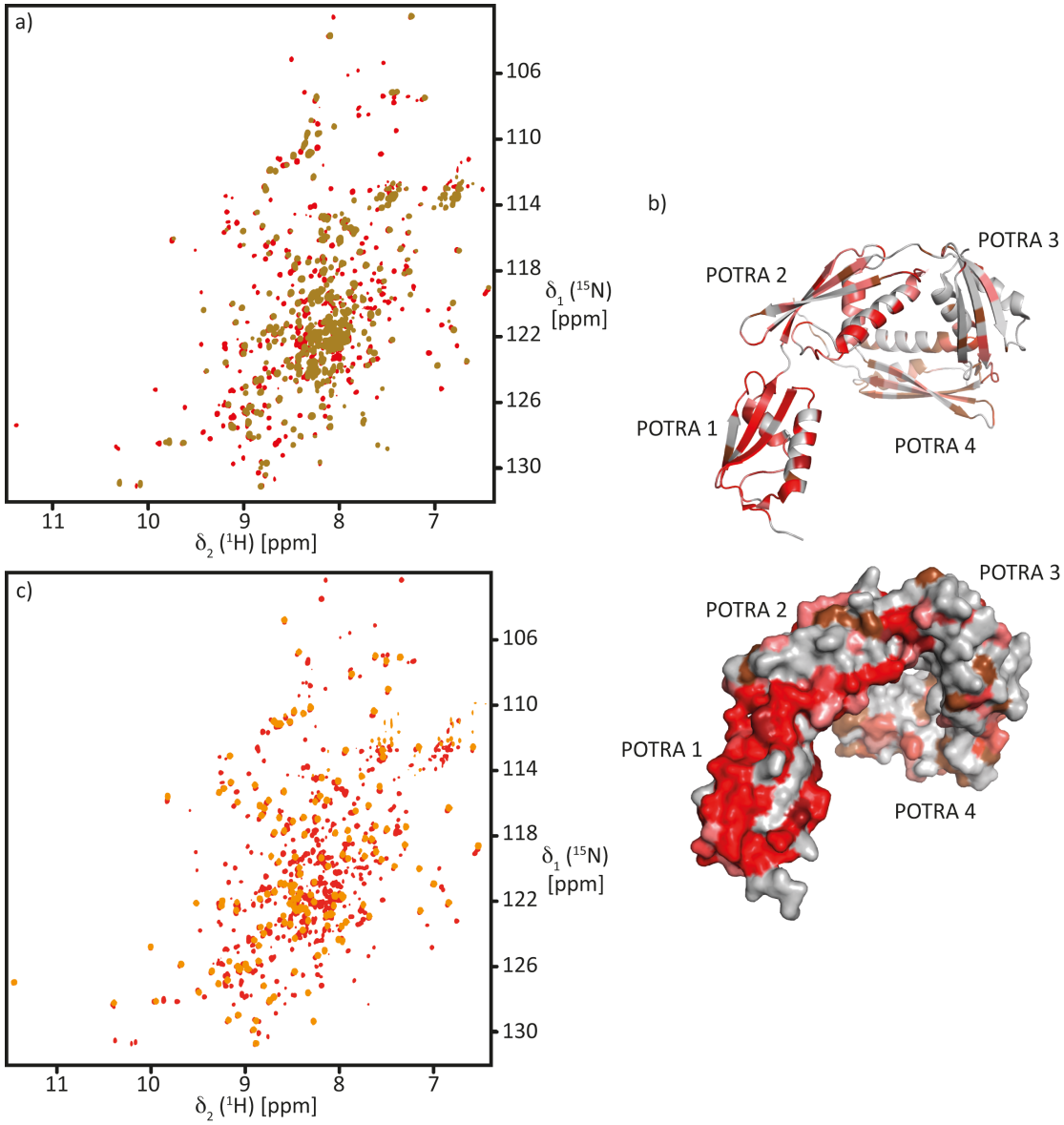


Figure 2.8. POTRA-SurA interaction. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N,^{2}H]-POTRA15 at 400 μM in NMR buffer (red) and in interaction with 1 molar equivalent of SurAfl (gold). b) Cartoon representation (top) and surface representation (bottom) of POTRA15 (PDB: 2QDF) where changes upon interaction with SurA are indicated by a red gradient for $\frac{I_{apo}}{I_{holo}}$ values between 1 (red) and 0 (brown). c) Overlay of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N,^{2}H]-POTRA15 at 400 μM in NMR buffer (red) and 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U* -^{15}N,^{2}H]-POTRA12 at 270 μM in NMR buffer (orange).

Large chemical shift changes between the apo and the holo [*U*-^{15}N,^{2}H]–POTRA15 NMR spectra were detected as well as disappearance of several resonances (Fig 2.8a) (B. Burmann, unpublished data). The mapping of these changes on the crystal structure of POTRA15 (PDB 2QDF) reveals that mainly the first two POTRA domains are affected upon SurA interaction (Fig 2.8b). Additional effects in the other POTRAs are located in the interfaces between the individual domains. Through the localization of its binding interface with SurA, a construct containing the two first POTRA domains was designed (POTRA12). 2D [^{15}N,^{1}H]-TROSY-HSQC spectrum recorded on [*U*-^{15}N]–POTRA12 overlaid with the corresponding POTRA15 (Fig 2.8c), indicating that the two domains alone have the same structure than within the full-length construct and that the truncation of domains 3 to 5 did not affect the structure of the individual domains. The 2D [^{15}N,^{1}H]-TROSY-HSQC spectrum of [*U*-^{15}N]–POTRA12 at a 1:1 ratio of POTRA:SurA display similar changes in chemical shifts as observed for the full-length POTRA (data not shown). On SurA, both 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N]–SurAfl at 1:1 SurA:POTRA15 and SurA:POTRA12 equivalents display chemical changes compared to the apo spectrum of SurAfl and the chemical shift perturbations observed are comparable upon addition of the different POTRA constructs (data not shown), showing that these two domains are sufficient for the interaction.

SurA interacts with the first domain of POTRA

The interaction between SurA and POTRA was suggested previously, on the basis of *in vivo* crosslinking experiments, to be located in the α2 helix of the first POTRA domain[16], consistent with our results obtained by NMR spectroscopy. To confirm the interaction site and obtain details at the atomic level, the two first POTRA domains were expressed independently (POTRA1 and POTRA2) and 2D [^{15}N,^{1}H]-TROSY-HSQC spectra with increasing amounts of SurA were recorded.

The 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of *[U*-^{15}N]–POTRA1 and [*U*-^{15}N]–POTRA2 showed well-dispersed resonances overlaying with the POTRA12 NMR spectrum, indicating a similar fold of these domains alone and linked together (Fig. 2.9a). Only the amine resonances of residues Q35 and R36 located in the loop between β1 and α1, residues N71 to V75 in the loops between α2 helix and β2 strand of POTRA 1, K89, E90, R91 and T93 located in the linker between the two domains and the residues E123 to I130 located in the loop between α1 and α2 helices of POTRA2 showed important changes in chemical shifts (Fig. 2.9a and b). Their changes in chemical shifts confirms that these residues are part of the domain interface. From the NMR

spectrum of POTRA12 all the assignments could be transferred to POTRA1 and POTRA2.

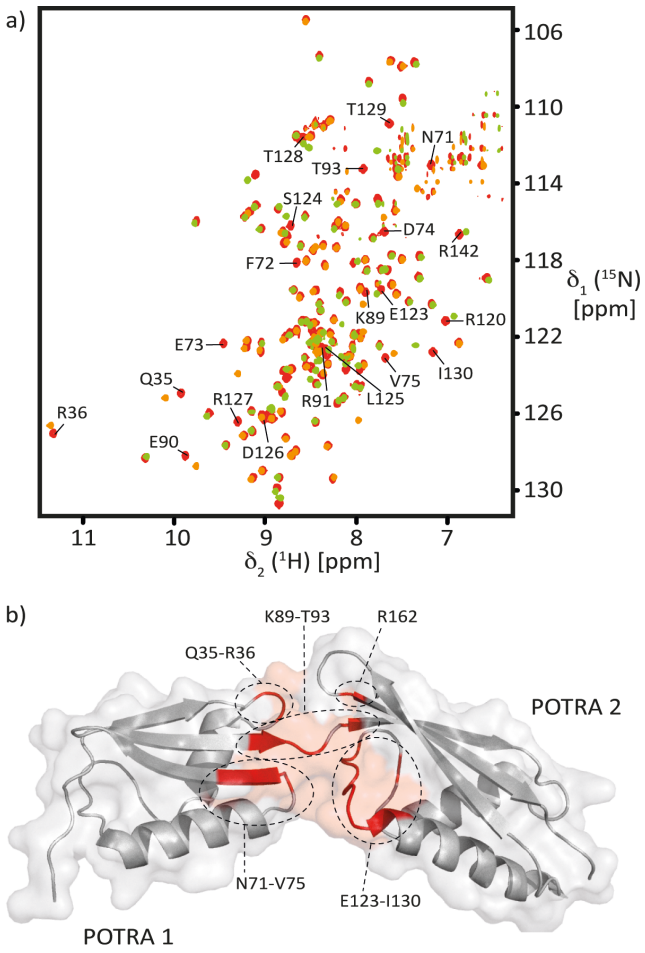


Figure 2.9. Individual POTRA domains. a) Overlay of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 300 μM [*U* -^{15}N]-POTRA12 (red), 270 μM [*U*-^{15}N]-POTRA1 (orange) and 270 μM [*U*-^{15}N]-POTRA2 (green) in NMR buffer. The sequence-specific resonance assignment of the residues undergoing chemical shift change in the isolated POTRA domains is indicated. b) Ribbon representation of the first and second POTRA domains (PDB: 2QDF), where the residues undergoing chemical shift change in the isolated POTRA domains are indicated in red.

Upon addition of 0.5 to 8 molar equivalent of [*U*-^{2}H]–SurAΔP2, no changes in chemical shifts were observed in ^{15}N-POTRA2 NMR spectra (Fig 2.10a). This indicates that this individual domain does not have a detectable affinity for SurA. However it is worth noting that POTRA 2 resonances within the bi-domain construct POTRA12 experience chemical shift changes upon SurA addition that are most likely triggered by POTRA1 (data not shown). These changes indicate an indirect interaction with SurA. The same experiments on the isolated POTRA1 revealed large chemical shift perturbations of the resonances upon addition of SurA (Fig 2.10b), in agreement

to the ones observed for the POTRA12 construct. POTRA1 was thus used as the minimal construct of interaction for all further studies.

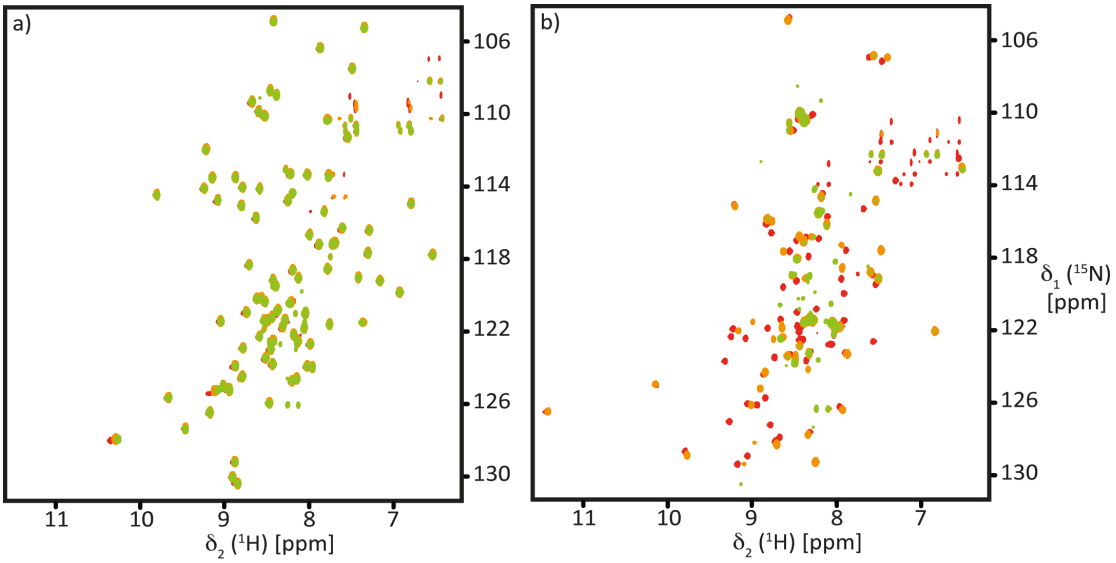


Figure 2.10. SurA interaction with single POTRA domains. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 270 μM [*U*-^{15}N]-POTRA2 in NMR buffer (red) and in interaction with 1 (orange) and 4 molar equivalent (green) of [*U*-^{2}H]-SurAΔP2. b) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 270 μM [*U*-^{15}N]-POTRA1 in NMR buffer (red) upon addition of 1 (orange) and 4 (green) molar equivalent of [*U* -^{2}H]–SurAΔP2.

Surface plasmon resonance confirms that POTRA1–SurAΔP2 is a suitable minimal construct for interaction studies

Surface plasmon resonance (SPR) measurements were performed to measure the affinity between the different POTRA and SurA constructs. POTRA1, POTRA2 and POTRA12 were immobilized on SPR chips by covalent coupling of their amine groups to the carboxymethylated dextran matrix. Then, SurAfl and SurAΔP2 constructs were loaded on the chip and the dissociation constants (K_D) were measured using the kinetic/affinity method[22]. The dissociation constant measured between POTRA12 and SurAΔP2 domains is equal to 12 ± 6 μM, comparable to the K_D of POTRA12–SurAfl interaction of 16 ± 1 μM (Fig. 2.11 and Table 2.1), in excellent agreement with the observed behavior in the NMR titrations (Fig 2.5d). As well, the specific interaction of POTRA1 with SurA was confirmed by affinity measurements, as POTRA2 did not show any detectable interaction with SurA, while POTRA1 and POTRA12 did. The dissociation

constants (K_D) between POTRA1 and SurAΔP2 and POTRA12 and SurAΔP2 are equal to 12 ± 4 μM and 12 ± 6 μM respectively, indicating a similar affinity for SurA, mediated through the first POTRA domain (Fig. 2.11 and Table 2.1). These results corroborate that SurA interacts with the first POTRA domain only.

	POTRA12	POTRA1	POTRA2	POTRA1R64A
SurAfl	16 ± 1 μM	13.4 ± 1 μM	> 100 μM	210 ± 44 μM
SurAΔP2	12 ± 6 μM	12 ± 4 μM	> 100 μM	> 500 μM

Table 2.1. Interaction constants (K_D) between the different constructs of POTRA and SurA measured by surface plasmon resonance (SPR) using the kinetic/affinity method.

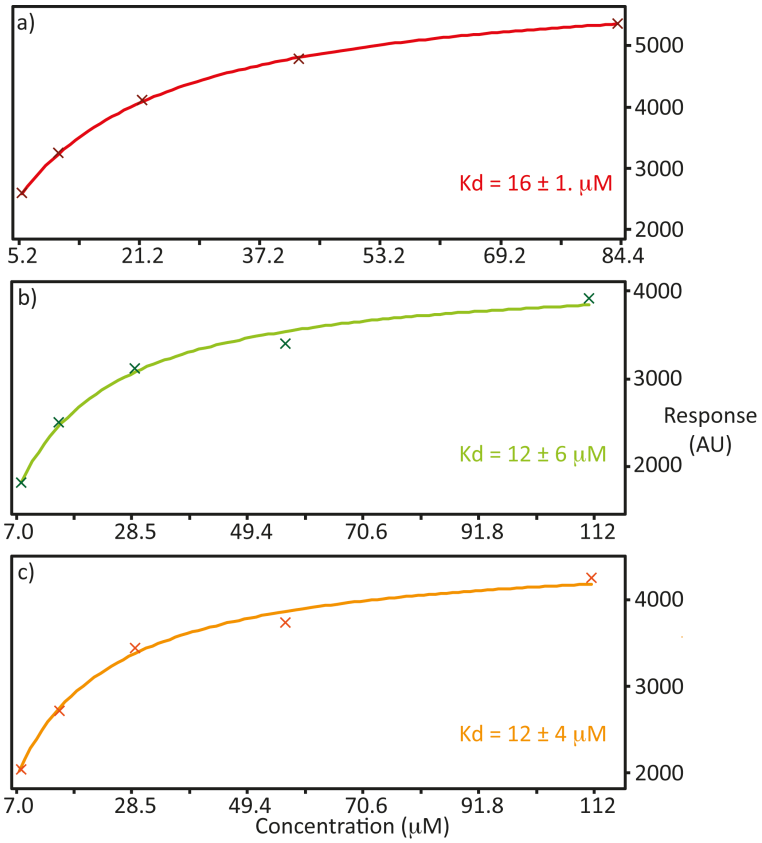


Figure 2.11. Determination of dissociation constants (K_D) between SurA and POTRA by surface plasmon resonance (SPR). a) Affinity curve of POTRA12 titrated with increasing amounts of SurAfl. b) Affinity curve of POTRA 12 titrated with increasing amounts of SurAΔP2. c) Affinity curve of POTRA1 titrated with increasing amounts of SurAΔP2. The different POTRA constructs were immobilized by amine coupling to the dextran matrix, then the different SurA constructs were loaded on the chip. The data points (crosses) were measured using the kinetic/affinity method[22]. The solid lines correspond to the fitted curves.

The hydrophobic core of POTRA1 is affected upon interaction with SurA

The large and numerous chemical shifts changes observed in the 2D [^{15}N, ^{1}H]-TROSY-HSQC spectrum of [*U*-^{15}N]–POTRA1 upon addition of SurAΔP2 (Fig. 2.12a) indicate a major structural rearrangement of the protein upon interaction.

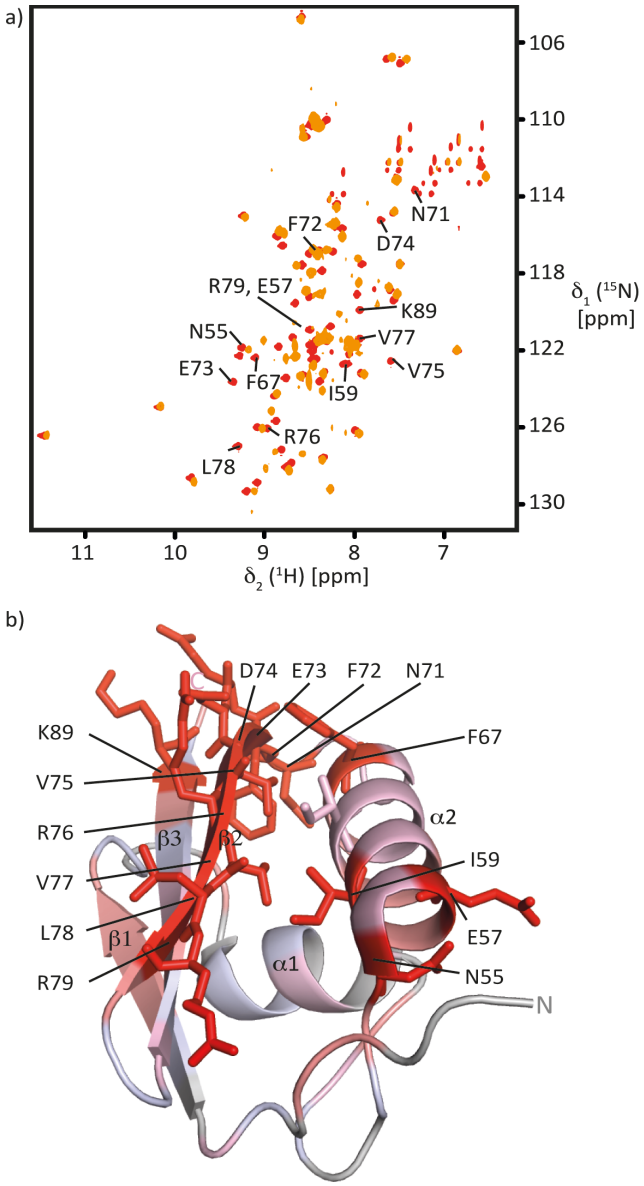


Figure 2.12. Changes on POTRA1 upon addition of SurAΔP2. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 270 μM [*U*-^{15}N]-POTRA1 in NMR buffer (red) and upon addition of 1 molar equivalent of [*U*-^{2}H]–SurAΔP2 (orange). The sequence-specific resonance assignment of the residues disappearing upon interaction with SurAΔP2 is indicated. b) Ribbon representation of POTRA1 (PDB: 2QDF) where changes upon addition of 1 molar equivalent of SurAΔP2 are indicated by a pink gradient for CSP values between 0 and 0.2 ppm and amino acid resonances disappearing are indicated in red. The residues disappearing upon interaction with SurAΔP2 are indicated.

In order to map the interaction surface of SurA on POTRA1 at the atomic level, chemical shift perturbations (CSP) were calculated between the NMR spectra of the apo form of POTRA1 and after addition of 1 molar equivalent of [*U*-^{2}H]–SurAΔP2. The CSP value were plotted on the crystal structure of POTRA 1 (PDB 2QDF), shown in Figure 2.12b, were the intensity of the CSP are depicted with a pink gradient. The major perturbations are located in the α2 helix as well as, to a lesser extend, in the β1 and β3 strands, while the α1 helix is less affected. In addition to these changes, residues from N71 to R79 constituting the β2 strand K89 in the β3 strands as well as N55 and E57, and I59 and F67 located in the α2 helix of POTRA1, facing the β2 strand, are line-broadened below the detection limit, already after addition of 0.5 molar equivalent of [*U*-^{2}H]–SurAΔP2 (Fig. 2.12b). The chemical shift perturbations observed on POTRA indicate changes on α2 helix as well as β1 and β3 strands, due to direct or indirect interaction with SurA, consistent with previous cysteine-specific *in vivo* cross-linking experiments proposing that an interaction of SurA with the α2 helix of POTRA1[16]. Workman *et al* also showed that the deletion of a residue of the α2 helix of POTRA1 leads to severe defect in the OMP biogenesis that can be partially recovered by the insertion of an amino acid that restores its wild-type helical turn. This functional defect can be due to the disruption of the hydrophobic core formed between residues of the α2 helix and β2 strand, which can be crucial for POTRA1 stability and function. This is also consistent with the conservation of the residues I59, D74, V75 and V77, forming a hydrophobic cluster[12]. The disappearance of resonances from residues forming this hydrophobic core, observed in our 2D [^{15}N,^{1}H]-TROSY-HSQC NMR spectra of [*U*-^{15}N]–POTRA1 upon addition of SurAΔP2, supports their importance in POTRA function. It reflects a chemical exchange of these moieties between two environments in the intermediate range of the NMR time scale where their kinetic constant of exchange (k_{ex}) about the order of the frequency difference between the resonances of these two states, prevents its detection. This exchange, observed in the presence of SurAΔP2, can be intermolecular, between the hydrophobic residues and SurA, upon interaction or intramolecular, induced by an allosteric SurA binding, where these residues will exchange between a bound conformation, forming the hydrophobic core and a free conformation at an intermediate rate. This suggests a possible mechanism for the transfer of the unfolded OMP from SurA to POTRA where the hydrophobic residues created a rim acting as a receptor for the unfolded substrate.

The first POTRA domain undergoes major structural rearrangements upon interaction with SurA

Analysis of the 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N]–POTRA1 in presence of 8 molar equivalents of SurAΔP2 also revealed the presence of new resonances, appearing in the random-coil region of the spectrum (7.5 < δ (^{1}H) < 8.5 ppm) (Fig. 2.13a). Their position corresponds to resonances of amide groups in a random-coil conformation, lacking defined secondary structure, indicating an unfolding of POTRA1 amino acids upon addition of SurA. This "unfolding" effect appeared to be specific, as it was observed only for POTRA1. The 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N]–POTRA2 alone do not display this set of new resonances when titrated with SurA (data not shown). To investigate the origin of this "unfolding" effect, the stability of the POTRA domains was investigated. Thermal melting curves of POTRA1, POTRA2 and POTRA12 constructs were measured using circular dichroism (CD) spectroscopy (Fig. 2.13b). Melting temperature (T_m), enthalpy of denaturation at T_m and Gibbs free energy of denaturation were obtained by non-linear least-squared fit of the measured data using a two-state model of the thermal unfolding transitions (Eq. 4 and 5 in Material and Methods section) (Table 2.2). The high melting temperature values obtained suggest stable POTRA folds for the three different constructs, in agreement with previously published values[14]. Comparison of the Gibbs free energy of the different POTRA constructs revealed a slight stabilization of POTRA1 by POTRA2, indicated by a difference in $\Delta G_{310.15}$ of $-0.58\ kJ \cdot mol^{-1}$. On the other hand, the addition of POTRA1 to POTRA2 induces a strong destabilization of the second domain indicated by a difference in $\Delta G_{310.15}$ of $4.6\ kJ \cdot mol^{-1}$.

	T_m (K)	$\Delta H_m\ (kJ \cdot mol^{-1})$	$\Delta G_{310.15}\ (kJ \cdot mol^{-1})$	$\Delta\Delta G_{310.15}(kJ \cdot mol^{-1})$
POTRA12	343	331.4	13.66	0
POTRA1	346.5	218	13.08	-0.58
POTRA2	343.5	291	18.28	4.6

Table 2.2. Stability data for the different POTRA constructs. The stability parameters were obtained from two-state analysis of the thermal unfolding transitions (Eq. 4 and 5 in Material and Methods section). For all constructs, the melting temperature (T_m), the enthalpy of denaturation at T_m, the Gibbs free energy of denaturation at 37°C ($\Delta G_{310.15}$) and the difference in Gibbs free energy to the POTRA12 construct ($\Delta\Delta G_{310.15}$) are given.

These values corroborate the observation done on the NMR spectra of POTRA1, POTRA2 and POTRA12 upon SurA titration (Fig. 2.13a) where the addition of SurA leads to "unfolding" of POTRA1 and POTRA12 but not POTRA2. This "unfolding" effect has to be investigated further to determine its precise role but, together with the possible chemical exchange between the residues constituting the hydrophobic core of POTRA1, it could suggest a mechanism of interaction and / or release of the OMP between POTRA and SurA proteins and transport of the OMP along the POTRA domains.

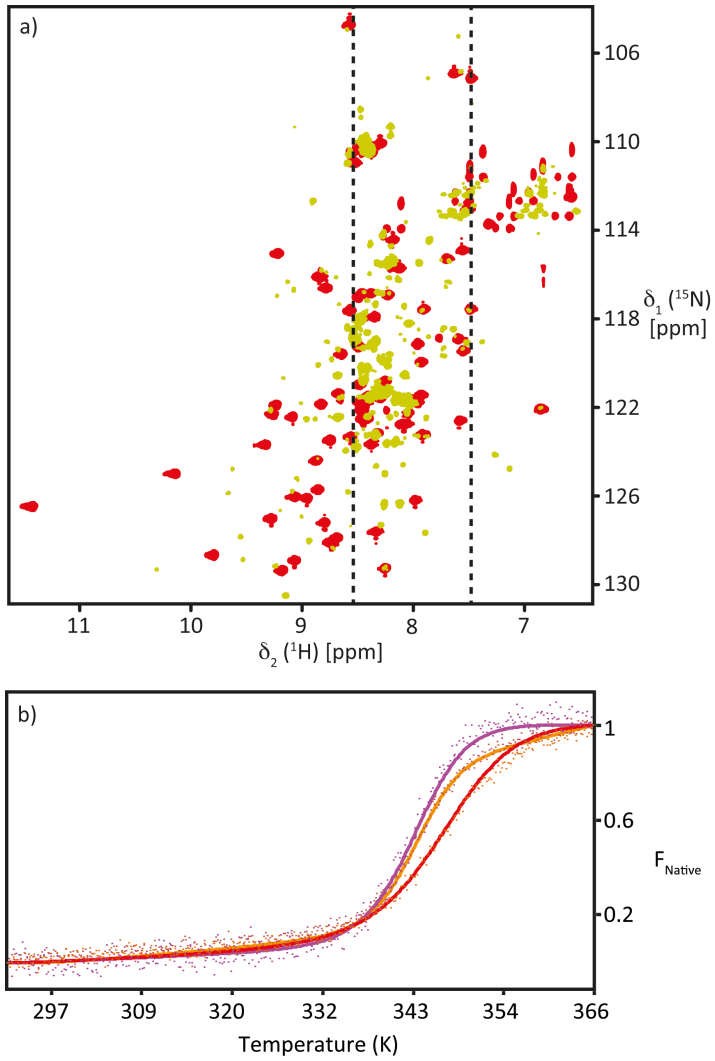


Figure 2.13. Stability of the different POTRA domains. a) 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 270 μM [*U*-^{15}N]–POTRA1 in NMR buffer (red) and upon addition of 8 molar equivalent of [*U*-^{2}H]–SurAΔP2 (yellow). The random coil region of the spectra is located between the dashed lines. b) Normalized melting curves of POTRA1 (red), POTRA2 (purple) and POTRA12 (orange) recorded by CD spectroscopy in CD buffer where the y axis corresponds to the fraction of native population. The solid lines correspond to the fitting curves, using a two-state model of the thermal unfolding transition (Eq. 4 and 5 in Material and Methods section).

Arginine 64, located in the α2 helix of POTRA1 plays a crucial role in the OMP delivery mechanism

In vivo studies of the interaction between POTRA and SurA showed that the deletion of arginine 64, located in the α2 helix of POTRA1, compromise the assembly of the OMP[15, 16]. This defect in the OMP assembly can be due to the crucial role of R64 but might as well be caused by an altered fold of the α2 helix upon R64 deletion. To investigate this aspect further, we designed a construct where arginine 64 of POTRA1 was mutated to an alanine (POTRA1R64A). The 2D [^{15}N, ^{1}H]-TROSY-HSQC spectrum of [*U*-^{15}N]–POTRA1R64A showed well-dispersed resonances overlaying with the POTRA1 NMR spectrum. Only residues in close vicinity of the mutated R64 experienced chemical-shift changes, which indicates that the mutation is not affecting the fold of the domain (data not shown). 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N]–POTRA1R64A in complex with SurAΔP2 at a 1:0.5 to 1:8 POTRA:SurA molar ratio were measured (Fig 2.14a) and chemical shift perturbations in the presence SurAΔP2 at a 1:1 molar ratio were measured and compared to the CSP observed for the wild type POTRA (Fig 2.14b). Overall, the same resonances of POTRA are affected upon interaction with SurAΔP2 but the CSP are smaller compared to the values obtained for wild type. This indicates a lower affinity between the two proteins, confirmed by the higher interaction constant (K_D), in the high micromolar range, measured by SPR (Fig. 2.11 and Table 2.1). Interestingly, the resonances constituting the hydrophobic core of POTRA1, disappearing in the wild type POTRA1 spectra upon titration of SurAΔP2, display here the largest chemical shift change (Fig. 2.14c). Furthermore, no "unfolding" effect was observed upon addition of SurA to a maximum of 8 molar equivalents, indicating that the residues from the hydrophobic core are affected by the presence of SurA but do not undergo conformational exchange in the intermediate range of the NMR timescale.

To confirm these observations, 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of [*U*-^{15}N,^{2}H]–SurAΔP2 in complex with POTRA1R64A at a 1:0.5 to 1:2 SurA:POTRA molar ratio were measured. As observed on POTRA side, the same residues are affected upon binding, indicating that the binding interface is not affected. Similarly, the chemical shift perturbations measured are smaller than the ones for wild-type POTRA indicating a lower affinity in the absence of arginine 64 (data not shown). Altogether, this indicates that R64 is not essential for the interaction between the two partners and that its deletion does not affect the binding interface but that it substantially contributes to the affinity between the two partners as well as in the mechanism of OMP delivery to the first domain of POTRA.

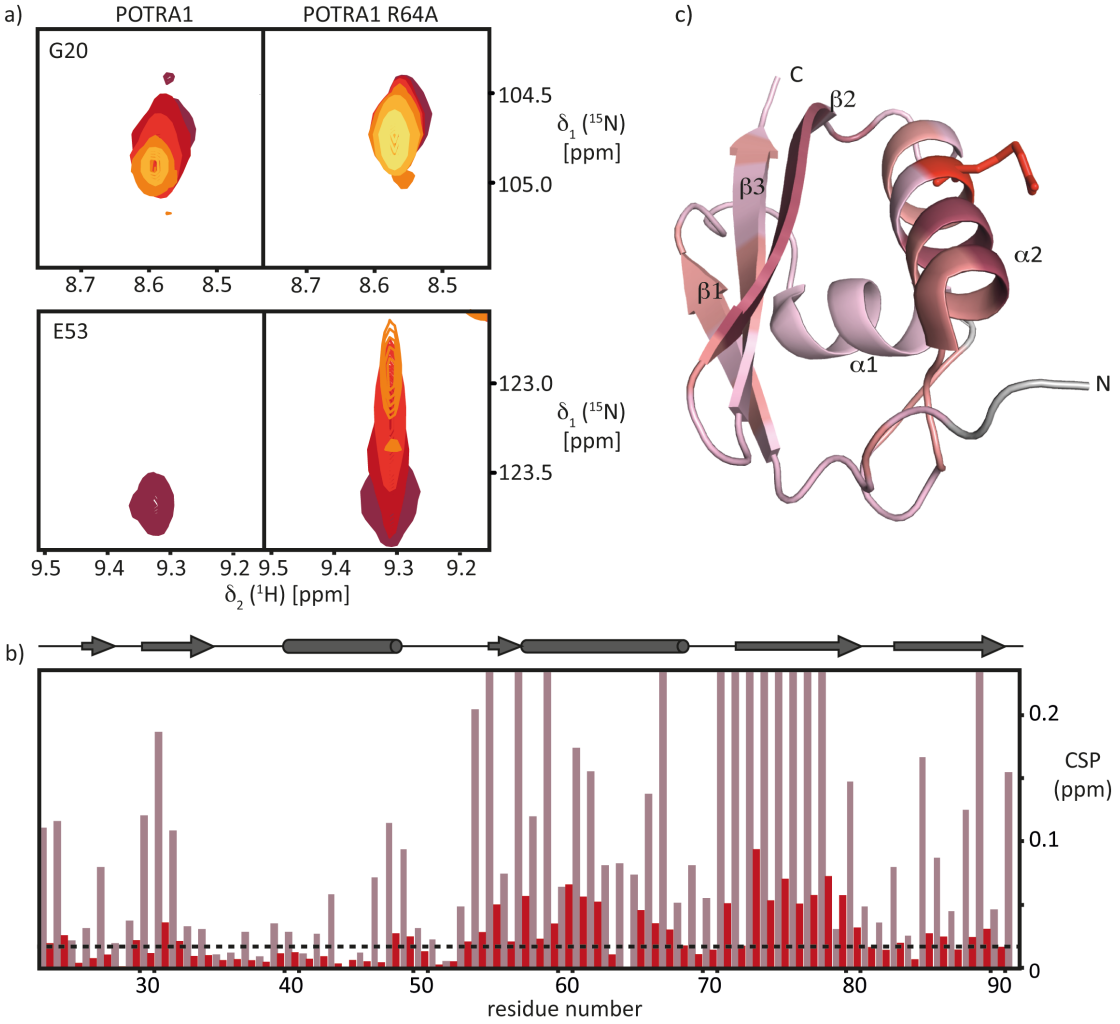


Figure 2.14. Role of arginine 64 in the interaction. a) Comparison of resonances of the residues G20 (top) and E53 (bottom) from 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of 270 μM [*U* -^{15}N]–POTRA1 (left) and [*U* -^{15}N]–POTRA1R64A (right) in NMR buffer (purple) and upon addition of 0.5 (dark red), 1 (red), 2 (dark orange), 4 (orange), and 8 (yellow) molar equivalent of [*U*-^{2}H]–SurAΔP2. b) CSP values of POTRA1 residues (red) and POTRA1R64A residues (light pink), in presence of 1 molar equivalent of SurAΔP2 plotted *vs.* the amino acid residue number of POTRA12. The secondary structure elements of the first domain of POTRA are indicated above. c) Ribbon representation of POTRA1R64A (PDB: 2QDF) where changes upon addition of 1 molar equivalent of SurAΔP2 are indicated by a pink gradient for CSP values between 0 and 0.1 ppm. The position of the mutated arginine to alanine is indicated in red.

Discussion

It has been shown that the interaction between POTRA and SurA in *E. coli* is a crucial step of the OMP refolding pathway, transferring the unfolded substrate from the chaperone to the POTRA domains of BamA for its fold into the outer membrane[15, 16]. Using NMR spectroscopy, we could reproduce the interaction between the two partners *in vitro*. The N- and C-terminal domains of

SurA are required for the interaction, while the two peptidyl-prolyl isomerase domains are not directly contributing. The role of these two domains is still unclear, the first PPIase domain exhibiting no PPIase activity and adopting a distorted fold compared to other PPIase domains, could be involved in substrate binding while the second one, exhibiting PPIase activity, could act as a second binding site for the peptide or exhibit PPIase activity[5, 7]. Preliminary results localized the interaction of POTRA on the N- and C-terminal β-strands of SurA. This β-sheet was suggested to act as a structural support of the protein[21] and the interaction of POTRA can act as a release mechanism for the OMP upon destabilization, in absence of external energy. Furthermore, we could show that SurA interacts with the first POTRA domain only. The residues affected by this interaction are located mainly on the α2 helix and the β1, 2 and 3 strands of POTRA1, consistent with previous observations showing that SurA is interacting with the α2 helix of POTRA 1[16]. A dynamic behavior of POTRA1 was also observed upon SurA interaction, where the hydrophobic core of the first POTRA domain formed between the residues of the α2 helix and the β2 strand seems to undergo conformational exchange between a bound state, forming hydrophobic interactions and a free state, at an intermediate rate in the NMR time scale. Finally, we could also show that in the absence of arginine 64, located in the α2 helix of POTRA 1, the affinity between the two partners is reduced substantially. Furthermore, the hydrophobic core is affected but did not undergo conformational exchange. This indicates that the arginine, albeit not the sole mediator of this interaction on POTRA side, plays an important role in the interaction and might in addition be crucial for the mechanism of release of the OMP from SurA to POTRA where the conformational exchange of the hydrophobic core could act as a receptor for the nascent OMP.

Our findings led us to propose a mechanism, depicted in Figure 2.15, for the delivery of the OMP to the POTRA domains of BamA for its fold into the outer membrane. There, SurA carrying the unfolded OMP interacts *via* its N- and C-terminal domains with the α2 helix of the first POTRA domain (i). Upon interaction, SurA docks to arginine 64, located in the α2 helix and this specific binding induces a conformational exchange of the hydrophobic core of POTRA1, where the hydrophobic residues swap between a favorable and stable bound state, and a free state where the hydrophobic moieties are exposed (ii). SurA destabilization, through interaction affecting its terminal β-sheet, leads to the release of the Omp that interacts with the exposed hydrophobic residues of POTRA (iii). The β-signal sequence of the OMP necessary for its folding into the outer membrane, consists of several hydrophobic residues and might be the interacting partner of the hydrophobic residues of POTRA. To reach the BamA barrel, the unfolded OMP slides along the

POTRA domains where the hydrophobic attraction between α2 helix and β2 strand will "push" the unfolded Omp to the next POTRA domain (iv). It was proposed earlier that the OMP was binding and subsequently sliding along the POTRA *via* β-augmentation[13]. This mechanism would not be in agreement with our proposed sliding mechanism on the hydrophobic residues, keeping the OMP unfolded until it reaches the β-barrel domain of BamA. To date, we do not have direct evidence to further prove this proposed mechanism. Mutations in the hydrophobic core of POTRA need to be introduced and the viability of these different constructs in the presence of the OMP need to be tested. In addition, measurements of the exposition of the hydrophobic residues to water, in presence of SurA could give hints on the possibility of this mechanism, presumably crucial in the OMP biogenesis.

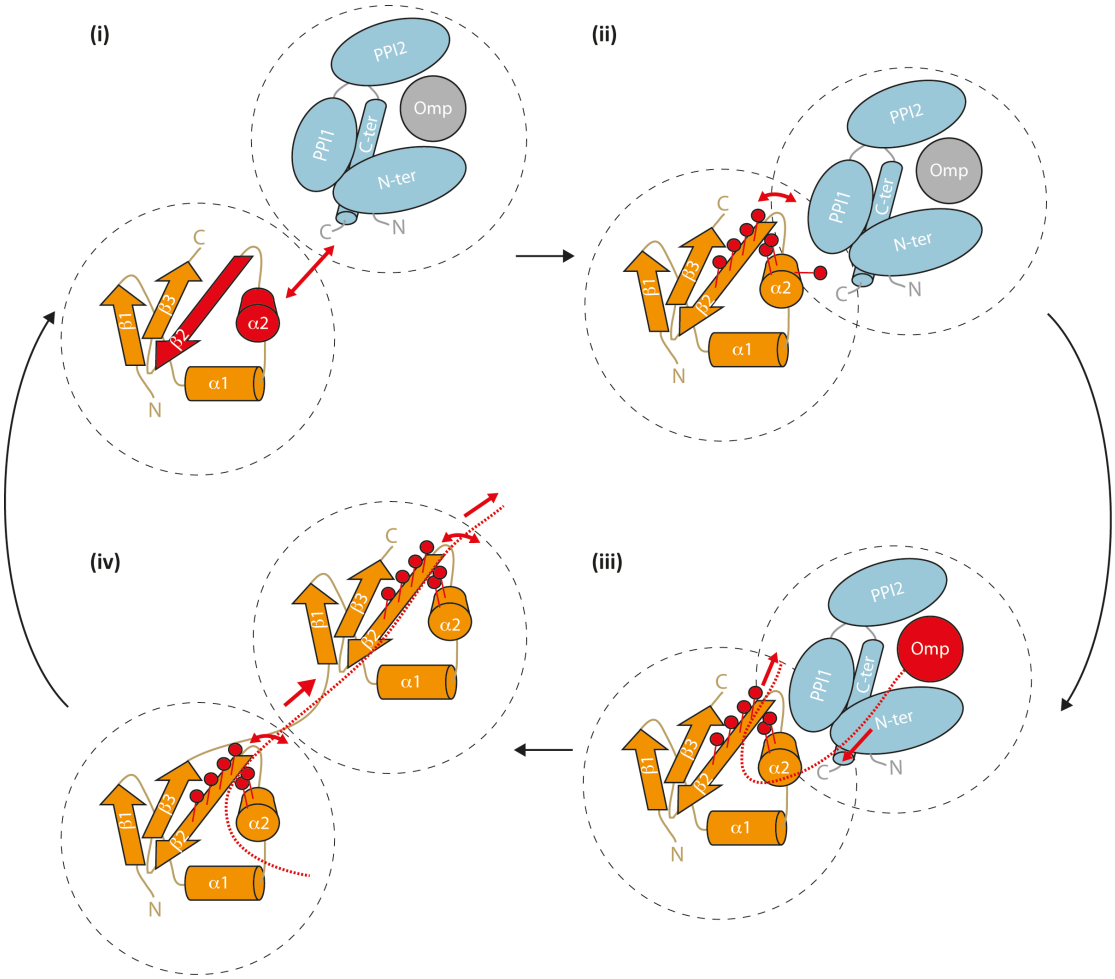


Figure 2.15. Proposed mechanism for the delivery of the unfolded OMP to the first POTRA domain of BamA. The elements involved in the mechanism at each steps are highlighted in red, SurA is depicted in blue and POTRA in orange. See text for details.

Material and Methods

Protein biochemistry

SurA constructs were cloned in pET28b vector (Novagen), containing a thrombin-cleavable N-terminal His_6-tag. *E. coli* BL21 (λ DE3) Lemo cells were transformed with the respective plasmid and grown at 37°C in medium containing 30 μg/mL kanamycin to an OD_{600} ~ 0.6 and then for an additional 30 min at 25°C. Expression was induced by 0.4 mM IPTG. Cells were harvested 18 – 20 h after induction and resuspended in buffer containing 25 mM HEPES, 300 mM NaCl and 10 mM Imidazol at pH 7.5 at a 4:1 buffer/pellet weight ratio and lyzed by two passes through a French press. The lysate was centrifuged for 1 h at 34,500 g at 4°C, subsequently applied to a 5 ml Ni^{2+}-HisTrap (GE Healthcare) column and eluted by an imidazole gradient. SurA elutes at 150 mM imidazol concentration. The elution fractions containing SurA were dialyzed overnight against Buffer A (25 mM Hepes, 150 mM NaCl, pH 7.5) at 4°C. Dialyzed SurA was denatured with 6 M Gdm/HCl, applied to Ni^{2+} beads and eluted with 200 mM imidazol. The eluted SurA was dialyzed overnight against Buffer A (25 mM HEPES, 150 mM NaCl, pH 7.5) at 4°C. Dialyzed SurA was concentrated in a Vivaspin concentrator MWCO 30,000 (Satorius) and applied to a HiLoad 16/600Superdex 75 or 200 pg (GE Healthcare). Afterwards, eluted fractions containing SurA were concentrated by ultrafiltration and stored at -80°C until use.

POTRA constructs were cloned in pET28b vector (Novagen), containing a thrombin-cleavable N-terminal His_6-tag. Expression and purification were done as described for SurA above with the change of purification under denaturing condition step to purification *via* anion exchange affinity with a 5 ml Q Hi-Trap FF (GE Healthcare) column and eluted by NaCl gradient. POTRA constructs elute in the flow through (FT).

The quick change II mutagenesis protocol (Stratagene) was used to clone the SurA-NC construct and to introduce the mutation R64A into POTRA1. PCR primers were ordered from Microsynth. The expression and purification of these two mutants were performed as their respective protocols described above.

Isotope labeling

[*U*-^{2}H], [*U*-^{2}H,^{15}N], and [*U*-^{2}H,^{15}N,^{13}C]-labeled SurA constructs were obtained by growing the expression cells in M9 minimal media[23] supplemented with ($^{15}NH_4$)Cl, [^{2}H,^{13}C]-glucose, [^{2}H,^{12}C]-glucose, and D_2O. [*U*-^{15}N]-labeled POTRA constructs were obtained by growing the expression

cells in M9 minimal media[23] supplemented with ($^{15}NH_4$)Cl. All isotopes were purchased from Sigma-Aldrich or Cambridge Isotope Labs.

NMR spectroscopy

NMR spectra were recorded at 37°C for SurA, POTRA and SurA–POTRA complexes on Bruker AscendII- 700 and Bruker Avance-900 spectrometer equipped with cryogenic triple-resonance probes. The 2D [^{15}N,^{1}H]-TROSY-HSQC of SurA and POTRA were recorded in a total experiment time of 2 h. The ^{1}H carrier was centered on the water resonance, the ^{15}N carrier at 119 ppm. The interscan delay was set to 1 s. In the direct dimension, 1024 complex points were recorded in an acquisition time of 91 ms, multiplied with a 75°-shifted sine bell, zero-filled to 2048 points and Fourier transformed. In the indirect dimension, 100 complex points were measured with a maximal evolution time of 4.44 ms, multiplied with a 75°-shifted sine bell, zero-filled to 256 points and Fourier transformed. For the SurA-POTRA titrations series, 2D [^{15}N,^{1}H]-TROSY-HSQC spectra were recorded with 0.5, 1, 2, 4 and 8 molar equivalents of the titrated protein. Chemical shift perturbations (CSP) of amide moieties were calculated as:

$$CSP = \sqrt{\left(\delta H_{ref} - \delta H\right)^2 + \left(\frac{\delta N_{ref} - \delta N}{5}\right)^2} \quad (1)$$

For the sequence-specific backbone assignment of [*U*-^{2}H,^{15}N,^{13}C]-labeled SurAΔP2, the following NMR spectra were recorded: 3D TROSY-HNCA, 3D TROSY-HNCACB[17], 3D TROSY-HNCO[18], 3D TROSY-HNCACO[19] and 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY[20] with a NOE mixing time of 200 ms. The 3D [^{1}H,^{1}H]-NOESY-^{15}N-TROSY spectrum was recorded using non-uniform sampling[24] (NUS) with 36 % of sparse sampling and 3456 hypercomplex points in both indirect dimensions. Spectrum reconstruction was achieved using MddNMR program[25]. For all spectra polynomial baseline correction was applied in all dimensions. NMR data were processed using PROSA[26], NMRPipe[27], and analyzed with CARA (cara.nmr.ch) and XEASY[28].

^{1}H projections of 2D [^{15}N,^{1}H]-TROSY-HSQC spectra of SurAΔP2 were used to derive the dimerization constant (K_D) of the protein, assuming a monomer-dimer equilibrium, where the monomeric molar fraction χ_m is given by Eq. 2 with C_{tot}, the total protein concentration[29].

$$\chi_m = \left[\sqrt{(8 \cdot C_{tot} \cdot K_D + K_D^2)} - K_D\right] / (4 \cdot C_{tot}) \quad (2)$$

The experimental values of χ_m were calculated as the ratio between the integrals at 8.5 to 10

and 7.5 to 8.5 ppm of the ^{1}H projections.

Circular dichroism spectroscopy

Circular dichroism (CD) data were recorded on a Chirascan spectrometer, using a 1 mm pathlength cuvette, in CD buffer (100 mM NaPi, pH 6.5) using the following protein concentration: POTRA1 at 12 μM, POTRA2 at 5 μM, POTRA12 at 3 μM. Melting curves were recorded form 20 to 98°C with an increase of 1°C / point and an averaging of 3 second per point. Data points were collected at 222 nm at 0.2 °C interval. The measured ellipticity θ was normalized into $[\theta]$ (deg.cm^2.dmol^{-1}) using the following equation where n is the number of peptide bonds in the protein, L the cuvette pathlength (mm) and C the protein concentration (μM):

$$[\theta] = \frac{\theta \cdot 10^6}{L \cdot n \cdot C} \qquad (3)$$

Curves were fitted using the following formula where N_{int} and N_{slope} were the slope and the intersection a $[\theta]$ = 0 of the native baseline and D_{int} and D_{slope} were the slope and the intersection a $[\theta]$ = 0 of the denatured baseline. ΔC_p was estimated *via* the accessible surface area (ASA) where $\Delta ASA = -907 + 93 \cdot n$ and $\Delta C_p = -251 + 0.19 \cdot \Delta ASA$ [30].

$$[\theta] = \frac{(N_{int}+N_{slope}\cdot T)+(D_{int}+D_{slope}\cdot T).e^{-\left(\frac{\Delta G}{RT}\right)}}{1+e^{-\left(\frac{\Delta G}{RT}\right)}} \qquad (4)$$

$$\Delta G = \Delta H_m \left(1 - \frac{T}{T_m}\right) + \Delta C_p \left[T - T_m - \left(T \cdot ln\frac{T}{T_m}\right)\right] \qquad (5)$$

Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were recorded on a Biacore T100 instrument (GE healthcare). Experiments were performed in SPR buffer (25 mM MES, 150 mM NaCl, pH6.5) at 25 °C. The ligands (POTRA constructs) (2 mg / mL) were immobilized by amine coupling to a CM5 chip surface (GE healthcare) at a flow rate of 2 μl/min and a contact time of 1200 second. The surface of the reference channel was inactivated by amine coupling of BSA (2 mg / mL).

The interaction of POTRA constructs with SurAfl and SurADP2 were measured using kinetic/affinity method from Biacore. SurA at increasing concentrations (0, 0.25, 0.5, 1, 2, 4 and 10 mg / ml) was loaded to the chip at a flow of 10 μL/min and a contact time of 600 s followed

by a dissociation step of 1200 s. Regeneration of the chip was achieved with 4 M Gdm/HCl solution at a flow of 10 μL/min and a contact time of 100 second followed by a regeneration step of 600 s with SPR buffer. SPR data were analyzed using the Biacore T100 evaluation software.

Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) was used to measure the size of proteins in solution using a labeling dye. SurAfl was labeled using a DyLight 488 NHS Ester (Thermoscientific). In brief, 500 μL of SurAfl at a concentration of 50 μM is mixed with 50 μg of DyLight 488 NHS Ester at incubated for 1 hour protected from light. To remove the excess of dye, the protein was then dialyzed four times against 1 L of AUC buffer (25 mM NaPi, 150 mM NaCl, pH 6.5). The labeled protein at a fixed concentration of 100 nM was mixed with SurAfl at concentrations of 15.7, 31.3, 62.5, 125, 250 and 500 μM in a 200 μL volume. The samples were loaded in AUC rotor cuvettes and measured under vaccum at 25°C. The sedimentation coefficient distribution c(s) was calculated for each concentration using Sedfit and the average size of the molecules in solution was determined for each concentration using the fitted frictional ratio (f/f_0).

The weight averaged s-values (s_W) were calculated from the c(s) distribution for each concentration using the program GUSSI and its distribution was fitted using a monomer-dimer model[29] (Eq. 6) where the sedimentation coefficient of the dimer (S_{dim}) was fixed to 6.2 s and the sedimentation coefficient of the monomer (S_{mono}) was floating.

$$s_W = S_{dim} + (S_{mono} - S_{dim}) \cdot \left[\left(-K_D + \sqrt{(K_D^2 + 8 \cdot C_{tot} \cdot K_D)}\right)/(4 \cdot C_{tot})\right] \quad (6)$$

References

[1] D. Missiakas, J. M. Betton, S. Raina, *Mol. Microbiol.* **1996**, *21*, 871–884.

[2] P. E. Rouviere, C. A. Gross, *Genes Dev.* **1996**, *10*, 3170–3182.

[3] A. Tormo, M. Almiron, R. Kolter, *J. Bacteriol.* **1990**, *172*, 4339–4347.

[4] S. Behrens, R. Maier, H. de Cock, F. X. Schmid, C. A. Gross, *EMBO J.* **2001**, *20*, 285–294.

[5] E. Bitto, D. B. McKay, *Structure* **2002**, *10*, 1489–1498.

[6] P. C. Stirling, S. F. Bakhoum, A. B. Feigl, M. R. Leroux, *Nat. Struct. Mol. Biol.* **2006**, *13*, 865–870.

[7] X. Xu, S. Wang, Y. X. Hu, D. B. McKay, *J. Mol. Biol.* **2007**, *373*, 367–381.

[8] E. Bitto, D. B. McKay, *J. Biol. Chem.* **2003**, *278*, 49316–49322.

[9] G. Hennecke, J. Nolte, R. Volkmer-Engert, J. Schneider-Mergener, S. Behrens, *J. Biol. Chem.* **2005**, *280*, 23540–23548.

[10] J. G. Sklar, T. Wu, D. Kahne, T. J. Silhavy, *Genes Dev.* **2007**, *21*, 2473–2484.

[11] L. Sanchez-Pulido, D. Devos, S. Genevrois, M. Vicente, A. Valencia, *Trends Biochem. Sci.* **2003**, *28*, 523–526.

[12] S. Kim, J. C. Malinverni, P. Sliz, T. J. Silhavy, S. C. Harrison, D. Kahne, *Science* **2007**, *317*, 961–964.

[13] P. Z. Gatzeva-Topalova, L. R. Warner, A. Pardi, M. C. Sousa, *Structure* **2010**, *18*, 1492–1501.

[14] T. J. Knowles, M. Jeeves, S. Bobat, F. Dancea, D. McClelland, T. Palmer, M. Overduin, I. R. Henderson, *Mol. Microbiol.* **2008**, *68*, 1216–1227.

[15] D. Bennion, E. S. Charlson, E. Coon, R. Misra, *Mol. Microbiol.* **2010**, *77*, 1153–1171.

[16] P. Workman, K. Heide, N. Giuliano, N. Lee, J. Mar, P. Vuong, D. Bennion, R. Misra, *J. Bacteriol.* **2012**, *194*, 3512–3521.

[17] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13585–13590.

[18] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich, *J. Am. Chem. Soc.* **1999**, *121*, 844–848.

[19] M. Salzmann, G. Wider, K. Pervushin, K. Wüthrich, *J. Biomol. NMR* **1999**, *15*, 181–184.

[20] E. R. P. Zuiderweg, S. W. Fesik, *Biochemistry* **1989**, *28*, 2387–2391.

[21] Q. Chai, B. Ferrell, M. Zhong, X. Zhang, C. Ye, Y. Wei, *PEDS* **2014**, *27*, 111–116.

[22] Biacore AB, *Kinetic and affinity analysis using BIA - Level 1*, **1997**.

[23] J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory, **1989**.

[24] Y. Manassen, G. Navon, *J. Magn. Reson.* **1988**, *79*, 291–298.

[25] V. Y. Orekhov, V. A. Jaravine, *Prog. NMR Spect.* **2011**, *59*, 271–292.

[26] P. Güntert, V. Dötsch, G. Wider, K. Wüthrich, *J. Biomol. NMR* **1992**, *2*, 619–629.

[27] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.

[28] C. Bartels, T. H. Xia, M. Billeter, P. Güntert, K. Wüthrich, *J. Biomol. NMR* **1995**, *6*, 1–10.

[29] C. T. Benfield, D. S. Mansur, L. E. McCoy, B. J. Ferguson, M. W. Bahar, A. P. Oldring, J. M. Grimes, D. I. Stuart, S. C. Graham, G. L. Smith, *J. Biol. Chem.* **2011**, *286*, 20727-20735.

[30] J. K. Myers, C. N. Pace, J. M. Scholtz, *Protein Sci.* **1995**, *4*, 2138–2148.

Chapter 3

NMR Spectroscopy as a Tool to Study Metabolic Pathways in the Cell

3.1 Introduction

Cellular metabolism is the sum of all the biochemical reactions taking place within a cell, including degradation of nutrients, synthesis of macromolecules or small precursors, such as amino acids or electron transfer[1]. Metabolism takes place in sequences of biochemical reactions called pathways that can be linear, extensively branched or cyclic. Pathways can have as precursor an earlier reaction[2] and can have different functions. The glycolytic pathway is a very important metabolic pathway. It provides energy to the cell in the form of ATP (adenosine triphosphate) and NADH (reduced nicotinamide adenine dinucleotide) by converting glucose into pyruvate (Fig. 3.1). The pathway occurs in nearly all organisms, both aerobic, consuming oxygen, and anaerobic, growing in the absence of oxygen. Pyruvate, the end product of the glycolytic pathway, is a branching point in the cell metabolism and can undergo diverse chemical reactions. In aerobic organisms, it is the precursor of acetyl-CoA in the citric acid cycle (tricarboxylic acid (TCA) cycle or Krebs cycle), providing energy to the cell in the form of ATP[3, 4] while in anaerobic organisms it is converted into lactate and ethanol[5] (Fig. 3.1). In both organisms it is also used to produce alanine via a transamination process. All metabolic reactions are catalyzed by enzymes, up or down regulating the different pathways through their sensitivity to external factors[6] and can be classified in two broad categories, catabolic and anabolic reactions, where the energy produced in the form of ATP during catabolic pathways by degrading molecules is used to drive anabolic processes[2].

In the cells, changes in metabolism upon infection is a critical feature that has to be characterized for most diseases treatment. NMR spectroscopy was shown to be a method of choice to study these changes in the cell[7]. It allows non-invasive monitoring of a wide range of processes, without introducing any chemical or mechanical perturbations into the system. NMR Studies of cell metabolism typically use ^{31}P, ^{13}C, ^{1}H and ^{15}N for their abundance in the biological compounds. As an example, the naturally abundant ^{31}P was used to assess the energy status of a cell by monitoring the intensity of resonances form adenosine triphosphate (ATP)[8] as well as to measure the precise intracllular pH[9].

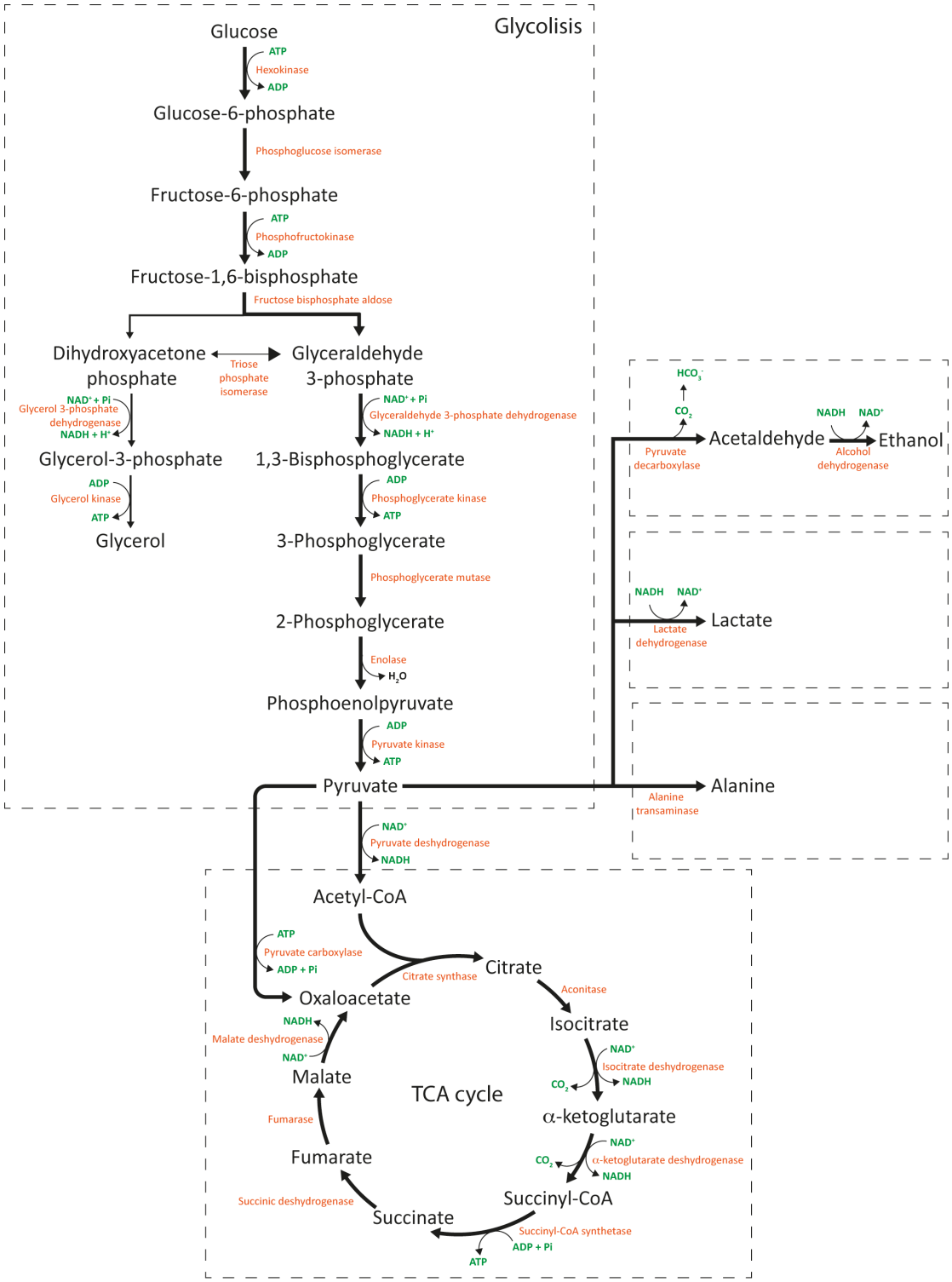


Figure 3.1 Metabolic pathways in the cell where the glycolytic pathway and the citric acid cycle (tricarboxylic acid (TCA) cycle or Krebs cycle are depicted as well as metabolites products from pyruvate. Enzymes, catalyzing the reactions, are indicated in orange.

In metabolomics studies, ^{13}C-NMR is broadly used to characterize various cell pathways, as carbons are found in most biological compounds. However, since the ^{13}C natural abundance is only 1.1 percent, ^{13}C labeling is typically required. Thereby, the redistribution of a ^{13}C-labeled nutrient among cellular metabolites is monitored using ^{13}C-NMR experiments, providing information on the subsequently labeled compounds, their labeling position within the molecule as well as their concentration[10, 11]. Due to its low gyromagnetic ratio, detection of ^{13}C labels is often limited to highly abundant compounds, but this issue can be overcome by its indirect observation via the more sensitive ^{1}H nuclei, recording ^{1}H-^{13}C correlation or X-filtered experiments[12]. There, 1D ^{12}C-filtered and 1D ^{13}C-filtered experiments allow the detection of protons bound to ^{12}C and ^{13}C nuclei respectively, providing quantitative information on fractional labeling of molecules. 2D correlation NMR experiments were also used to characterize metabolites, improving the spectral resolution by editing a second dimension[13, 14]. These experiments are however time-consuming since data acquisition of the second dimension requires longer measurement times. These long times could lead to modification of cell metabolism in the NMR tube as well as to the detection of molecules originating from an average metabolism that does not corresponds to a specific physiological state of the cell. However, this issue can be partially overcome by the use of fast acquisition spectroscopy[15], which decreases the measurement time to few seconds[16]. NMR-based metabolism studies were broadly performed on intact cells[17-20] allowing observation of active metabolite processes and intracellular milieu but, due to the relative insensitivity of the method, require a large number of cells for quantitative measurements. Thus, metabolomics by NMR is mainly limited to cell lines that can be grown in culture condition, such as cancer cells even though NMR-based metabolic studies of other systems have been performed, including macrophages[21, 22]. The technique thus also suffers from cell packing, lack of oxygen, accumulation of waste products or pH changes in the NMR tube that can modify the metabolism. In parallel, metabolite studies of extracted intracellular media[19, 23, 24] and extracellular media (metabolic footprinting)[25, 26] is a well-established procedure taking advantage of better resolution, higher sensitivity and possible addition of reference compounds for unambiguous assignments. However, the destruction of the cell sample and the impossibility to monitor continuous metabolite processes can lead to erroneous results and imply the collection of numerous samples at different time points to cover a complete metabolic process.

Despite significant technological and biochemical advancement, the application of NMR spectroscopy to measure metabolic processes is limited by the intrinsically low sensitivity of the method. To overcome this issue, various techniques have been proposed to enhance the nuclear

spin polarization among which optical pumping[27] or dynamic nuclear polarization (DNP)[28]. These hyperpolarization methods have the potential to create a polarization close to unity but suffer form intrinsic physical limitations. Recently, dissolution dynamic nuclear polarization (d-DNP) NMR spectroscopy has been developed allowing real time observation of living cells with high sensitivity by enabling close to unit polarization of organic molecules in a liquid solution[29] and was shown to be a method of choice to characterize metabolic pathways in living *S. cerevisiae*[30], *E. coli*[31, 32] or breast cancer cells[33].

In this chapter, standard NMR spectroscopy was used to characterize changes in metabolism of cells upon infection by the *S. flexneri* pathogen (section 3.2), providing new insights into this highly infectious process. In section 3.3, d-DNP NMR spectroscopy was applied for the first time to macrophages, where metabolic pathways could be observed in real time, demonstrating the possible application of this technique to a large range of systems as well as its major upcoming contribution to the metabolic field.

References

[1] P. C. Hinkle, R. E. Mc Carty, *Sci. Am.* **1978**, *238*, 104–117.

[2] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry, 5th edition*, New York: W. H. Freeman and Company, **2002**.

[3] H. A. Krebs, P. D. J. Weitzman, *Krebs' citric acid cycle: half a century and still turning*, London: Biochemical Society, **1987**.

[4] J. M. Lowenstein, *Methods in Enzymology, Volume 13: Citric Acid Cycle*, Boston: Academic Press, **1969**.

[5] G. Tortora, B. Funke, C. Case, *Microbiology: An Introduction. 5th ed*, Menlo Park, CA: Bejamin/Cummings, **1995**.

[6] F. C. Neidhardt, J. L. Ingraham, M. Schaechter, *Physiology of the bacterial cell. A molecular approach*, Sinauer associates, Inc., Sunderland, MA, **1990**.

[7] R. J. Gillies, *NMR In Physiology and Biomedicine*, Academic Press, Inc., **1994**.

[8] K. Ugurbil, D. L. Guernsey, T. R. Brown, P. Glynn, N. Tobkes, I. S. Edelman, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 4843–4847.

[9] B. S. Szwergold, T. R. Brown, J. J. Freed, *J. Cell. Physio.* **1989**, *138*, 227–235.

[10] M. F. Chauvin, F. Megnin-Chanet, G. Martin, J. M. Lhoste, G. Baverel, *J. Biol. Chem.* **1994**, *269*, 26025–26033.

[11] A. W. Jans, D. Leibfritz, *NMR Biomed.* **1989**, *1*, 171–176.

[12] G. Otting, K. Wüthrich, *Q. Rev. Biophys.* **1990**, *23*, 39–96.

[13] J. W. Welch, K. Bhakoo, R. M. Dixon, P. Styles, N. R. Sibson, A. M. Blamire, *NMR Biomed.* **2003**, *16*, 47–54.

[14] H. Wen, Y. J. An, W. J. Xu, K. W. Kang, S. Park, *Angew. Chem. Int. Ed.* **2015**, *54*, 5374–5377.

[15] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014–8015.

[16] A. Motta, D. Paris, D. Melck, *Anal. Chem.* **2010**, *82*, 2405–2411.

[17] J. A. den Hollander, T. R. Brown, K. Ugurbil, R. G. Shulman, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 6096–6100.

[18] R. B. Moon, J. H. Richards, *J. Biol. Chem.* **1973**, *25*, 7276–7278.

[19] G. Navon, S. Ogawa, R. G. Shulman, T. Yamane, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 888–891.

[20] K. Ugurbil, T. R. Brown, J. A. den Hollander, P. Glynn, R. G. Shulman, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 3742–3746.

[21] C. Ducrocq, M. Lenfant, G. H. Werner, B. Gillet, J. C. Beloeil, *Biochem. Biophys. Res. Commun.* **1987**, *147*, 519–525.

[22] F. Seguin, A. Le Pape, *J. Immunol. Methods* **1994**, *4*, 179–187.

[23] S. J. Berners-Price, M. E. Sant, R. I. Christopherson, P. W. Kuchel, *Magn. Reson. Med.* **1991**, *18*, 142–158.

[24] U. Sonnewald, N. Westergaard, P. Jones, A. Taylor, H. S. Bachelard, A. Schousboe, *J. Neurochem.* **1996**, *67*, 2566–2572.

[25] J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, D. B. Kell, *Nat. Biotechnol.* **2003**, *21*, 692–696.

[26] D. B. Kell, M. Brown, H. M. Davey, W. B. Dunn, I. Spasic, S. G. Oliver, *Nat. Rev. Microbiol.* **2005**, *3*, 557–565.

[27] B. M. Goodson, *J. Magn. Reson.* **2002**, *155*, 157–216.

[28] V. S. Bajaj, C. T. Farrar, M. K. Hornstein, I. Mastovsky, J. Vieregg, J. Bryant, B. Elena, K. E. Kreischer, R. J. Temkin, R. G. Griffin, *J. Magn. Reson.* **2003**, *160*, 85–90.

[29] J. H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning, K. Golman, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10158–10163.

[30] S. Meier, M. Karlsson, P. R. Jensen, M. H. Lerche, J. O. Duus, *Mol. BioSyst.* **2011**, *7*, 2834–2836.

[31] S. Meier, P. R. Jensen, J. O. Duus, *FEBS Lett.* **2011**, *585*, 3133–3138.

[32] S. Meier, P. R. Jensen, J. O. Duus, *ChemBioChem* **2012**, *13*, 308–310.

[33] T. Harris, H. Degani, L. Frydman, *NMR Biomed.* **2013**, *26*, 1831–1843.

3.2 *Shigella* Reroutes Host Cell Central Metabolism to Obtain High-flux Nutrient Supply for Vigorous Intracellular Growth

Original publication

David Kentner, Giuseppe Martano, Morgane Callon, Petra Chiquet, Maj Brodmann, Olga Burton, Asa Wahlander, Paolo Nanni, Nathanaël Delmotte, Jonas Grossmann, Julien Limenitakis, Ralph Schlapbach, Patrick Kiefer, Julia A. Vorholt, Sebastian Hiller, and Dirk Bumann

***Shigella* reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth**

Proc. Natl. Acad. Sci., **2014**, *111*, 9929–9934

David Kentner, Giuseppe Martano and Morgane Callon contributed equally to this work. The research was conducted by David Kentner, Patrick Kiefer, Julia A. Vorholt, Sebastian Hiller, and Dirk Bumann. The NMR experiments and their analysis were performed by Morgane Callon.

Shigella reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth

David Kentner[a,1,2], Giuseppe Martano[b,1], Morgane Callon[a,1], Petra Chiquet[a], Maj Brodmann[a], Olga Burton[a], Asa Wahlander[c], Paolo Nanni[c], Nathanaël Delmotte[b], Jonas Grossmann[c], Julien Limenitakis[a], Ralph Schlapbach[c], Patrick Kiefer[b], Julia A. Vorholt[b], Sebastian Hiller[a], and Dirk Bumann[a,2]

[a]Biozentrum, University of Basel, 4056 Basel, Switzerland; [b]Institute of Microbiology, Eidgenössiche Technische Hochschule Zürich, 8093 Zürich, Switzerland; and [c]Functional Genomics Center Zurich, 8057 Zürich, Switzerland

Edited by Philippe J. Sansonetti, Institut Pasteur, Paris, France, and approved May 28, 2014 (received for review April 16, 2014)

Shigella flexneri **proliferate in infected human epithelial cells at exceptionally high rates. This vigorous growth has important consequences for rapid progression to life-threatening bloody diarrhea, but the underlying metabolic mechanisms remain poorly understood. Here, we used metabolomics, proteomics, and genetic experiments to determine host and *Shigella* metabolism during infection in a cell culture model. The data suggest that infected host cells maintain largely normal fluxes through glycolytic pathways, but the entire output of these pathways is captured by *Shigella*, most likely in the form of pyruvate. This striking strategy provides *Shigella* with an abundant favorable energy source, while preserving host cell ATP generation, energy charge maintenance, and survival, despite ongoing vigorous exploitation. *Shigella* uses a simple three-step pathway to metabolize pyruvate at high rates with acetate as an excreted waste product. The crucial role of this pathway for *Shigella* intracellular growth suggests targets for antimicrobial chemotherapy of this devastating disease.**

infectious diseases | host–pathogen interactions

Infectious diseases typically arise when pathogens grow to high tissue loads, causing extensive damage and immunopathology. An outstanding example is *Shigella flexneri*, which rapidly grow from a small infectious dose of 10–100 bacteria (1) to intestinal loads causing life-threatening bloody diarrhea (bacillary dysentery) within a few hours (2, 3). This vigorous *Shigella* growth occurs inside human colon epithelial cells and requires an integrated *Shigella* pathogenesis program, including a type three secretion system encoded on the *Shigella* virulence plasmid. Using this system, *Shigella* translocates enzymes into the host cell cytosol, where they target key cellular functions, allowing *Shigella* to enter the host cell and escape bacterial killing by innate immune responses (4). After *Shigella* reaches the host cell cytosol, many virulence factors are down-regulated (5), and *Shigella* starts rapid proliferation.

Biomass generation at such high rates depends on extensive exploitation of intracellular host nutrients (6). The host cell cytoplasm contains hundreds of metabolites, but it is unclear which of these potential nutrients *Shigella* uses, how the host cell can supply them at sufficiently high rates to support rapid *Shigella* growth, and why host cells can sustain viability while being vigorously exploited by intracellular *Shigella*. For related enteroinvasive *Escherichia coli*, previous research has shown that glucose and other host metabolites, such as diverse amino acids, can be incorporated into the biomass of these closely related pathogens (7). However, quantitative data are still lacking, and energy production, which is usually a major part of nutrient use (8), could not be analyzed because of technical limitations.

In general, pathogen metabolism has been recognized as a fundamentally important aspect of infectious diseases, but available data are mostly restricted to qualitative presence/absence of enzymes in pathogen genomes and metabolite or gene expression profiles in various infection models (9–13). Comprehensive quantitative studies on pathogen nutrition, metabolism, and growth are largely lacking. This limited knowledge reflects, in part, the fact that suitable methodologies are just becoming available. In this study, we combined various metabolomics approaches, proteomics, and microbial genetics to elucidate the metabolic basis of *Shigella* rapid growth in infected human host cells.

Results

Shigella **Grow Rapidly Inside HeLa Host Cells.** We infected HeLa epithelial cells with *S. flexneri* 2a 2457T *icsA*. The *icsA* mutation prevents spread between host cells (14, 15), thereby simplifying analysis of intracellular growth. In this model, *Shigella* grew rapidly with a generation time of 37 ± 4 min (Fig. S1 and Movie S1), close to maximal axenic growth rates in rich broth and faster than almost all other pathogens in their respective host environments. Infected HeLa cells remained intact until around 3.5–4 h postinfection, when their cytoplasm was packed with more than 100 *Shigella*. The HeLa cells subsequently detached and disintegrated, which was observed previously (16).

Host Central Metabolism Remains Functional During *Shigella* Infection. Rapid intracellular *Shigella* growth likely causes a substantial

Significance

Shigella **causes devastating bloody diarrhea. Rapid disease progression results from exceptionally fast *Shigella* growth inside human gut epithelial cells, but how *Shigella* can obtain nutrients at such high rates from its host cell has been unclear. Here, we show that infected host cells maintain normal central metabolism for energy production and host cell survival. However, *Shigella* captures the entire host metabolism output and degrades it further to acetate. This striking strategy provides *Shigella* with an abundant supply of a favorable energy source, while preserving host cell viability for prolonged exploitation. The crucial role of acetate metabolism for *Shigella* growth suggests potential new targets for antimicrobial chemotherapy.**

Author contributions: D.K., P.K., J.A.V., S.H., and D.B. designed research; D.K., G.M., M.C., P.C., M.B., O.B., A.W., P.N., N.D., and P.K. performed research; J.L. and R.S. contributed new reagents/analytic tools; D.K., G.M., M.C., P.C., M.B., O.B., A.W., P.N., N.D., J.G., P.K., S.H., and D.B. analyzed data; and D.K., J.A.V., S.H., and D.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]D.K., G.M., and M.C. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: david.kentner@gmx.de or dirk.bumann@unibas.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406694111/-/DCSupplemental.

metabolic burden on the infected host cell. Metabolite quantification in infected and uninfected cells identified some metabolites with differential concentrations (Tables S1 and S2), but surprisingly, the energy charge, a measurement of relative ATP, ADP, and AMP levels, did not change significantly on infection (uninfected cells, 0.83 ± 0.03; infected cells, 0.80 ± 0.05). This observation showed that infected cells largely maintain their energy production, despite ongoing exploitation by *Shigella*.

Quantitative metabolite analyses of pathogen-infected host cells encounter the general technical challenge that pathogen and host cell contents are not easily separable within the required time frame. Available separation techniques would inevitably cause a time delay of up to minutes, during which rapid reactions, such as ATP hydrolysis, would substantially alter metabolite concentrations. To avoid such artifacts, we quenched the cell culture immediately and determined combined metabolite concentrations from infected cells and *Shigella*. Because of the much larger volume of human cells compared with *Shigella*, most of these data are probably dominated by host metabolites. As an example, we experimentally determined *Shigella* adenosine phosphate (AXP) contents in various axenic cultures (glucose or pyruvate as sole energy/carbon source provided at 0.1 or 1 g L^{-1}). The results showed that 50 *Shigella* cells contained 0.25–0.69 fmol ATP, 0.19–0.35 fmol ADP, and 0.04–0.05 fmol AMP. Even when subtracting these potential *Shigella* contributions from the combined AXP levels of infected HeLa cells, the HeLa-only AXP values would still yield an energy charge of 0.79 ± 0.02, suggesting a very minor impact of *Shigella* AXP on calculated host cell energy charge values, which was expected based on the different cell volumes of HeLa and *Shigella*.

Various mechanisms could enable infected host cells to maintain their energy charge, despite *Shigella* exploitation. In particular, infected cells might increase nutrient uptake from the extracellular environment (17). However, under the experimental conditions used here, uninfected and infected cells consumed glucose at similar rates (9.0 ± 1.1 vs. 9.3 ± 1.3 fmol/min per cell), whereas uptake of glutamine, another potentially major nutrient for mammalian cells, remained below 0.5 fmol/min per cell.

To determine metabolic fluxes involved in host cell ATP production, we switched unlabeled glucose in the external medium to uniformly labeled (*U*-) ^{13}C glucose and monitored subsequent ^{13}C incorporation in diverse metabolites using mass spectrometry (MS) (Fig. 1*A* and Table S3). Uninfected HeLa cells showed uptake and catabolism of glucose through Embden–Meyerhof and pentose phosphate pathways but very little feeding into tricarboxylic acid (TCA) cycle intermediates, indicating predominant ATP generation through fermentation, which was previously shown for HeLa and other cancer cell lines (18, 19). Interestingly, infected HeLa cells, which contained around 50 rapidly growing *Shigella* at 2.5 h postinfection, had almost

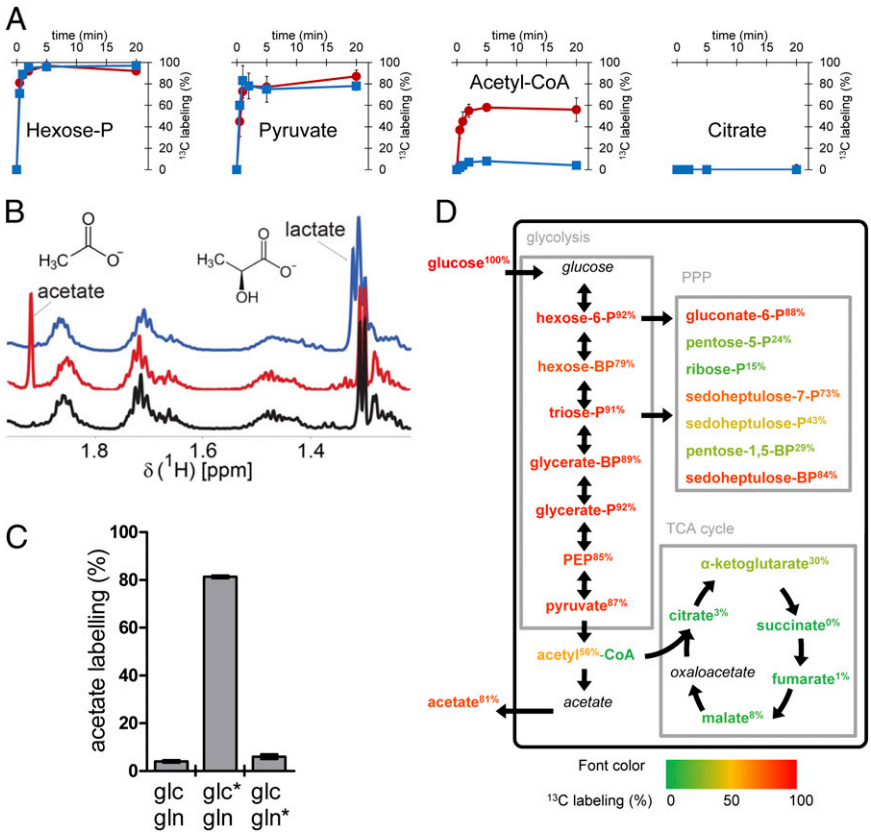


Fig. 1. (*A*) ^{13}C-incorporation kinetics for various metabolites in uninfected (blue) or infected (red) HeLa cells after feeding of *U*-^{13}C glucose. The ^{13}C-labeling percentage is calculated as the fraction of the labeled species (2-^{13}C for acetyl-CoA and *U*-^{13}C for others) relative to the total metabolite pool. Data are represent means and SDs of six independent experiments (Table S3 shows the full dataset). (*B*) Extracellular metabolites before (black) and 3 h after (red) infection as detected by 1D ^{1}H NMR (Fig. S3 shows identification of acetate by 2D [^{13}C,^{1}H] NMR). Uninfected HeLa cells are shown for comparison (blue). The medium contained *U*-^{13}C glucose. (*C*) ^{13}C labeling of extracellular acetate in medium containing *U*-^{13}C–labeled compounds (glc*, glucose; gln*, glutamine) or unlabeled compounds (glc, glucose; gln, glutamine). Data represent means and SDs of three independent experiments. (*D*) Overview of ^{13}C labeling of metabolites from *U*-^{13}C glucose in infected HeLa cells. Font color and superscripts indicate the percentage of the labeled species (2-^{13}C for acetyl-CoA and *U*-^{13}C for others) after 20 min (all intracellular metabolites) or 1 h (acetate) of feeding with ^{13}C glucose. Data represent means from six independent experiments. PPP, pentose phosphate pathway and related metabolites.

identical label incorporation kinetics in Embden–Meyerhof and pentose phosphate pathway intermediates and marginal flux into the TCA cycle. This flux distribution suggested that host cells maintained their energy charge during infection by using the main preexisting energy-producing glycolytic pathways. Proteome comparisons revealed that, in general, host cell enzyme abundance changed little during infection (Fig. S2), arguing against a major reorganization of the host metabolic network during infection. This proteomics result was consistent with previous data on unaltered host cell amino acid biosynthesis during infection with related enteroinvasive *E. coli* (7).

Infection Reroutes Carbon Flux to Acetate Excretion. The apparently minor impact of infection on central host metabolism was surprising, because *Shigella* obviously must redirect major metabolic fluxes and capture abundant metabolites to sustain its rapid proliferation. Indeed, our ^{13}C glucose isotope-tracking experiments showed one remarkable difference between infected and uninfected cells: a substantial fraction (56 ± 11%) of the acetyl moiety in acetyl-CoA was rapidly labeled exclusively in infected cells (Fig. 1*A* and Table S3). Interestingly, this ^{13}C was hardly transferred to TCA cycle intermediates, suggesting metabolization through other pathways. To identify potential end products, we analyzed metabolites in the cell culture medium by NMR spectroscopy. Uninfected cells converted glucose to lactate and pyruvate (Figs. 1*B* and 2*A*), which was expected for this cell line (20). Surprisingly, lactate and pyruvate excretion was completely abolished in *Shigella*-infected cells (Figs. 1*B* and 2*A*). Instead, these cells excreted acetate at a ratio of 2.1 ± 0.3 acetate per consumed glucose (Figs. 1*B* and 2*A* and Fig. S2, identification of acetate by NMR). Experiments using media containing ^{13}C glucose or ^{13}C glutamine showed that most acetate originated from glucose but not glutamine (Fig. 1*C*).

Together, these data showed that, in infected cells, the dominant glycolytic pathways were fully functional but that their entire output was rerouted to acetate (Fig. 1*D*). This finding was in marked contrast to *Shigella*-infected macrophages that become metabolically inactive and rapidly die (21) because of IpaB-mediated pyroptosis (22).

***Shigella* Produces Acetate from Pyruvate.** Acetate is the dominant waste product of rapidly growing *Enterobacteriaceae*, such as *Shigella* and *E. coli* (23). Specifically, *Shigella* uses phosphotransacetylase (PTA) to convert acetyl-CoA to acetyl phosphate and acetate kinase (ACKA) to convert acetyl phosphate and ADP to acetate and ATP. To test the role of this *Shigella* acetate-generating pathway during infection, we constructed *Shigella pta* and *ackA* mutants. Both *Shigella* mutants had substantial intracellular growth defects (40–60% of parental strain) (Fig. 2*B* and Table S4). Complementation of the *pta* mutant with an episomal *pta* allele rescued normal growth, indicating that the *Shigella pta* phenotype was caused by the deletion of *pta*. The reduced mutant growth rates resulted in strongly diminished *Shigella* loads after 3.5 h of infection (13 ± 2 *Shigella pta* and 8 ± 1 *Shigella ackA* per HeLa cell vs. 73 ± 9 parental *Shigella*), consistent with longer generation times. Over an entire day of infection, these longer generation times could lead to some 500,000-fold lower *Shigella* loads for mutants vs. the parental strain (assuming continuous growth). These data show that the acetate pathway is active and essential for normal *Shigella* intracellular growth. Similar data for polarized Caco-2 cells and human umbilical vein endothelial cells (HUVECs) (Table S4) suggested analogous *Shigella* metabolic strategies in a colon epithelial cell line as well as primary human cells.

HeLa cells infected with *Shigella pta* or *ackA* excreted lactate and pyruvate but very little acetate (Fig. 2*A*). These data suggested that most of the acetate excreted from HeLa cells that were infected with the *Shigella* parental strain was a product of

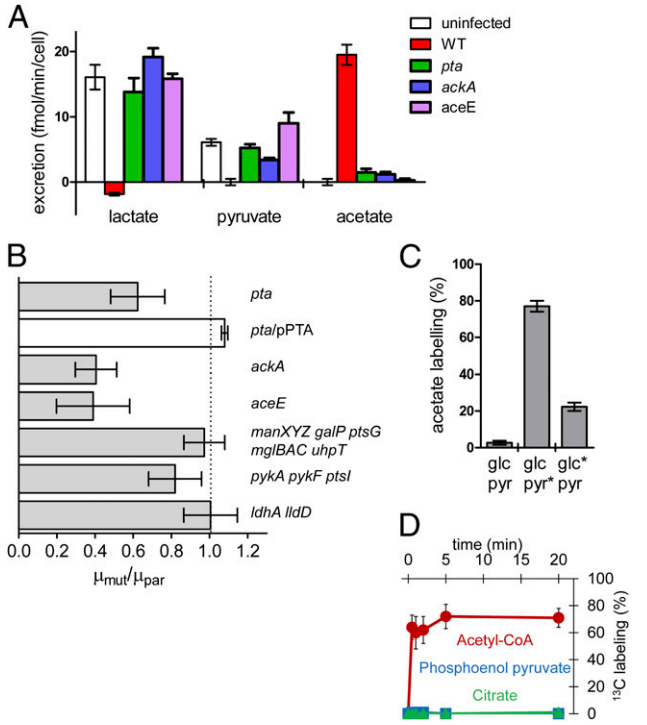


Fig. 2. (*A*) Net metabolite excretion of uninfected HeLa cells and HeLa cells infected with various *Shigella* strains between 2 and 3 h postinfection determined by NMR spectroscopy. Data represent means and SDs of three independent experiments. (*B*) Intracellular generation time of *Shigella* mutant strains (μ_{mut}) relative to the parental strain (μ_{par}; *pta*/pPTA, *Shigella pta* mutant complemented *in trans* with an episomal *pta* allele). A value of 1.0 (dashed line) corresponds to unimpaired growth. Data represent means and SDs of 34–56 replicate wells. Detailed data and statistical analysis are in Table S4. (*C*) ^{13}C labeling of extracellular acetate in medium containing U-^{13}C–labeled compounds (glc*, 25 mM glucose; pyr*, 5 mM pyruvate) or unlabeled compounds (glc, 25 mM glucose; pyr, 5 mM pyruvate). Data represent means and SDs of three independent experiments. (*D*) ^{13}C-incorporation kinetics of acetyl-CoA, phosphoenolpyruvate, and citrate in infected HeLa cells after feeding U-^{13}C pyruvate in the presence of a fivefold excess of unlabeled glucose. Labeling percentage refers to the fraction of 2-^{13}C (acetyl-CoA) or *U*-^{13}C (PEP and citrate) metabolite relative to the total metabolite pool. Data represent means and SDs of three independent experiments (full dataset is in Table S5).

Shigella pyruvate to acetate conversion, which largely depends on a functional PTA–ACKA pathway (23). The low residual acetate production of cells infected with *Shigella pta* or *ackA* mutants could reflect alternative *Shigella* pathways [such as pyruvate oxidation mediated by pyruvate oxidase (POXB)] or host cell activities (23). HeLa cells incubated together with gentamicin-killed *Shigella* excreted no acetate, again consistent with acetate as a product of *Shigella* metabolism.

Together, these data suggest that infected cells switched from lactate excretion to acetate excretion as a result of *Shigella* capturing host metabolites and converting them to acetate primarily through the PTA–ACKA pathway.

Pyruvate Is Taken Up by *Shigella* from the Host Cytosol. To determine potential host metabolites that fuel *Shigella* acetate generation, we mutated additional metabolic genes. *Shigella aceE* deficient for pyruvate dehydrogenase and thus, impaired in conversion of pyruvate to acetyl-CoA showed a strong growth defect and largely abolished acetate production, suggesting that pyruvate was a major source of acetyl-CoA and acetate in *Shigella* (Fig. 2*B* and Table S4).

MICROBIOLOGY

Shigella could obtain this pyruvate directly from the host cell cytoplasm or generate it itself from other host cell metabolites through glycolysis and/or lactate oxidation. A *Shigella manXYZ galP ptsG mglBAC uhpT* mutant unable to use glucose and mannose as well as hexose phosphates exhibited unimpaired growth (Fig. 2*B* and Table S4). A *pykA pykF ptsI* mutant defective for the last step of glycolysis [conversion of phosphoenolpyruvate (PEP) to pyruvate], which cannot grow on any glycolysis intermediate upstream of pyruvate in vitro, had a moderate growth defect (77 ± 10% of the parental strain's intracellular growth rate) (Fig. 2*B* and Table S4). These data indicated the importance of the PEP–pyruvate conversion but also showed substantial growth of *Shigella*, even without the use of any host glycolysis intermediate. This mutant phenotype was consistent with unabated ^{13}C-label incorporation in host cell metabolites from hexose phosphate all of the way down to pyruvate, suggesting at most, minor consumption of host cell intermediates upstream of pyruvate by *Shigella*. These data are in agreement with previous studies showing no growth of closely related *E. coli* on key glycolysis intermediates 3-phosphoglycerate, 2-phosphoglycerate, and phosphoenolpyruvate (24), and toxic effects of excessive uptake of organophosphates (25). Finally, *Shigella ldhA lldD* incapable of lactate utilization also maintained normal growth, arguing against lactate as a critical source of pyruvate (Fig. 2*B* and Table S4). Taken together, these data suggested direct uptake of pyruvate by *Shigella* from the host cell cytoplasm as the main supply route. A major role of pyruvate and its apparently complete conversion into acetate was also consistent with the fact that pyruvate (oxidation state +2) is a highly acetogenic substrate compared with nutrients with a lower oxidation state, such as glucose (23).

Initially, it was surprising that just a few *Shigella* cells could apparently capture the entire glycolysis output of much larger HeLa cells. However, *Shigella* can efficiently use external pyruvate as a sole carbon and energy source. Indeed, NMR spectroscopy revealed that *Shigella* consumed pyruvate in vitro at a rate of 0.35 ± 0.06 fmol min^{-1} per *Shigella* in M9 minimal medium containing 1 mM pyruvate and nicotinic acid and 0.1% tryptic soy broth as supplements. Based on these data, some 50 *Shigella* can indeed consume all pyruvate generated by host metabolism in an infected HeLa cell (around 20 fmol min^{-1}; see above).

Direct validation of *Shigella* pyruvate uptake during intracellular growth would require a mutant with blocked pyruvate uptake. Unfortunately, several yet unidentified high-affinity pyruvate transporters exist in *Shigella*/*E. coli* (26). We deleted multiple tentative candidates [*yhjE* (27); *actP*, a homolog of *mctC* (28); *ycaM* and *ybaT*, putative transporter genes up-regulated in *E. coli* on pyruvate-containing media (29)], but the resulting combination mutant was still able to grow rapidly on pyruvate in vitro, showing the presence of other yet unidentified pyruvate import mechanisms.

As an alternative approach, we added ^{13}C pyruvate to the external medium of infected HeLa cells and tracked the labeled carbon. NMR analysis of the extracellular medium revealed efficient conversion of ^{13}C pyruvate to ^{13}C acetate, even in the presence of an excess of unlabeled glucose (Fig. 2*C*), consistent with preferred use of pyruvate over glucose and glycolysis pathway intermediates. Indeed, the only detectable intracellular metabolite that rapidly acquired substantial amounts of ^{13}C in infected cells was acetyl-CoA, whereas directly adjacent metabolites PEP and citrate were not labeled or labeled at much lower rate (Fig. 2*D* and Table S5). Their poor labeling rules out Embden–Meyerhof and pentose phosphate pathways or TCA cycle intermediates as relevant sources of the strongly labeled excreted acetate and supports a direct conversion of host-derived pyruvate to acetyl-CoA and acetate by *Shigella*.

Together, these data suggest major *Shigella* nutrient use through a short pathway: pyruvate (host cytoplasm) → pyruvate (*Shigella*) → acetyl-CoA (*Shigella*) → acetyl-phosphate (*Shigella*) → acetate (*Shigella*) → acetate (host cytoplasm) → acetate (medium) (Fig. 3). For each pyruvate molecule, this pathway yields one ATP in the conversion of acetyl-phosphate to acetate. Because one molecule of glucose is essentially fully converted into two excreted acetate molecules, this pathway predominantly provides energy, but not carbon, for *Shigella* growth. The limited role of pyruvate as a carbon source might explain why this major pathway escaped detection in previous metabolic studies of related pathogens using methods focusing on pathogen biomass incorporation (10).

Intracellular *Shigella* Consume Oxygen by Respiration. Oxidation of glucose to pyruvate and pyruvate to acetate yields reducing equivalents that might, in part, be used for *Shigella* biosynthetic pathways, such as lipogenesis. In addition, infected cells consumed oxygen at a rate of 0.07 ± 0.02 fmol/min per *Shigella*, whereas uninfected cells had undetectable oxygen consumption in our assay. This oxygen consumption was mostly caused by *Shigella* respiration, because cells infected with *Shigella ubiC menA* (lacking ubiquinone and menaquinone required for aerobic respiration) consumed much less oxygen (0.003 ± 0.001 fmol/min per *Shigella*). This observation is consistent with commonly observed active respiration during acetate production in *Enterobacteriaceae* (23). Additional work is required to identify which of the many potential intermediate electron donors/acceptors are involved in the various host and *Shigella* redox reactions.

Additional Nutrients Support *Shigella* Growth. In addition to the major energy source pyruvate, *Shigella* might obtain other nutrients from the host cell cytoplasm to meet its biomass demands. We analyzed *Shigella* mutants with utilization defects for 18 diverse metabolites that are known to be available in human cell cytoplasm (30). However, all mutants except one had growth rates indistinguishable from the parental strain (Tables S6 and S7). The weak but significant growth defect of *fadD fadK* suggests a potential small contribution of host fatty acids to *Shigella* growth. We also analyzed growth phenotypes of 11 auxotrophic *Shigella* mutants that depend on external supplementation of specific nutrients for growth (Tables S6 and S7). High intracellular growth rates of these auxotrophs showed *Shigella* access to diverse amino acids in sufficient amounts to fully meet their respective *Shigella* biomass needs, which was previously observed for closely related pathogens (10). Asparagine, proline, and purine nucleosides were also available but only in limiting amounts, requiring additional *Shigella* biosynthesis for full

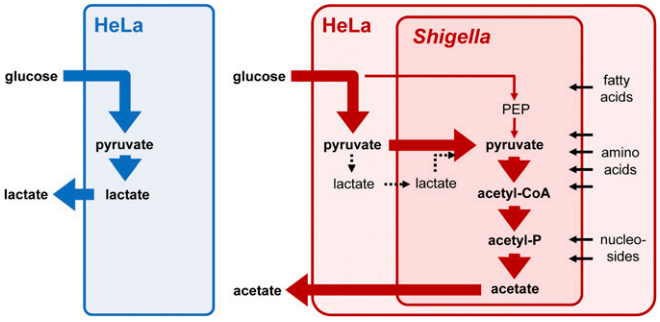


Fig. 3. Metabolic overview of *Shigella*-infected HeLa cells. Uninfected cells (blue) convert glucose to lactate. In infected cells (red), *Shigella* takes up glucose-derived pyruvate from the host cytoplasm and converts it to acetate, which is excreted into the environment. In addition, *Shigella* uses host fatty acids, amino acids, and purine nucleosides. Thick and thin arrows indicate major and minor carbon flux routes, respectively. The dashed arrows represent reactions not supported by available evidence.

growth. Availability of these (and possibly, additional) diverse nutrients represented another important contribution to support vigorous intracellular *Shigella* growth. However, compared with the massive throughput of pyruvate use, uptake of additional nutrients for *Shigella* biomass production must be rather low: based on acetate excretion, each *Shigella* consumed, on average, 1.2 pg pyruvate but generated total new biomass of only 0.4 pg (31) within one generation time (37 min).

We incorporated our data for *Shigella* pyruvate consumption, respiration, and additional biomass supplementation in an in silico genome-scale metabolic model of closely related *E. coli* (32). According to this model, the measured fluxes would enable *Shigella* growth with generation times in the range of 33–54 min (based on error margins of the underlying experimental data), which is in good agreement with our experimentally determined generation time of 37 ± 4 min, showing that nutrients identified in this study can largely explain the outstandingly rapid intracellular *Shigella* growth as one precondition for the rapid and dramatic disease progression in dysentery.

Discussion

Rapid disease progression in bacillary dysentery is the result of vigorous *Shigella* growth in gut epithelial cells. In addition to a large number of virulence factors that interfere with host signaling networks, this fast growth requires massive *Shigella* exploitation of host metabolism for high-flux nutrient supply. However, infected cells remain viable for several hours until they contain more than 100 bacteria. In this study, we show that *Shigella* does not affect the main host cell energy production pathways—Embden–Meyerhof and pentose phosphate pathways—but instead, consumes their entire output at the level of pyruvate. For this purpose, *Shigella* uses a short pyruvate-to-acetate pathway that *Enterobacteriaceae* typically use during rapid growth under nutrient-rich conditions (23). *S. flexneri* cannot reuse acetate later, because it lacks acetyl-CoA synthetase that is present in closely related *E. coli* (33).

This pathway generates only a single ATP for each pyruvate molecule (and some additional ATP through oxidation of NADH), and it is far less efficient in terms of ATP generation compared with the TCA cycle (up to 14 ATP per pyruvate) that predominates under nutrient-poor conditions. However, the low-yield acetate pathway has lower enzyme costs and membrane space requirements and can operate at much faster rates compared with low-throughput, high-yield TCA cycle/oxidative phosphorylation, which requires bulky enzymatic machinery in the bacterial inner membrane. The acetate pathway can, thus, cope with high nutrient supply rates that would saturate the respiratory chain (overflow metabolism) (34). Similar switches to low-yield/high-rate pathways are observed for many microbial cells (Crabtree effect) (23) as well as cancer cells (Warburg effect) (20) during rapid growth.

The use of abundant pyruvate predominantly through this low-yield, high-speed acetate pathway represents a typical metabolic pattern of *Enterobacteriaceae* in general (23). Interestingly, this default pathway could offer several additional important benefits to *Shigella* during intracellular growth in infected human cells: (*i*) use of a favorable abundant energy source, pyruvate, at high rates; (*ii*) exploitation of the dominant preexisting host metabolic pathways as high-flux supply lines without requiring major host cell metabolism alterations that could cause delays and limit supply; and (*iii*) preservation of host cell energy charge by glycolytic ATP generation, allowing extended host cell survival, despite vigorous exploitation. Taken together, a short metabolic pathway that is commonly present in many bacteria enables *Shigella* to efficiently exploit major nutrient supply routes in infected host cells. This preadaptation might explain why *Shigella* can thrive as a voracious pathogen with only minor metabolic adaptations to the host cell intracellular environment compared with closely related extracellular commensals (33, 35). In addition to the major energy source pyruvate, *Shigella* accessed diverse host metabolites for direct biomass incorporation. Computational modeling revealed that all these nutrients together were sufficient to support the experimentally observed rapid *Shigella* intracellular growth.

An important caveat of our study was the use of cancer cell lines as host cells that ferment glucose to lactate to satisfy their energy needs (the Warburg effect). Dysentery is a disease of the colon, and the epithelial cells in this location, the colonocytes, have a different metabolism. Some 80% of ATP produced in colonocytes comes from mitochondrial oxidation of butyrate, which is derived from gut microbiota. However, experiments with germ-free mice have shown that, in the absence of microbiota-derived butyrate, colonocytes switch from burning butyrate to fermenting glucose to lactate (36). It is possible that the flushing action of diarrhea during dysentery removes much of the microbiota, thereby lowering butyrate levels in the colon and triggering colonocyte metabolism similar to HeLa cell metabolism in our in vitro model. Additional studies are required to clarify this issue.

Comprehensive quantitative data on pathogen nutrition, metabolism, and growth in host environments are largely lacking. However, comparison of the *Shigella* data with our recent study on *Salmonella enterica* metabolism in a mouse typhoid fever model (37) reveals commonalities and striking differences. *Shigella*, like *Salmonella* and many other pathogens (37), exploits the fact that most biomass components are readily available in infected host environments. Uptake of these biomass components can save substantial biosynthesis costs. However, *Shigella* has access to a high-flux supply for pyruvate that provides sufficient energy to drive fast growth, whereas intracellular *Salmonella* has access to many diverse but only scarce energy sources that together just support slow nutrient-limited growth (37). This striking difference may reflect the fact that *Shigella* reside directly in the host cytoplasm, whereas intracellular *Salmonella* are surrounded by a phagosomal membrane with apparently poor permeability in macrophages (their main target cell type during systemic infections). Additional studies with other pathogens might clarify the general relevance of these distinct metabolic patterns.

Although host pyruvate supply enables vigorous *Shigella* proliferation, the heavy dependence on pyruvate metabolism makes *Shigella* also vulnerable to metabolic perturbation. Indeed, our mutant data for the corresponding enzymes show dramatically reduced *Shigella* loads compared with the parental strain, even within short infection times. The residual slow growth of these mutants and their altered waste product spectrum show that *Shigella* can use alternative pathways, such as pyruvate oxidase, but these pathways are far less efficient for supporting growth as expected (23). Importantly, the key enzymes phosphotransacetylase (PTA) and acetate kinase (ACKA) have no human homologs, suggesting that specific inhibition of *Shigella* enzymes without adverse effects on human metabolism might be possible. *Shigella* acetate metabolism enzymes could, thus, represent promising targets to control bacillary dysentery.

Materials and Methods

Strains and Plasmids. HeLa Kyoto cells (38) and HUVECs were provided by Cécile Arrieumerlou and Claudia Mistl (Biozentrum, University of Basel). Caco-2 cells (ATCC-HTB-37) were obtained from the American Type Culture Collection. *S. flexneri* 2a 2457T *icsA* was provided by M. B. Goldberg (Massachusetts General Hospital, Boston). All mutants in this work were derived from this *icsA* strain (referred to as the parental strain). Mutants were constructed as described (39) using a λ-red recombinase-mediated allelic replacement system introducing a flippase recognition target sites-flanked chloramphenicol resistance marker (40) followed by purification using phage P1 transduction. Combinations of multiple deletions were made by P1 transduction into mutants, in which the chloramphenicol resistance was removed using flippase. For complementation of the *pta* mutant, we cloned *pta* together with its native promoter (intergenic region upstream of the *ackA-pta* operon) on a medium copy number plasmid (pA15 ori). All strains

carried plasmid pNF106 (SC101 ori) for doxycycline/anhydrotetracycline-inducible expression of *gfp* (39).

Labeled Nutrients. [*U*-99% ^{13}C]-labeled glucose, [*U*-99% ^{13}C; *U*-99% ^{15}N]-labeled glutamine, and [*U*-99% ^{13}C]-labeled pyruvate were purchased from Cambridge Isotopes Laboratories.

***Shigella* Infections.** HeLa cells, CaCo-2 cells, and HUVECs were infected with polylysine-treated *Shigella*. After 30 min, cells were washed and treated with gentamicin. Intracellular growth of GFP-expressing *Shigella* was determined using flow cytometry and plating. Protocols for infections and growth determination are described in detail in *SI Materials and Methods*.

Proteomics of Infected Cells. Cells were harvested, digested with trypsin/chymotrypsin, and analyzed by MS as described in *SI Materials and Methods*.

Measurement of Intracellular Metabolites by MS. Uninfected and infected cells were quenched with acetonitrile/methanol/formic acid. Samples were analyzed by MS. Detailed protocols are described in *SI Materials and Methods*.

Measurement of Glucose Uptake Rates. Glucose consumption was determined using the EnzyChrom Glucose Assay Kit (Bioassay Systems) as described in *SI Materials and Methods*.

Measurement of Excreted Extracellular Metabolites. Supernatants of uninfected and infected cells were analyzed by 1D ^{1}H NMR spectroscopy as described in *SI Materials and Methods*.

Oxygen Consumption Measurements. Oxygen consumption of infected and uninfected cells was determined using Oxoplates (OP96C; PreSens) and a fluorescent plate reader as described in *SI Materials and Methods*.

Modeling of Intracellular *Shigella* Metabolism. Experimental metabolite uptake rates and mutant phenotypes were combined with a modified *E. coli* metabolism reconstruction (32) to model intracellular *Shigella* metabolism using Flux-Balance Analysis as described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank M. B. Goldberg (Massachusetts General Hospital) as well as Cécile Arrieumerlou and Claudia Mistl (Biozentrum, University of Basel) for providing materials, Nicole Freed (Biozentrum, University of Basel) for help with live cell imaging, Philipp Christen (Eidgenössiche Technische Hochschule Zürich) for support with sampling and HPLC-MS, Claudia Fortes and Philipp Schläfli (Functional Genomics Center Zurich) for assistance in proteomic analysis, and Endre Laczko and Tshering Altherr (Functional Genomic Center Zurich) for support in GC-MS. Funding was provided by SystemsX, The Swiss Initiative in Systems Biology (Project "BattleX").

1. DuPont HL, Levine MM, Hornick RB, Formal SB (1989) Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis* 159(6):1126–1128.
2. Schroeder GN, Hilbi H (2008) Molecular pathogenesis of *Shigella* spp.: Controlling host cell signaling, invasion, and death by type III secretion. *Clin Microbiol Rev* 21(1):134–156.
3. Perdomo OJ, et al. (1994) Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis. *J Exp Med* 180(4):1307–1319.
4. Phalipon A, Sansonetti PJ (2007) *Shigella*'s ways of manipulating the host intestinal innate and adaptive immune system: A tool box for survival? *Immunol Cell Biol* 85(2):119–129.
5. Campbell-Valois FX, et al. (2014) A fluorescent reporter reveals on/off regulation of the *Shigella* type III secretion apparatus during entry and cell-to-cell spread. *Cell Host Microbe* 15(2):177–189.
6. Ray K, Marteyn B, Sansonetti PJ, Tang CM (2009) Life on the inside: The intracellular lifestyle of cytosolic bacteria. *Nat Rev Microbiol* 7(5):333–340.
7. Götz A, Eylert E, Eisenreich W, Goebel W (2010) Carbon metabolism of enterobacterial human pathogens growing in epithelial colorectal adenocarcinoma (Caco-2) cells. *PLoS ONE* 5(5):e10586.
8. Andersen KB, von Meyenburg K (1980) Are growth rates of *Escherichia coli* in batch cultures limited by respiration? *J Bacteriol* 144(1):114–123.
9. Fuchs TM, Eisenreich W, Heesemann J, Goebel W (2012) Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. *FEMS Microbiol Rev* 36(2):435–462.
10. Eisenreich W, Dandekar T, Heesemann J, Goebel W (2010) Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nat Rev Microbiol* 8(6):401–412.
11. Rohmer L, Hocquet D, Miller SI (2011) Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol* 19(7):341–348.
12. Abu Kwaik Y, Bumann D (2013) Microbial quest for food in vivo: 'Nutritional virulence' as an emerging paradigm. *Cell Microbiol* 15(6):882–890.
13. Kafsack BF, Llinás M (2010) Eating at the table of another: Metabolomics of host-parasite interactions. *Cell Host Microbe* 7(2):90–99.
14. Bernardini ML, Mounier J, d'Hauteville H, Coquis-Rondon M, Sansonetti PJ (1989) Identification of icsA, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc Natl Acad Sci USA* 86(10):3867–3871.
15. Goldberg MB, Theriot JA (1995) *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proc Natl Acad Sci USA* 92(14):6572–6576.
16. Sansonetti PJ, Ryter A, Clerc P, Maurelli AT, Mounier J (1986) Multiplication of *Shigella flexneri* within HeLa cells: Lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. *Infect Immun* 51(2):461–469.
17. Mantis N, Prévost MC, Sansonetti P (1996) Analysis of epithelial cell stress response during infection by *Shigella flexneri*. *Infect Immun* 64(7):2474–2482.
18. Munger J, et al. (2008) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat Biotechnol* 26(10):1179–1186.
19. Goldberg EB, Colowick SP (1965) The role of glycolysis in the growth of tumor vells. 3. Lactic dehydrogenase as the site of action of oxamate on the growth of cultured cells. *J Biol Chem* 240:2786–2790.
20. Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science* 324(5930):1029–1033.
21. Sansonetti PJ, Mounier J (1987) Metabolic events mediating early killing of host cells infected by *Shigella flexneri*. *Microb Pathog* 3(1):53–61.
22. Senerovic L, et al. (2012) Spontaneous formation of IpaB ion channels in host cell membranes reveals how *Shigella* induces pyroptosis in macrophages. *Cell Death Dis* 3:e384.
23. Wolfe AJ (2005) The acetate switch. *Microbiol Mol Biol Rev* 69(1):12–50.
24. Saier MH, Jr., Wentzel DL, Feucht BU, Judice JJ (1975) A transport system for phosphoenolpyruvate, 2-phosphoglycerate, and 3-phosphoglycerate in *Salmonella typhimurium*. *J Biol Chem* 250(13):5089–5096.
25. Kadner RJ, Murphy GP, Stephens CM (1992) Two mechanisms for growth inhibition by elevated transport of sugar phosphates in *Escherichia coli*. *J Gen Microbiol* 138(10):2007–2014.
26. Kreth J, Lengeler JW, Jahreis K (2013) Characterization of pyruvate uptake in *Escherichia coli* K-12. *PLoS ONE* 8(6):e67125.
27. Huhn S (2011) Identifikation und Charakterisierung von bakteriellen Carbonsäure-Transportern. PhD thesis (Cologne Univ, Cologne, Germany).
28. Jolkver E, et al. (2009) Identification and characterization of a bacterial transport system for the uptake of pyruvate, propionate, and acetate in *Corynebacterium glutamicum*. *J Bacteriol* 191(3):940–948.
29. Göhler AK, et al. (2011) More than just a metabolic regulator—elucidation and validation of new targets of PdhR in *Escherichia coli*. *BMC Syst Biol* 5:197.
30. Wishart DS, et al. (2013) HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res* 41(Database issue):D801–D807.
31. Neidhardt FC (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ASM Press, Washington, DC), 2nd Ed.
32. Feist AM, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3:121.
33. Monk JM, et al. (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci USA* 110(50):20338–20343.
34. van Hoek MJ, Merks RM (2012) Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Syst Biol* 6:22.
35. Bliven KA, Maurelli AT (2012) Antivirulence genes: Insights into pathogen evolution through gene loss. *Infect Immun* 80(12):4061–4070.
36. Donohoe DR, Wali A, Brylawski BP, Bultman SJ (2012) Microbial regulation of glucose metabolism and cell-cycle progression in mammalian colonocytes. *PLoS ONE* 7(9):e46589.
37. Steeb B, et al. (2013) Parallel exploitation of diverse host nutrients enhances *Salmonella* virulence. *PLoS Pathog* 9(4):e1003301.
38. Neumann B, et al. (2006) High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat Methods* 3(5):385–390.
39. Mary C, et al. (2012) Functional identification of APIP as human *mtnB*, a key enzyme in the methionine salvage pathway. *PLoS ONE* 7(12):e52877.
40. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97(12):6640–6645.

Supporting Information

Kentner et al. 10.1073/pnas.1406694111

SI Materials and Methods

HeLa Infections. HeLa cells were cultivated in DMEM (Sigma) containing 10 mM Hepes, 25 mM glucose, and 4 mM glutamine. HeLa cells were initially seeded in DMEM supplemented with 10% (vol/vol) fetal calf serum and penicillin/streptomycin solution (100 units/mL; Invitrogen). At 16 and 3.5 h before infection, medium was exchanged for serum-free DMEM containing glucose and glutamine (5 mM pyruvate was also included in some experiments). *Shigella* was grown to exponential phase in tryptic soy broth, coated with poly-L-lysine, and added at a multiplicity of infection (MOI) of 75 (metabolomics and proteomics) or 10 (growth rate assays). *Shigella* was centrifuged onto HeLa cells (600 × *g* for 5 min). At 30 min postinfection, we added gentamicin (gent; 100 μg/mL) to kill extracellular bacteria and 1 μg/mL anhydrotetracycline (aTc) to induce *gfp* expression in live intracellular bacteria. In metabolomics experiments, we used 200 ng/mL doxycycline instead of aTc to avoid addition of the aTc solvent ethanol. At an MOI of 75, typically around 90% of the HeLa cells became infected and reached a *Shigella* load of ~50 in infected cells at 2.5 h postinfection, which was determined by comparing GFP amounts of infected cells and individual *Shigella* released by shear stress using flow cytometry.

Infections of Caco-2 Cells and Human Umbilical Vein Endothelial Cells. We infected Caco-2 and human umbilical vein endothelial cells and analyzed them for intracellular *Shigella* growth as described for HeLa cells with the following differences: human umbilical vein endothelial cells were cultivated in Gibco Medium 199 (product no. 22340020; Life Technologies) and infected with an MOI of 200. Caco-2 cells were grown in DMEM on 3-μm cell culture filter inserts (product no. 353492; Becton Dickinson) to facilitate infection from the basal side of this polarized cell line, mimicking the natural route of *Shigella* infection of gut epithelial cells. Differentiation of Caco-2 cells was achieved using the BioCoat Intestinal Epithelium Differentiation Environment Kit (product no. 355057; Becton Dickinson). To infect Caco-2 cells, we flipped inserts upside-down, placed custom-made plastic caps on the inserts, and added poly-L-lysine–coated *Shigella* (MOI of 100 or more). After centrifugation (1 min at 600 × *g*), inserts were flipped back and filled with 0.5 mL DMEM.

Intracellular Growth Rate Determination. HeLa cells were infected with *Shigella* at an MOI of 10 (to minimize multiple infections of the same host cell). Gent and aTc were added at 30 min postinfection. Cells were trypsinized and fixed at 2.5 or 3.5 h postinfection. *Shigella* loads of infected cells were determined by comparing GFP amounts in infected cells and individual *Shigella* liberated by shear stress using flow cytometry. These GFP-based data closely agree with the colony-forming units (CFUs) as determined by plating, and both methods also yield similar values for the fold increase between two time points postinfection (Fig. S1). Furthermore, live cell imaging shows that almost all GFP-*Shigella* rapidly replicate during intracellular infections (Movie S1). These data suggest that GFP particles represent viable *Shigella*, similar to previous observations for *Salmonella* (1).

Our infection cultures might also contain some dead *Shigella*, because we killed extracellular *Shigella* with gent after 30 min to synchronize the infection. Importantly, we induced *gfp* expression only after this addition of gent. Extracellular *Shigella* is immediately inhibited in its protein synthesis and thus, remains GFP-negative and dies (we confirmed blocked GFP synthesis in control experiments). In contrast, intracellular *Shigella* is protected from extracellular gent, starts to express GFP, and stays alive.

Taken together, all GFP-*Shigella* seems to be alive, whereas potentially present dead *Shigella* remains GFP-negative. These data support the use of flow cytometry high-throughput measurements for assessing intracellular *Shigella* growth. We determined growth rates by calculating the increase in *Shigella* loads between 2.5 and 3.5 h postinfection. HeLa cells infected with the parental strain contained 24 ± 4 and 73 ± 9 *Shigella* at 2.5 and 3.5 h postinfection, respectively, corresponding to a generation time of 37 ± 4 min. Data for *Shigella* mutants are represented as relative rates compared with the parental strain.

Proteomics of Infected Cells. HeLa cells were infected at an MOI of 75 and sampled at 3 h postinfection together with uninfected controls (six biological replicates each). Cells were washed two times with PBS, treated with a 4 °C quenching solution [EDTA-free protease inhibitor mixture (Roche), 6 M urea, 0.1 M NH_4HCO_3], and harvested by scraping. Samples were sonicated three times for 45 s and stored at −80 °C. For each sample, 500 μg protein was digested with trypsin/chymotrypsin (2), desalted, dried, and reconstituted in 3% (vol/vol) acetonitrile and 0.2% (vol/vol) formic acid for a final peptide concentration of 0.25 μg/μL. MS analyses were carried out on an LTQ-Orbitrap Velos Mass Spectrometer (Thermo Scientific) coupled to an Eksigent-Nano-HPLC System (Eksigent Technologies). For each run, 1 μg peptides were loaded on a self-made C18 column (75 μm × 150 mm, 3 μm, 200 Å) and eluted with a gradient from 3% to 35% (vol/vol) acetonitrile in 142 min (flow rate = 250 nL/min). The acquisition method consisted of one survey scan (MS) followed by 20 dependent scans (MS/MS) looped throughout the run. All samples were run two times. Proteins were identified using the Mascot 2.3.02 Search Engine (Matrix Science) against an in-house database containing human (Swissprot), *S. flexneri 2457T*, *Escherichia coli* MG1655, and reversed decoy sequences as well as the common contaminants. Peptides identified with a Mascot ion score higher than 40 were accepted, resulting in a protein false discovery rate of 1.65%. Label-free quantification was performed using Progenesis LC-MS v.4.0 (Nonlinear Dynamics), normalizing on all human features.

Measurement of Intracellular Metabolites by MS. For each condition, we prepared three (six for isotope tracking with ^{13}C glucose) independent biological replicates on different days. HeLa cells were grown on 28-mm coverslips to allow quick washing and quenching of cells. Samples were taken at 2.5 h postinfection. For quantification of metabolite concentrations, coverslips were briefly washed in rapidly stirred ddH_2O at 37 °C and dropped into 8 mL quenching solution (40% acetonitrile, 40% methanol, 20% 0.5 M formic acid, all vol/vol) at −20 °C together with a ^{13}C standard prepared from *Methylobacterium extorquens* and *Saccharomyces cerevisiae* containing ^{13}C-labeled metabolites of known concentration. Total time for washing and quenching was 2 s. Quenched samples were sonicated, freeze-dried, and dissolved in ddH_2O. To remove debris, samples were centrifuged, and the supernatant was transferred to new tubes and centrifuged a second time.

To determine *Shigella* adenosine phosphates content, *Shigella* cultures were grown until an OD_{600} of 0.1 in M9 minimal medium containing 0.1 or 1 g L^{-1} glucose or 0.1 or 1 g L^{-1} pyruvate. *Shigella* cells were sampled during exponential growth by fast filtration without washing (3, 4). For quenching and metabolite extraction, filters were transferred into −20 °C cold quenching solution (acetonitrile:methanol:0.5 M formic acid at 60:20:20).

For analysis of ^{13}C incorporation kinetics using ^{13}C glucose or ^{13}C pyruvate, coverslips were washed in DMEM lacking glucose or pyruvate, transferred to DMEM containing ^{13}C glucose or ^{13}C pyruvate, and incubated for 30 s, 1 min, 2 min, 5 min, or 20 min before washing in ddH_2O and quenching (which was always done at 2.5 h postinfection). Unquenched control samples were trypsinized, fixed, and analyzed by flow cytometry to estimate infection efficiency and *Shigella* load. Additional control samples were used to count the number of HeLa cells per coverslip.

All metabolites except pyruvate were measured by liquid chromatography–mass spectrometry (MS). Samples were diluted 1:4 (vol:vol) with a tributylamine solution (1.7 mM acetic acid, 3.28 mM tributylamine adjusted to pH 9.2 with ammonium hydroxide). Analytical measurements were performed using a nanoliquid chromatography system (nano-2D Ultra LC; Eksigent Technologies Inc.) coupled to a tandem mass spectrometer (LTQ Orbitrap; Thermo Fisher Scientific) as described previously (5) with slight modifications. The stationary phase was replaced by HaLo C18 particles (2.7 μm; New Objective Inc.), and the tributylamine solution had a slightly lower pH of 9.2. Xcalibur, version 2.1 (Thermo Fisher Scientific) and eMZed (6), version 1.3 were used for data acquisition and analysis, respectively.

Pyruvate was measured by gas chromatography-MS. Samples were derivatized with trimethylsilane and analyzed on a Waters GCT Premier Spectrometer (Waters) equipped with Restek GC Column Rxi-5Sil MS (Restek Corporation). Acquisition and data analysis were performed with MassLinx software (Waters) as described previously (7).

Measurement of Glucose Uptake Rates. Three independent biological replicates of infected or uninfected HeLa cells were prepared on different days. Supernatant samples were taken at 2 and 3 h postinfection, and glucose was quantified using the EnzyChrom Glucose Assay Kit (Bioassay Systems). The culture medium contained 2.5 mM (instead of 25 mM) glucose to achieve a measureable decrease in concentration between 2 and 3 h postinfection. Lowering the glucose concentration did not affect lactate or acetate excretion in uninfected or infected cells, and for uninfected cells, uptake rates at 2.5 and 25 mM determined over 10 h were similar. Uptake of glutamine, measured with the EnzyChrom Glutamine Assay Kit (Bioassay Systems), was below the detection threshold of 0.5 fmol/min per cell for both uninfected and infected HeLa cells.

Measurement of Excreted Extracellular Metabolites. Three independent biological replicates of infected or uninfected HeLa cells were prepared on different days. At 20 min postinfection, cells were switched to DMEM containing gent, doxycyclin, and unlabeled or uniformly labeled ^{13}C glucose, glutamine, and/or pyruvate. Supernatant samples were taken at 20 min, 2 h, and 3 h postinfection. After medium sampling, infection efficiency and *Shigella* load were determined by flow cytometry, and HeLa cells were counted for each sample.

All NMR samples were prepared by diluting 100 μL supernatant medium to a total volume of 500 μL with 5% (vol/vol) D_2O, 20 μM trimethylsilylproprionate, and 0.05% (wt/vol) sodium azide in the final sample. All experiments were recorded at 30 °C on Bruker Avance III Spectrometers with field strengths of 600, 700, or 800 MHz. The spectra were calibrated using trimethylsilylproprionate as chemical shift reference (8). To quantify the amounts of unlabeled and labeled metabolites, 1D ^1H NMR experiments with ^{12}C and ^{13}C filtering were recorded (9). NMR data were processed and analyzed using Topspin. For concentration determination, the PULCON method (10) was used combined with interactive intensity scaling relative to reference spectra of pure compound samples with uniformly labeled ^{13}C or natural abundance labeling.

Oxygen Consumption Measurements. Infections were done as for growth rate assays, except that 96-well plates containing an oxygen-sensitive fluorescence sensor (Oxoplates OP96C; PreSens) were used. After addition of gent/aTc at 20 min postinfection, the medium in each well was overlaid with 100 μL mineral oil, and the plate was sealed with VIEWseal Transparent Film (Greiner Bio-One) to prevent oxygen exchange between medium and ambient air. Fluorescence was continuously recorded in a plate reader for 4 h, and oxygen concentrations were calculated according to the manufacturer's protocol.

Modeling of Intracellular *Shigella* Metabolism. We started with a genome-scale in silico reconstruction for closely related *E. coli* (11) and removed genes present in *E. coli* but absent in *S. flexneri*, including *cadA*, *nadB*, and *acs*. The resulting reconstruction was used as an approximation for *Shigella*. *Shigella* growth was modeled by Flux-Balance Analysis using the COBRA 2.0 Toolbox (12) in a MATLAB environment. Uptake rates for pyruvate and oxygen as well as acetate excretion rates were derived from our experimental data. Uptake rates for amino acids, nucleosides, and nicotinic acid were estimated based on intracellular growth rates of *Shigella* auxotrophs as recently described for *Salmonella* (13).

1. Barat S, et al. (2012) Immunity to intracellular *Salmonella* depends on surface-associated antigens. *PLoS Pathog* 8(10):e1002966.
2. Fischer F, Wolters D, Rögner M, Poetsch A (2006) Toward the complete membrane proteome: High coverage of integral membrane proteins through transmembrane peptide detection. *Mol Cell Proteomics* 5(3):444–453.
3. Bolten CJ, Kiefer P, Letisse F, Portais JC, Wittmann C (2007) Sampling for metabolome analysis of microorganisms. *Anal Chem* 79(10):3843–3849.
4. Bennett BD, Yuan J, Kimball EH, Rabinowitz JD (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat Protoc* 3(8):1299–1311.
5. Kiefer P, Delmotte N, Vorholt JA (2011) Nanoscale ion-pair reversed-phase HPLC-MS for sensitive metabolome analysis. *Anal Chem* 83(3):850–855.
6. Kiefer P, Schmitt U, Vorholt JA (2013) eMZed: An open source framework in Python for rapid and interactive development of LC/MS data analysis workflows. *Bioinformatics* 29(7):963–964.
7. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 1(1):387–396.
8. Pohl L, Eckle M (1969) Cyclosilane-D18 a new 1h-NMR standard for measurements up to 200 degrees C. *Angew Chem Int Ed Engl* 8(5):380.
9. Otting G, Wüthrich K (1990) Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: Combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q Rev Biophys* 23(1):39–96.
10. Dreier L, Wider G (2006) Concentration measurements by PULCON using X-filtered or 2D NMR spectra. *Magn Reson Chem* 44(Spec No):S206–S212.
11. Feist AM, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3(2007): 121.
12. Schellenberger J, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat Protoc* 6(9):1290–1307.
13. Steeb B, et al. (2013) Parallel exploitation of diverse host nutrients enhances *Salmonella* virulence. *PLoS Pathog* 9(4):e1003301.

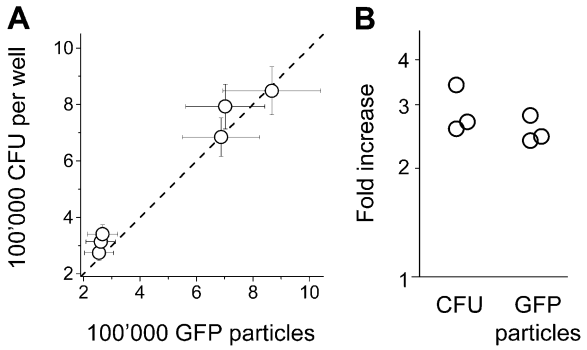


Fig. S1. Comparison of flow cytometry and plating for analysis of *Shigella* infections. (*A*) Number of GFP particles as detected by flow cytometry and colony-forming units (CFUs) as determined by plating per well of infected HeLa cells. The data represent means and SDs of 10 wells each for two time points (2.5 and 3.5 h postinfection) from three independent experiments. (*B*) Fold increase in CFUs or GFP particles between 2.5 and 3.5 h postinfection as calculated from data shown in *A*.

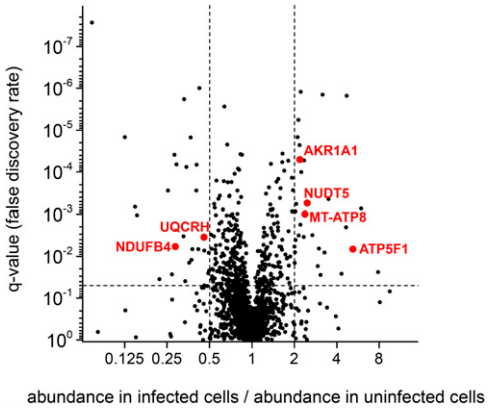


Fig. S2. Changes in protein levels of HeLa cells on *Shigella* infection. The levels of 1,848 host proteins were measured in infected and uninfected HeLa by MS-based shot gun proteomics; 58 proteins (3.1%) had at least twofold abundance differences (27 less abundant in infected cells and 31 more abundant in infected cells) and q values < 0.05. These differentially abundant proteins included six metabolic enzymes (red circles). Data represent geometric means (abundance ratios) and q values calculated from six independent experiments. AKR1A1, Aldo-keto reductase family 1 member A1 (oxidoreductase with broad specificity); ATP5F1, mitochondrial ATP synthase subunit 8 (involved in energy production); MT-ATP8, mitochondrial ATP synthase subunit 8 (involved in energy production); NDUFB4, mitochondrial NADH dehydrogenase (ubiquinone) 1 β-subcomplex, subunit 4 (involved in energy production through aerobic respiration); NUDT5, Nudix type 5 motif hydrolase (degrades toxic ADP ribose, a potential NAD^+ degradation product, and other toxic nucleoside derivatives); UQCRH, mitochondrial cytochrome *b*-c1 complex subunit 6 (involved in energy production through aerobic respiration).

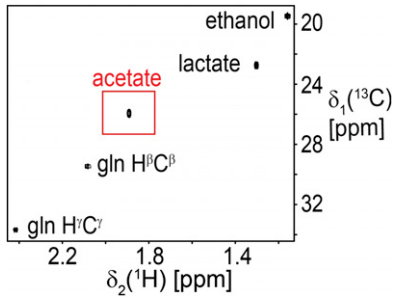


Fig. S3. Identification of extracellular metabolites from infected HeLa cells by 2D [^{13}C,^{1}H] heteronuclear multiple quantum coherence NMR. The subspectrum with ^{13}C-enriched acetate (red square) is plotted at higher-contour base level to account for the increased signal intensity relative to the other natural abundance cross-peaks. gln, Glutamine.

Table S1. Intracellular metabolite levels of infected and uninfected HeLa cells: metabolites with absolute quantification

Metabolite	Concentration (x × mol/cell)			Log10 (infected/uninfected; range; P value)
	x	Uninfected	Infected	
Hexose 1-phosphate	10^{-17}	33 ± 7	12 ± 3	−0.42 (−0.58 to −0.30; 0.00)
UTP	10^{-16}	39 ± 11	15 ± 5	−0.41 (−0.66 to −0.25; 0.02)
Pantothenic acid	10^{-17}	143 ± 50	62 ± 24	−0.36 (−0.68 to −0.18; 0.05)
cADP ribose	10^{-18}	186 ± 56	91 ± 32	−0.31 (−0.58 to −0.15; 0.00)
Fumarate	10^{-17}	110 ± 32	55 ± 10	−0.30 (−0.48 to −0.17; 0.01)
CTP	10^{-17}	153 ± 44	78 ± 32	−0.29 (−0.59 to −0.12; 0.08)
UDP	10^{-17}	141 ± 58	75 ± 40	−0.27 (−0.76 to −0.05; 0.01)
Malate	10^{-16}	28 ± 7	15 ± 4	−0.25 (−0.43 to −0.13; 0.03)
ATP	10^{-16}	88 ± 23	51 ± 22	−0.23 (−0.54 to −0.06; 0.01)
Hexose 6-phosphate	10^{-17}	121 ± 28	72 ± 3	−0.23 (−0.34 to −0.14; 0.01)
Mesaconic acid	10^{-18}	44 ± 24	27 ± 5	−0.21 (−0.58 to −0.01; 0.12)
Sedoheptulose 7-phosphate	10^{-17}	26 ± 13	16 ± 4	−0.20 (−0.55 to −0.01; 0.02)
2-Deoxy-ATP	10^{-18}	88 ± 22	56 ± 19	−0.20 (−0.44 to −0.04; 0.00)
ADP	10^{-16}	26 ± 10	17 ± 9	−0.20 (−0.66–0.02; 0.03)
ADP-pentose	10^{-20}	129 ± 66	85 ± 72	−0.18 (−2.50–0.12; 0.95)
Glycerate bisphosphate	10^{-18}	112 ± 27	76 ± 40	−0.16 (−0.54–0.03; 0.05)
GTP	10^{-16}	19 ± 4	13 ± 6	−0.14 (−0.43–0.03; 0.02)
CDP	10^{-17}	21 ± 9	16 ± 8	−0.14 (−0.64–0.09; 0.23)
GDP	10^{-17}	65 ± 27	48 ± 24	−0.13 (−0.60–0.09; 0.68)
AMP	10^{-17}	84 ± 30	67 ± 31	−0.10 (−0.49–0.10; 0.11)
Glutathione oxidized	10^{-16}	104 ± 18	83 ± 26	−0.10 (−0.29–0.04; 0.69)
Triose phosphate (DHAP, GAP)	10^{-17}	58 ± 24	48 ± 26	−0.09 (−0.59–0.14; 0.09)
Ethylmalonic acid	10^{-17}	14 ± 5	13 ± 3	−0.05 (−0.29–0.10; 0.66)
UMP	10^{-17}	13 ± 5	12 ± 5	−0.04 (−0.42–0.15; 0.39)
Aconitate	10^{-18}	36 ± 13	33 ± 7	−0.04 (−0.27–0.12; 0.82)
Gluconate phosphate	10^{-17}	43 ± 14	41 ± 7	−0.02 (−0.22–0.12; 0.01)
GMP	10^{-17}	12 ± 4	12 ± 6	0.00 (−0.39–0.20; 0.12)
Citrate	10^{-16}	11 ± 4	12 ± 1	0.01 (−0.17–0.14; 0.65)
TDP	10^{-19}	70 ± 31	73 ± 34	0.02 (−0.43–0.23; 0.13)
CMP	10^{-18}	45 ± 16	64 ± 20	0.16 (−0.12–0.33; 0.01)
α-Ketoglutarate	10^{-17}	70 ± 19	133 ± 23	0.28 (0.11–0.40; 0.01)
Hexose 1,6-bisphosphate	10^{-16}	11 ± 3	23 ± 11	0.32 (−0.02–0.51; 0.01)
Sedoheptulose bisphosphate	10^{-18}	82 ± 16	246 ± 62	0.48 (0.31–0.60; 0.00)
Glycerate phosphate	10^{-17}	24 ± 5	136 ± 8	0.75 (0.65–0.84; 0.00)
Glyceric acid	10^{-17}	23 ± 6	159 ± 20	0.85 (0.70–0.95; 0.03)
PEP	10^{-18}	50 ± 11	437 ± 3	0.94 (0.84–1.02; 0.00)
Pentose 5-phosphate	10^{-17}	21 ± 5	285 ± 25	1.14 (1.01–1.23; 0.00)
Ribose 5-phosphate	10^{-17}	34 ± 10	891 ± 186	1.42 (1.22–1.56; 0.00)

Data represent means and SDs of three independent replicates. DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate.

Table S2. Intracellular metabolite levels of infected and uninfected HeLa cells: metabolites with relative quantification

Metabolite	Signal (a.u.)		Log10 (infected/uninfected; range; *P* value)
	Uninfected	Infected	
Serine phosphate	1,160 ± 232	29 ± 5	−1.6 (−1.73 to −1.5; 0.00)
AICAR	253 ± 50	54 ± 44	−0.67 (−1.44 to −0.41; 0.01)
Propionyl-CoA	139 ± 41	46 ± 21	−0.48 (−0.82 to −0.29; 0.90)
UDP-*N*-acetylglucosamine	28 ± 9	11 ± 5	−0.42 (−0.75 to −0.23; 0.03)
dTTP	139 ± 46	56 ± 24	−0.4 (−0.74 to −0.21; 0.00)
Glyceryl-phosphate	28 ± 8	11 ± 4	−0.4 (−0.66 to −0.24; 0.03)
UDP-diphosphoglucurate	23 ± 8	12 ± 6	−0.29 (−0.75 to −0.08; 0.45)
NAD	153 ± 43	83 ± 28	−0.26 (−0.51 to −0.11; 0.00)
FAD	31 ± 9	19 ± 8	−0.22 (−0.51 to −0.04; 0.08)
UDP-Hexose	25 ± 7	15 ± 6	−0.22 (−0.52 to −0.04; 0.03)
Succinyl-CoA	36 ± 11	28 ± 7	−0.12 (−0.34–0.02; 0.12)
NADH	36 ± 14	27 ± 11	−0.12 (−0.48–0.07; 0.51)
Sedoheptulose phosphate	15 ± 3	16 ± 5	0.04 (−0.17–0.17; 0.98)
Acetyl-CoA	10 ± 3	18 ± 3	0.23 (0.03–0.37; 0.00)
N-acetyl-glutamate	67 ± 21	125 ± 10	0.27 (0.1–0.39; 0.00)
Succinate	36 ± 6	119 ± 22	0.51 (0.39–0.61; 0.00)
Pentose 1,5-bisphosphate	57 ± 35	461 ± 271	0.9 (0.1–1.17; 0.00)
MECDP	—	12 ± 2	—
dTDP-D-fucose/rhamnose	—	17 ± 6	—

Data represent means and SDs of three independent replicates. AICAR, 5-aminoimidazole-4-carboxamide ribonucleotide; dTDP, thymidine diphosphate; dTTP, thymidine triphosphate; FAD, flavin adenine dinucleotide; MECDP, 2-C-methyl-D-erythritol-2.4-cyclodiphosphate (*Shigella* metabolite not present in uninfected cells).

Table S3. Metabolite labeling after feeding 25 mM uniformly labeled ^{13}C glucose

			Labeling infected HeLa (%)					Labeling uninfected HeLa (%)				
Metabolite	No. of carbon	Mass shift	30 s	1 min	2 min	5 min	20 min	30 s	1 min	2 min	5 min	20 min
Embden–Meyerhof pathway												
Hexose 6-phosphate	6	+6	81 ± 2	88 ± 3	92 ± 2	97 ± 5	92 ± 1	71 ± 4	89 ± 2	95 ± 1	96 ± 1	97 ± 0
Hexose 1,6-bisphosphate	6	+6	41 ± 4	48 ± 5	54 ± 2	69 ± 3	79 ± 2	49 ± 3	75 ± 4	82 ± 3	91 ± 1	85 ± 2
Triose phosphate (DHAP, GAP)	3	+3	38 ± 3	55 ± 8	68 ± 16	78 ± 4	91 ± 3	50 ± 8	79 ± 9	91 ± 3	97 ± 1	97 ± 2
Glycerate bisphosphate	3	+3	40 ± 15	48 ± 14	78 ± 17	86 ± 3	89 ± 3	45 ± 10	84 ± 3	90 ± 5	95 ± 2	99 ± 1
Glycerate phosphate	3	+3	45 ± 9	74 ± 9	78 ± 14	90 ± 1	92 ± 1	49 ± 16	72 ± 7	88 ± 5	98 ± 1	98 ± 2
PEP	3	+3	17 ± 13	63 ± 18	84 ± 3	85 ± 4	85 ± 6	27 ± 4	48 ± 3	73 ± 4	99 ± 0	98 ± 3
Pyruvate	3	+3	45 ± 14	73 ± 6	78 ± 12	77 ± 4	87 ± 6	60 ± 4	83 ± 14	78 ± 1	75 ± 12	78 ± 4
Acetyl-CoA	23	+2	37 ± 8	45 ± 9	55 ± 6	58 ± 3	56 ± 11	2 ± 1	4 ± 1	7 ± 3	8 ± 1	4 ± 2
Pentose phosphate pathway (and related metabolites)												
Gluconate 6-phosphate	6	+5	1 ± 2	6 ± 7	6 ± 4	5 ± 3	6 ± 4	1 ± 1	4 ± 2	6 ± 3	8 ± 3	9 ± 4
		+6	51 ± 9	64 ± 9	80 ± 5	86 ± 5	88 ± 2	33 ± 8	66 ± 10	79 ± 6	86 ± 5	88 ± 4
Pentose 5-phosphate	5	+5	5 ± 1	9 ± 1	13 ± 2	14 ± 3	24 ± 8	28 ± 1	67 ± 4	79 ± 1	78 ± 9	90 ± 2
Ribose 5-phosphate	5	+5	0 ± 0	0 ± 0	1 ± 1	3 ± 2	15 ± 4	7 ± 3	33 ± 11	56 ± 2	57 ± 50	81 ± 4
Sedoheptulose 7-phosphate	7	+7	4 ± 6	35 ± 4	44 ± 12	68 ± 14	73 ± 23	2 ± 2	28 ± 4	76 ± 6	86 ± 5	83 ± 3
Sedoheptulose phosphate	7	+7	0± 0	9 ± 10	24 ± 6	30 ± 5	43 ± 15	0 ± 0	3 ± 3	15 ± 2	24 ± 3	36 ± 1
Pentose 1,5-bisphosphate	5	+5	5± 1	12 ± 2	14 ± 2	14 ± 1	29 ± 4	9 ± 5	49 ± 7	70 ± 2	90 ± 5	92 ± 6
Sedoheptulose bisphosphate	7	+7	16 ± 15	43 ± 12	71 ± 14	85 ± 10	84 ± 9	1 ± 13	49 ± 14	87 ± 3	90 ± 3	92 ± 0
TCA cycle												
Citrate	6	+2	1 ± 0	1 ± 1	2 ± 1	3 ± 2	3 ± 2	0 ± 1	1 ± 1	0 ± 0	0 ± 0	1 ± 0
		+6	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	2 ± 1	0 ± 1	0 ± 0	0 ± 0	0 ± 1
		int	2 ± 2	2 ± 0	2 ± 1	3 ± 1	3 ± 2	1 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0
α-Ketoglutarate	5	+2	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		+5	2 ± 1	6 ± 3	9 ± 1	18 ± 2	30 ± 4	0 ± 0	0 ± 0	0 ± 0	2 ± 0	21 ± 1
		int	2 ± 1	6 ± 2	9 ± 10	19 ± 7	31 ± 4	0 ± 0	0 ± 1	0 ± 1	2 ± 2	22 ± 11
Succinate	4	+2	0 ± 0	0 ± 0	0 ± 0	0 ± 1	2 ± 1	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		+4	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		int	1 ± 0	1 ± 0	1 ± 0	1 ± 0	2 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0
Fumarate	4	+2	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		+4	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		int	0 ± 0	0 ± 0	0 ± 0	0 ± 0	2 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 0
Malate	4	+2	0 ± 0	1 ± 0	1 ± 0	0 ± 0	1 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 0	2 ± 0
		+4	0 ± 0	0 ± 0	1 ± 1	3 ± 3	8 ± 5	0 ± 0	0 ± 0	0 ± 1	0 ± 0	2 ± 0
		int	1 ± 0	1 ± 0	2 ± 0	4 ± 2	11 ± 3	1 ± 0	1 ± 1	1 ± 0	1 ± 1	6 ± 2
Other												
AMP	10	+5	0 ± 0	0 ± 0	1 ± 2	8 ± 1	18 ± 3	0 ± 0	0 ± 0	0 ± 0	1 ± 2	1 ± 0
ADP	10	+5	1 ± 1	3 ± 3	1 ± 1	6 ± 5	19 ± 1	0 ± 0	0 ± 0	4 ± 6	1 ± 1	2 ± 0
ATP	10	+5	1 ± 1	0 ± 0	0 ± 1	5 ± 5	6 ± 6	0 ± 0	0 ± 0	1 ± 1	0 ± 0	0 ± 2
GMP	10	+5	0 ± 0	1 ± 1	2 ± 1	6 ± 2	20 ± 7	0 ± 1	1 ± 1	2 ± 3	1 ± 1	7 ± 4
GDP	10	+5	0 ± 0	0 ± 0	1 ± 1	7 ± 1	14 ± 1	0 ± 0	0 ± 0	2 ± 2	0 ± 0	5 ± 0
GTP	10	+5	0 ± 0	0 ± 0	1 ± 2	9 ± 3	20 ± 6	0 ± 0	0 ± 0	0 ± 3	0 ± 0	6 ± 0
CDP	9	+5	0 ± 0	0 ± 0	0 ± 0	5 ± 4	13 ± 5	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
UMP	9	+5	1 ± 0	4 ± 0	1 ± 1	11 ± 1	16 ± 5	0 ± 0	0 ± 0	3 ± 0	1 ± 0	5 ± 1

Table S3. Cont.

Metabolite	No. of carbon	Mass shift	Labeling infected HeLa (%)					Labeling uninfected HeLa (%)				
			30 s	1 min	2 min	5 min	20 min	30 s	1 min	2 min	5 min	20 min
UTP	9	+5	0 ± 0	0 ± 1	1 ± 2	6 ± 4	12 ± 3	0 ± 0	0 ± 0	2 ± 3	1 ± 0	9 ± 0
2-Deoxy-ATP	10	+5	0 ± 0	0 ± 0	0 ± 0	0 ± 0	26 ± 14	0 ± 0	0± 0	0 ± 0	0 ± 0	0 ± 2
AICAR	9	+3	3 ± 4	2 ± 3	11 ± 5	8 ± 2	7 ± 5	3 ± 1	10 ± 3	7 ± 1	16 ± 3	6 ± 2
		+5	0 ± 0	8 ± 4	10 ± 16	58 ± 7	65 ± 8	0 ± 0	0 ± 0	18 ± 16	17 ± 7	84 ± 1
dTDP-D-fucose/rhamnose	16	+6	31 ± 7	58 ± 5	64 ± 10	73 ± 4	57 ± 13	—	—	—	—	—
		+11	0 ± 0	0 ± 0	0 ± 0	4 ± 4	27 ± 9	—	—	—	—	—
Hexitol phosphate	6	+6	30 ± 5	49 ± 11	45 ± 8	67 ± 6	76 ± 5	—	—	—	—	—
Hexose 1-phosphate	6	+6	34 ± 10	47 ± 15	47 ± 11	56 ± 10	65 ± 9	15 ± 3	25 ± 4	36 ± 8	37 ± 9	49 ± 5
Triose phosphate	3	+3	38 ± 3	55 ± 8	68 ± 16	78 ± 4	91 ± 3	50 ± 8	79 ± 9	91 ± 3	97 ± 1	97 ± 2
UDP-hexose	15	+5	7 ± 3	14 ± 9	13 ± 6	20 ± 3	49 ± 4	1 ± 0	2 ± 1	11 ± 7	19 ± 2	59 ± 3
		+11	0 ± 0	0 ± 0	0 ± 1	3 ± 3	5 ± 3	0 ± 0	0 ± 0	1 ± 1	0 ± 1	1 ± 0
MECDP	5	+3	0 ± 0	0 ± 0	0 ± 0	2 ± 1	7 ± 6	—	—	—	—	—
		+5	1 ± 1	0 ± 0	1 ± 1	10 ± 7	47 ± 10	—	—	—	—	—

This table includes only metabolites that had 5% labeling at any time point. Additional unlabeled metabolites were detected (Tables S1 and S2 have full lists of all detected metabolites). Data represent means and SDs of six independent replicates. int, Integrated labeling; AICAR, 5-aminoimidazole-4-carboxamide ribonucleotide; DHAP, dihydroxyacetone phosphate; dTDP, thymidine diphosphate; dTTP, thymidine triphosphate; GAP, glyceraldehyde 3-phosphate; MECDP, 2-C-methyl-D-erythritol-2.4-cyclodiphosphate (*Shigella* metabolite not present in uninfected cells); PEP, phosphoenolpyruvate; TCA, tricarboxylic acid.

Table S4. Intracellular growth of *Shigella* mutants with deficiencies around the pyruvate node

	Shigella strain				
Name	Genotype*	Defective step	μ_{mut}/μ_{par} (n)	P value	q Value
HeLa					
SHP236	Δ*manXYZ* Δ*galP* Δ*ptsG* Δ*mglBAC uhpT::kan*	Glucose, mannose, hexose-P uptake	0.97 ± 0.11 (33)	0.16	n.s.
SHM10	Δ*ptsI* Δ*pykA pykF::kan*	PEP ↔ pyruvate	0.77 ± 0.10 (32)	0.0001	<0.001
SHP550	*ppsA::cam*	Pyruvate → PEP	1.07 ± 0.10 (21)	0.007	0.009
SHP555	*aceE::kan*	Pyruvate ↔ acetyl-CoA	0.39 ± 0.19 (43)	0.0002	<0.001
SHP578	*pflB pflD tdcE ybiYW yfiD*	Pyruvate ↔ acetyl-CoA	1.02 ± 0.07 (16)	0.21	n.s.
SHP437	*ldhA::kan lldD::cam*	Lactate ↔ acetyl-CoA	1.01 ± 0.14 (47)	0.78	n.s.
SHP374	*pta:kan*	Acetyl-CoA ↔ acetyl-P	0.62 ± 0.14 (56)	0.0001	<0.001
SHP410	*ackA::kan*	Acetyl-P ↔ acetate	0.41 ± 0.11 (34)	0.0001	<0.001
Caco-2					
SHP555	*aceE::kan*	Pyruvate ↔ acetyl-CoA	0.66 ± 0.30 (18)	0.0002	<0.001
SHP374	*pta:kan*	Acetyl-CoA ↔ acetyl-P	0.64 ± 0.24 (13)	0.0002	<0.001
HUVEC					
SHP555	*aceE::kan*	Pyruvate ↔ acetyl-CoA	0.21 ± 0.22 (6)	0.0003	<0.001
SHP374	*pta:kan*	Acetyl-CoA ↔ acetyl-P	0.51 ± 0.19 (8)	0.0001	<0.001

HUVEC, human umbilical vein endothelial cell; n.s., q value > 0.05; PEP, phosphoenolpyruvate.

*All mutants were derivatives of the *icsA* parental strain; μ_{mut}/μ_{par} is the growth rate of the mutant (μ_{mut}) relative to the parental strain (μ_{par}). Data represent means and SDs of n replicate culture wells. Statistical differences were analyzed by a two-tailed t test followed by Benjamini–Hochberg correction for multiple testing.

Table S5. Metabolite labeling with 5 mM uniformly labeled ^{13}C pyruvate/25 mM unlabeled glucose

			Labeling infected HeLa (%)					Labeling uninfected HeLa (%)				
Metabolite	No. of carbon	Mass shift	30 s	1 min	2 min	5 min	20 min	30 s	1 min	2 min	5 min	20 min
Embden–Meyerhof pathway												
Hexose 6-phosphate	6	+6	1 ± 0	1 ± 0	1 ± 0	1 ± 1	1 ± 1	1 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0
Hexose 1,6-bisphosphate	6	+6	2 ± 4	0 ± 0	0 ± 0	7 ± 11	0 ± 0	1 ± 0	1 ± 0	1 ± 1	0 ± 0	0 ± 0
Triose phosphate (DHAP, GAP)	3	+3	0 ± 0	0 ± 0	1 ± 1	0 ± 0	0 ± 1	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Glycerate bisphosphate	3	+3	9 ± 3	14 ± 1	23 ± 3	27 ± 18	54 ± 4	12 ± 2	24 ± 2	41 ± 1	56 ± 3	65 ± 3
Glycerate phosphate	3	+3	2 ± 1	1 ± 1	1 ± 1	0 ± 1	1 ± 1	0 ± 1	0 ± 0	1 ± 1	0 ± 0	6 ± 11
PEP	3	+3	1 ± 1	1 ± 1	1 ± 1	0 ± 1	0 ± 0	2 ± 4	1 ± 3	1 ± 3	0 ± 0	1 ± 1
Acetyl-CoA	23	+2	64 ± 9	60 ± 12	62 ± 10	72 ± 9	71 ± 7	16 ± 11	27 ± 6	39 ± 8	71 ± 8	74 ± 8
Pentose phosphate pathway (and related metabolites)												
Gluconate 6-phosphate	6	+6	1 ± 1	0 ± 0	1 ± 1	1 ± 1	1 ± 0	0 ± 0	1 ± 0	0 ± 0	0 ± 0	0 ± 0
Pentose 5-phosphate	5	+5	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 2	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Ribose 5-phosphate	5	+5	1 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Sedoheptulose 7-phosphate	7	+7	2 ± 0	2 ± 0	1 ± 0	1 ± 0	0 ± 0	0 ± 0	1 ± 0	1 ± 0	1 ± 0	1 ± 0
Sedoheptulose phosphate	7	+7	2 ± 1	2 ± 0	2 ± 0	2 ± 1	2 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0
Pentose 1,5-bisphosphate	5	+5	1 ± 1	1 ± 1	1 ± 0	1 ± 1	1 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Sedoheptulose bisphosphate	7	+7	2 ± 1	2 ± 0	2 ± 0	2 ± 1	2 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0	2 ± 0
TCA cycle												
Citrate	6	+2	10 ± 2	14 ± 1	23 ± 3	32 ± 6	43 ± 6	13 ± 2	24 ± 1	38 ± 1	51 ± 2	52 ± 3
		+6	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		int	5 ± 1	2 ± 0	2 ± 1	3 ± 1	3 ± 2	1 ± 0	1 ± 0	1 ± 0	1 ± 0	1± 0
α-Ketoglutarate	5	+2	1 ± 0	1 ± 0	2 ± 0	6 ± 0	18 ± 2	1 ± 0	1 ± 0	2 ± 0	6 ± 2	17 ± 0
		+5	2 ± 1	6 ± 3	9 ± 1	18 ± 2	30 ± 4	0 ± 0	0 ± 0	0 ± 0	2 ± 0	21 ± 1
		int	0 ± 0	0 ± 0	1 ± 0	2 ± 0	11 ± 1	0 ± 0	0 ± 0	1 ± 0	2 ± 1	9 ± 1
Succinate	4	+2	0 ± 0	0 ± 0	1 ± 0	3 ± 1	9 ± 2	0 ± 0	0 ± 0	0 ± 0	2 ± 1	7 ± 1
		+4	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		int	1 ± 0	1 ± 0	1 ± 0	2 ± 1	6 ± 1	1 ± 0	1 ± 0	1 ± 0	2 ± 0	4 ± 0
Fumarate	4	+2	0 ± 0	0 ± 0	0 ± 1	1 ± 1	7 ± 1	1 ± 0	1 ± 2	1 ± 1	1 ± 1	3 ± 1
		+3	0 ± 0	1 ± 1	1 ± 1	2 ± 1	6 ± 1	0 ± 0	0 ± 0	0 ± 0	2 ± 1	5 ± 1
		+4	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
		int	0 ± 0	1 ± 0	1 ± 0	2 ± 1	8 ± 1	1 ± 0	1 ± 1	1 ± 0	2 ± 1	6 ± 1
Malate	4	+2	1 ± 0	1 ± 0	1 ± 0	2 ± 0	10 ± 2	1 ± 0	1 ± 0	1 ± 0	2 ± 0	7 ± 1
		+3	2 ± 0	3 ± 1	3 ± 0	6 ± 1	8 ± 2	1 ± 0	2 ± 0	2 ± 0	5 ± 1	8 ± 1
		+4	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1 ± 0
		int	2 ± 0	3 ± 0	4 ± 0	6 ± 0	13 ± 3	2 ± 0	3 ± 0	3 ± 0	5 ± 1	11 ± 1
Other												
Aconitic acid	6	+2	4 ± 2	10 ± 2	18 ± 4	25 ± 10	38 ± 11	8 ± 2	18 ± 3	32 ± 5	43 ± 13	41 ± 13
Mesaconic acid	5	+4	0 ± 0	0 ± 0	0 ± 0	0 ± 1	5 ± 2	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
MECDP	5	+2	1 ± 3	4 ± 4	6 ± 4	10 ± 9	53 ± 4	—	—	—	—	—

This table includes only metabolites that had 5% labeling at any time point. Additional unlabeled metabolites were detected (Tables S1 and S2 have full lists of all detected metabolites). Data represent means and SDs of three independent replicates. int, Integrated labeling; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; MECDP, 2-C-methyl-D-erythritol-2.4-cyclodiphosphate (*Shigella* metabolite not present in uninfected cells); PEP, phosphoenolpyruvate; TCA, tricarboxylic acid.

Table S6. Intracellular growth of *Shigella* nutrient utilization mutants

Strain name	Genotype*	Utilization defect (defective for use of...)	μ_{mut}/μ_{par} (n)	P value	q Value
SHP078	*amtB::kan*	Ammonium	1.03 ± 0.04 (10)	0.08	n.s.
SHP079	*lldP::kan yghK::cam*	Lactate	0.95 ± 0.05 (9)	0.01	n.s.
SHP084	*glnHPQ::kan*	Glutamine	1.00 ± 0.05 (8)	0.82	n.s.
SHP126	*uhpT::kan*	Hexose phosphate	1.03 ± 0.13 (21)	0.31	n.s.
SHP127	Δ*rbsC* Δ*rbsD xylFGH::kan*	D-RIBOSE	1.02 ± 0.06 (9)	0.31	n.s.
SHP128	*sdaC::kan*	Serine	0.97 ± 0.08 (7)	0.36	n.s.
SHP129	*kgtP::kan*	α-Ketoglutarate	1.04 ± 0.08 (4)	0.38	n.s.
SHP130	Δ*ytfRT* Δ*galP fucP::cam mglBAC::kan*	Galactose	1.01 ± 0.11 (18)	0.61	n.s.
SHP155	Δ*gltJKL gltS::kan gltP::cam*	Glutamate	0.94 ± 0.08 (10)	0.04	n.s.
SHP168	Δ*manXYZ* Δ*galP ptsG::cam mglBAC::kan*	Glucose and mannose	1.02 ± 0.09 (21)	0.41	n.s.
SHP169	*ptsI:kan*	PTS carbohydrates	1.03 ± 0.07 (12)	0.17	n.s.
SHP171	Δ*fucP* Δ*manXYZ fruBKA::kan*	Fructose	1.05 ± 0.10 (4)	0.42	n.s.
SHP196	Δ*yhcL* Δ*dctA* Δ*dcuC* Δ*dcuA dcuB::kan*	C4 dicarboxylates	0.96 ± 0.08 (9)	0.19	n.s.
SHP240	Δ*gntU* Δ*gntP kdgT::cam gntT::kan*	Gluconate	1.01 ± 0.05 (11)	0.72	n.s.
SHP263	*nupG::kan nupC::cam*	Nucleosides	0.96 ± 0.09 (7)	0.26	n.s.
SHP384	Δ*glpT ugpBAEQ::kan*	Glycerol-phosphate	0.95 ± 0.12 (31)	0.02	n.s.
SHP280	Δ*ggt gsiABCD::kan*	Glutathione	0.94 ± 0.11 (9)	0.13	n.s.
SHP492	*fadK::cam fadD::kan*	Fatty acid	0.90 ± 0.05 (12)	0.0001	0.002
SHP408	*citE::kan*	Citrate	1.01 ± 0.07 (24)	0.63	n.s.

n.s., q value > 0.05.

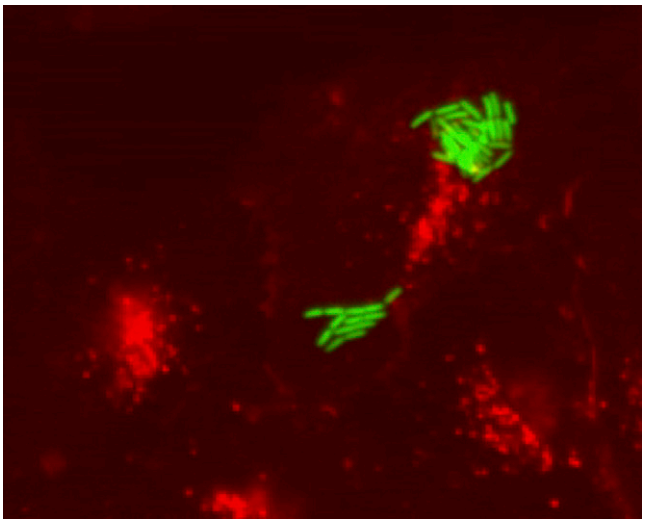
*All mutants were derivatives of the *icsA* parental strain; μ_{mut}/μ_{par} is the growth rate of the mutant (μ_{mut}) relative to the parental strain (μ_{par}). Data represent means and SDs of n replicate culture wells. Statistical differences were analyzed by two-tailed t test followed by Benjamini–Hochberg correction for multiple testing. PTS, phosphotransferase system.

Table S7. Intracellular growth of auxotrophic mutants

Strain name	Genotype*	Biosynthesis defect (auxotrophic for...)	μ_{mut}/μ_{par} (n)	P value	q Value
SHP300	*glnA::cam*	Glutamine	1.06 ± 0.11 (10)	0.12	n.s.
SHP239	*hisG::kan*	Histidine	0.93 ± 0.10 (10)	0.051	n.s.
SHP259	*serA::kan*	Serine	1.02 ± 0.10 (6)	0.58	n.s.
SHP260	*asnB::kan asnA::cam*	Asparagine	0.49 ± 0.11 (32)	<0.0001	<0.001
SHP261	*thrC::kan*	Threonine	0.99 ± 0.12 (4)	0.91	n.s.
SHP270	*argH::kan lysA::cam*	Arginine, lysine	1.03 ± 0.12 (4)	0.68	n.s.
SHP305	*metA::cam*	Methionine	1.01 ± 0.03 (10)	0.51	n.s.
SHP336	*ilvE::kan leuB::cam*	Leucine, valine, isoleucine	1.03 ± 0.08 (10)	0.27	n.s.
SHP375	Δ*tyrA trpA::kan pheA::cam*	Phenylalanine, tyrosine, tryptophan	0.97 ± 0.12 (10)	0.34	n.s.
SHP475	*proC::kan*	Proline	0.56 ± 0.14 (10)	<0.0001	<0.001
SHP198	*purH::kan*	Purines	0.86 ± 0.06 (9)	<0.0001	<0.001

n.s., q value > 0.05.

*All mutants were derivatives of the *icsA* parental strain; μ_{mut}/μ_{par} is the growth rate of the mutant (μ_{mut}) relative to the parental strain (μ_{par}). Data represent means and SDs of n replicate culture wells. Statistical differences were analyzed by two-tailed t test followed by Benjamini–Hochberg correction for multiple testing.



Movie S1. Intracellular growth of GFP-expressing *Shigella* (green) in HeLa cells stained with Wheat Germ Agglutination-Texas Red (red).

Movie S1

3.3 Direct Observation of Metabolic Events in Living Macrophages by Dissolution Dynamic Nuclear Polarization NMR Spectroscopy

Abstract

Within the last ten years, the development of dissolution dynamic nuclear polarization (d-DNP) NMR spectroscopy provided the sensitivity to follow real time metabolic processes in the cell. So far, it was successfully applied to several systems such as *E. coil* or *S. cerevisiae*, allowing the characterization of metabolic pathways such as glycolysis or the TCA cycle. Here, for the first time, dissolution-DNP NMR spectroscopy was applied to macrophages. We show that hyperpolarized [*U*-^{2}H,-^{13}C]-glucose and $^{13}C_1$-pyruvate metabolisation can be followed in real time. The end products of metabolic pathways are detected, at increasing concentration with time, in the NMR spectra. Our findings show that dissolution-DNP NMR spectroscopy can be applied to a broad range of cellular systems, and can become routinely applied for metabolic studies in the cell. In the case of macrophages, we aim to characterize the metabolic changes induced by activation from a basal state to the M1 phenotype.

Introduction

Standard NMR spectroscopy is a powerful tool to characterize metabolites and metabolic flux in the cell, allowing non-invasive measurement of cellular compounds[1-9]. However, the intrinsically low sensitivity of the method generally limits its use to specific cell lines or highly concentrated cell media. To overcome this issue, a range of methods was proposed to increase the nuclear spin polarization, called hyperpolarization methods. In particular optical pumping[10], para-hydrogen-induced polarization[11] and dynamic nuclear polarization[12] were shown to increase polarization close to unit but are still limited to specific samples and applications. In the last ten years, dissolution dynamic nuclear polarization (d-DNP) NMR spectroscopy was developed, enabling close to unit polarization of organic molecules in a liquid solution[13] by hyperpolarization of the nuclear spins in the solid state, transferring the high electron polarization to other nuclei using microwave irradiation at low temperature, followed by rapid dissolution. This method enables to bring the polarized cold sample into a liquid state preserving its nuclear polarization[14], enhancing the single-scan liquid-state ^{13}C NMR signal-to-noise ratio by factors of 10,000 or more[13]. Dissolution-DNP NMR spectroscopy has high potential in the metabolomics filed, where administration of hyperpolarized metabolites with ^{13}C enrichment to cell suspensions has been used to characterize metabolism associated with glycolysis of *S. cerevisiae*[15], *E. coli*[16, 17], breast cancer cells[18] or perfused rat heart[19]. Recent developments, as the use of deuterated molecules or cross polarization transfer to increase polarization levels and lifetime[20], extends its use to a large range of molecular processes in the cell.

Here we applied for the first time dissolution-DNP NMR spectroscopy to living macrophages. Macrophages are the “big eaters” of the immune system that engulf apoptotic cells and pathogens and produce immune effector molecules in a process called phagocytosis[21]. Upon infection, they undergo specific activation profiles from a basal to a classical state (M1) encouraging inflammation and stimulating the immune system or an alternative state (M2) decreasing inflammation and encouraging tissue repair, modifying their metabolism[22, 23]. Metabolic changes associated to the activation to a M1 phenotype (Fig. 3.2) are induced by bacterial derived products such as lipopolysaccharides[24] (LPS), as well as by signals associated with infection from which interferon gamma (IFN-γ) is among the most important[25]. Mainly, it down-regulates mitochondrial oxidation[22] while glycolysis is up-regulated resulting in high rates of glucose consumption and conversion of pyruvate to lactate[26, 27]. As well, the pro-inflammatory cytokine production is up regulated[28]. Altogether, these metabolic events provide rapid energy and reducing equivalents to the cell, which are required for bactericidal activity.

Here, we show, for the first time, that d-DNP NMR could be used to study real-time metabolism of living macrophages, using [*U*-^{2}H,-^{13}C]-glucose and $^{13}C_1$-pyruvate as hyperpolarized nutrients.

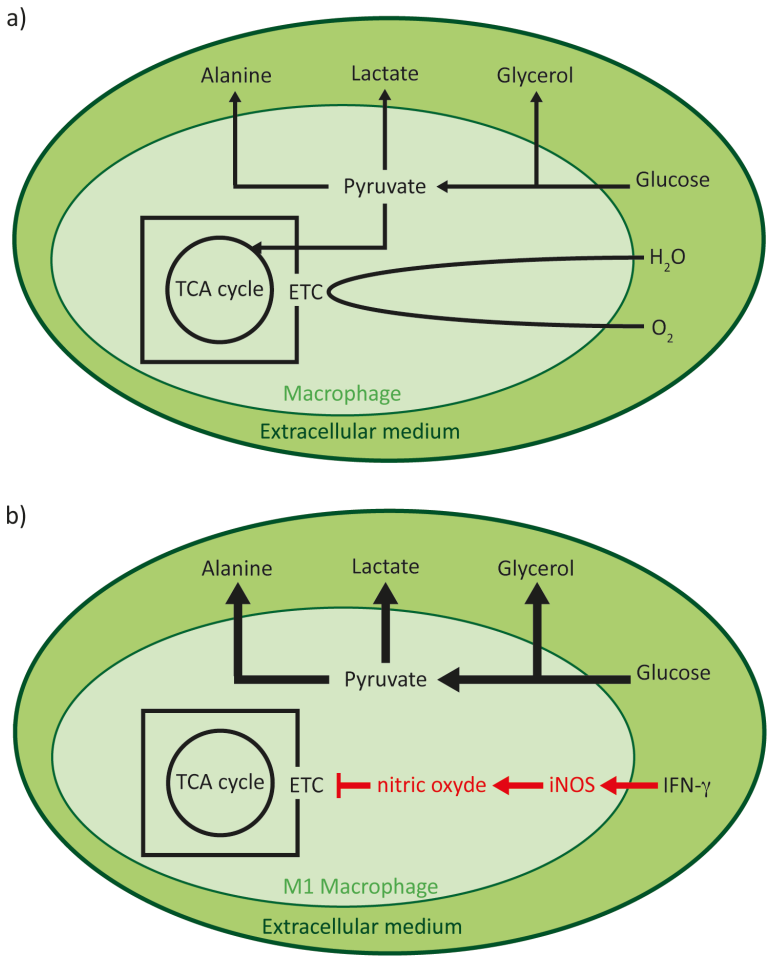


Figure 3.2. Overview of the macrophages metabolism. a) Simplified view of macrophage metabolism in the basal state. b) Simplified view of macrophage metabolism in the M1 state. The up-regulation of the glycolytic pathway is indicated by bold arrows.

Results

The short T_1 of [*U*-^{13}C]-glucose limits its use as a hyperpolarized nutrient

In an attempt to visualize the different metabolites of the glycolytic pathway in macrophages, a similar procedure than previously developed by Meier *et al.* was applied[15]. Hyperpolarized

glucose, efficiently incorporated and metabolized by the cell at a rate of 1 – 3 nmol / min / million cell, was used as a nutrient.

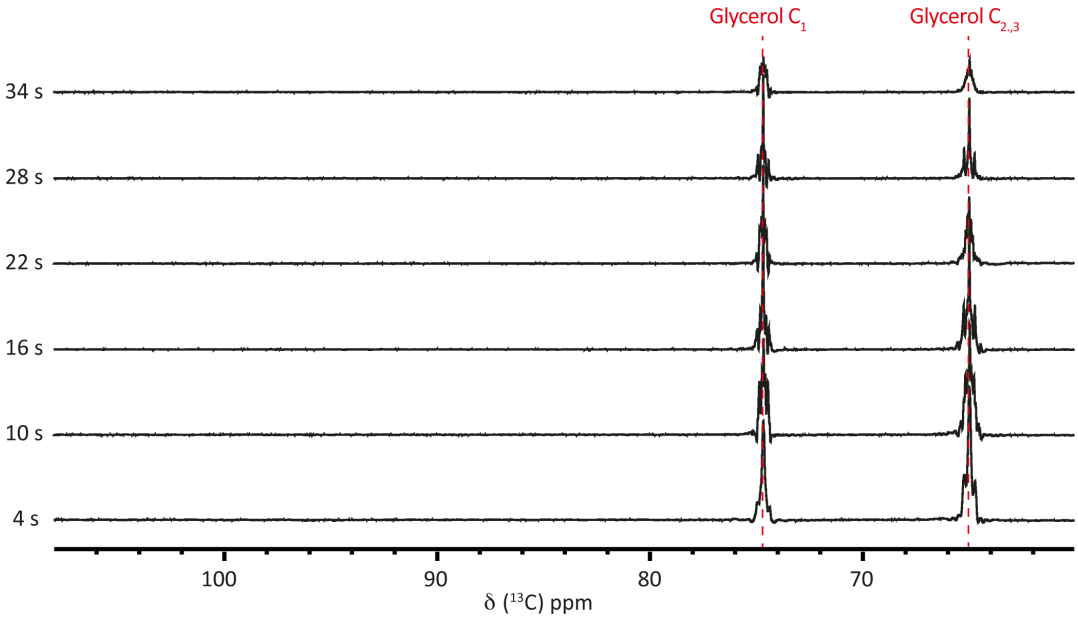


Figure 3.3. Zoom of time series of 1D ^{13}C spectra of living macrophages fed with hyperpolarized [*U*-^{13}C]-glucose, recorded every 6 seconds. The lower spectrum is recorded 4 seconds after injection of hyperpolarized [*U*-^{13}C]-glucose. Glycerol, as an excreted product, is highlighted in red.

867 mM *U*-^{13}C glucose was hyperpolarized in the solid state and subsequently quickly dissolved by superheated D_2O. After dissolution, the hyperpolarized sample was automatically injected into a suspension of 80 million macrophages in the basal state, residing in a 500 MHz NMR spectrometer in order to detect chemical transformation of the substrate by the cells. Subsequently, a series of 1D ^{13}C spectra was recorded every second for a total time of 256 seconds. The spectra displayed glycerol resonances only, used as a glassing agent during the hyperpolarized sample preparation (Fig. 3.3). The absence of detectable hyperpolarized signal, besides glycerol, was attributed to the fast T_1 of the protonated *U*-^{13}C glucose, of about 6 seconds, leading to a rapid loss of the hyperpolarization by cross-relaxation via protons, before reaching the cell suspension. In comparison, the perdeuterated glycerol displayed a longer T_1 allowing its detection in the NMR spectra.

Hyperpolarized [*U*-^{2}H,-^{13}C] glucose as a nutrient to follow glycolysis cycle in macrophages

In order to provide a sufficiently long T_1 for d-DNP NMR measurement, [*U*-^{2}H,-^{13}C]-glucose, displaying a T_1 two times longer than the protonated version, was used as hyperpolarized nutrient for macrophages. The same experiment was performed, where 867 mM [*U*-^{2}H,-^{13}C]-glucose was hyperpolarized in the solid state, dissolved by superheated D_2O and automatically injected into a suspension of 80 million macrophages in the basal state, residing in a 500 MHz NMR spectrometer. Subsequently, a series of 1D ^{13}C spectra was recorded at 37°C during 256 seconds after dissolution (Fig 3.4a).

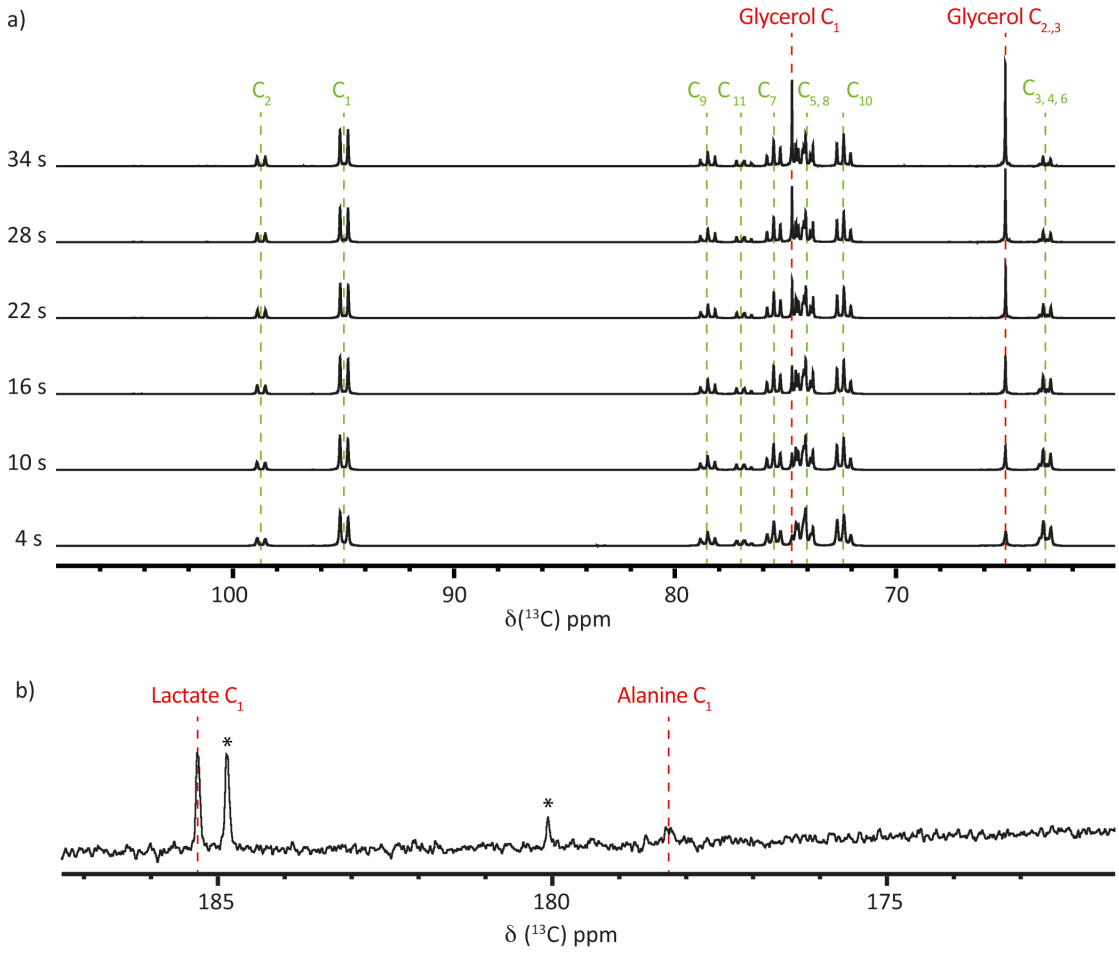


Figure 3.4. Macrophages fed with hyperpolarized [*U*-^{2}H,-^{13}C]-glucose. a) Zoom of time series of 1D ^{13}C spectra of living macrophages fed with hyperpolarized [*U*-^{2}H,-^{13}C]-glucose, recorded every 6 seconds, normalized to the signal of glucose. The lower spectrum is recorded 4 seconds after injection of hyperpolarized [*U*-^{2}H,-^{13}C]-glucose. Glucose resonances are highlighted in green and excreted products in red. b) Zoom of 1D ^{13}C spectra of living macrophages fed with hyperpolarized [*U*-^{2}H,-^{13}C]-glucose, recorded at thermal equilibrium 30 min after injection of hyperpolarized [*U*-^{2}H,-^{13}C]-glucose. Resonances labeled with * correspond to impurities.

There the resonances of the hyperpolarized glucose were detected as well as glycerol derived from glucose, possibly through the conversion of the glycolysis metabolite dihydroxyacetone phosphate via glycerol-3-phosphate. The increase of the glycerol resonances over time indicated its accumulation in the sample. Weak resonances, undergoing intensity increase over time were detected as well at 67.8, 69.5, 81, 83.5, 101.2, 105.5 and 106.5 ppm (data not shown) but could not be assigned to any of the metabolites that are part of the glycolytic pathway. 1D ^{13}C spectrum recorded on the macrophages suspension at "thermal equilibrium", in the absence of hyperpolarization, 30 minutes after dissolution (Fig. 3.4b), displayed resonances from lactate and alanine.

Intermediate metabolites of the glycolytic pathway are not detected in this experiment because of their low concentration, as they are constantly produced and metabolized by the cells, while the end product glycerol is accumulating over time. Surprisingly, lactate was not detected during the first 90 seconds of measurement whereas it appeared as an end product. This observation indicates a longer and / or less efficient conversion of glucose to pyruvate / lactate than to glycerol. Comparison to different cell systems indicates that the glycolytic pathway in macrophages is longer than in *E.coli* or *S. cerevisiae* where pyruvate and lactate production were observed during the first seconds of measurements[15, 16].

Hyperpolarized $^{13}C_1$-pyruvate used as a nutrient by macrophages

In an attempt to visualize the downstream metabolites of the glycolytic pathway, $^{13}C_1$-pyruvate was used as a hyperpolarized nutrient. Its T_1 of 70 seconds, 6 times longer than [*U*-^{2}H,-^{13}C]-glucose, enables long measurement times of the hyperpolarized compounds. 1.5 M $^{13}C_1$-pyruvate was hyperpolarized to 28 % in the solid state and subsequently quickly dissolved by superheated D_2O. After dissolution, the hyperpolarized sample was automatically injected into a suspension of 80 million macrophages in the basal state, residing in a 500 MHz NMR spectrometer in order to detect chemical transformation of the substrate by the cells. Subsequently, a series of 1D ^{13}C spectra was recorded every second for a total time of 1280 second. Figure 3.5 displays the sum of the initial 30 1D ^{13}C spectra recorded on macrophage suspension fed with hyperpolarized pyruvate. Resonances of pyruvate C_1 (172.9 ppm) and ^{13}C natural abundant pyruvate C_2 (207.9 ppm) were detected as well as pyruvate hydrate (181.2 ppm) produced from pyruvate in our setup condition[29]. Resonances of lactate C_1 (185.1 ppm), alanine C_1 (178.5 ppm), bicarbonate (HCO_3^-, 160.6 ppm) and CO_2 (125.2 ppm) were visible in the NMR spectra. The conversion of pyruvate signal into several products was detected in real time

and occurred within few seconds (Fig. 3.6a). Metabolites that were quickly formed and yielded signals were lactate C_1, alanine C_1, bicarbonate and CO_2 directly produced from pyruvate. They were detected only 10 seconds after $^{13}C_1$-pyruvate injection, indicating its conversion to downstream products of about that time. Lactate approached its maximum signal level within 80 seconds after pyruvate injection (Fig. 3.6b). As an end product, it should accumulate inside and outside the cell, and its decrease in concentration observed in the NMR spectra after 80 seconds could indicates a metabolic change in the NMR tube leading to a consumption of lactate by the cells. Alanine, bicarbonate and CO_2 were formed considerably slower and in fewer amounts than lactate and their production was observed during all the measurement time frame.

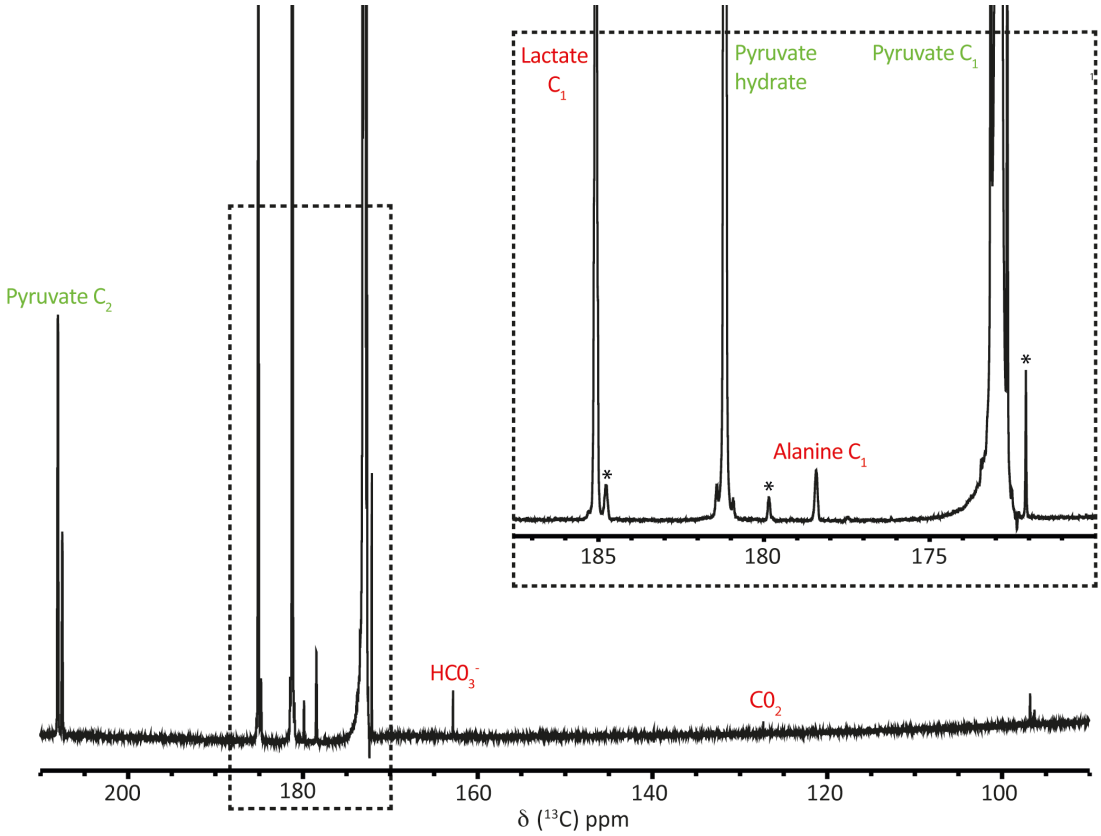


Figure 3.5. Metabolism of macrophages suspension fed with hyperpolarized pyruvate. Sum of the initial 30 1D ^{13}C spectra recorded on living macrophage fed with hyperpolarized $^{13}C_1$-pyruvate. Excreted products are highlighted in red and resonances of the hyperpolarized nutrient in green. Resonances labeled with * correspond to impurities.

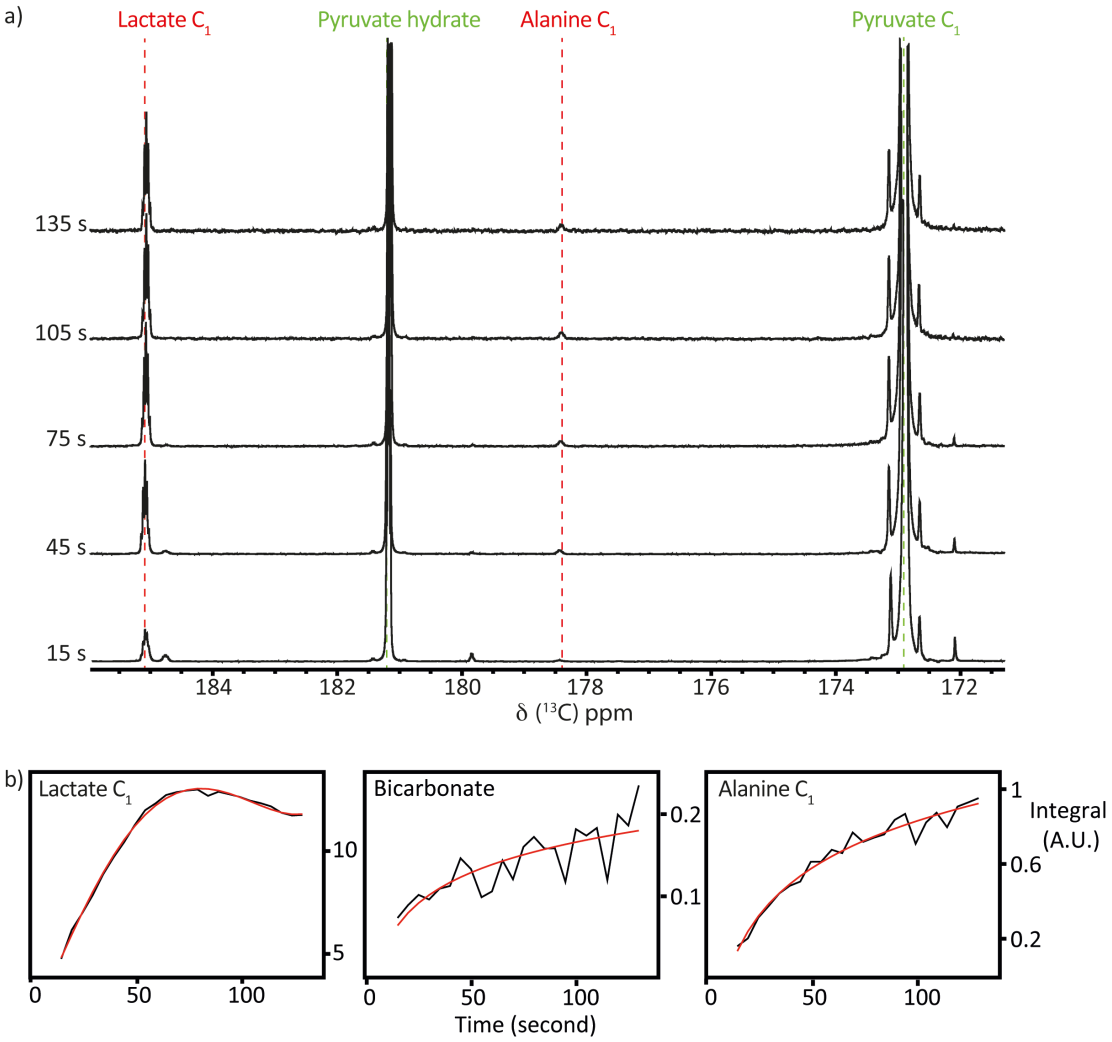


Figure 3.6. Metabolism of living macrophages fed with hyperpolarized $^{13}C_1$-pyruvate. a) Zoom of time series of 1D ^{13}C spectra of living macrophages fed with hyperpolarized $^{13}C_1$-pyruvate, recorded every 30 seconds, normalized to the signal of pyruvate C_1. The lower spectrum is recorded 15 seconds after injection of hyperpolarized $^{13}C_1$-pyruvate. Excreted products are highlighted in red and resonances of the hyperpolarized nutrient in green. b) Evolution of the metabolites produced from $^{13}C_1$-pyruvate in living macrophages over time normalized to the signal of pyruvate C_1. The evolution of CO_2 is not shown due to its low intensity, close to the noise level. The red lines correspond to the fitting curve.

Discussion

Within the last ten years, hyperpolarization NMR allows to monitor real time metabolites in a large range of cell systems[15-19]. Compared to standard NMR measurements, it enables the observation of fast metabolite processes and increases the sensitivity of the NMR experiments

of at least a factor 10,000 compared to standard ^{13}C measurements[13]. Recently, the development of new polarization techniques and the use of specifically labeled nutrients to be hyperpolarized extended its use to a broad range of metabolic processes in living cells[20]. Here dissolution-DNP NMR spectroscopy was applied for the first time to study the metabolism of living macrophages in real time. The use of deuterated nutrients, necessary for measurement of hyperpolarized metabolites signal was demonstrated, where [*U*-^{2}H,-^{13}C]-glucose and $^{13}C_1$-pyruvate allowed the detection of downstream metabolite products in real time in the cell suspension. Mainly, the resonances observed in the 1D ^{13}C NMR spectra corresponded to end products of metabolic pathways, where their accumulation at high amounts in the cell suspension allowed their detection. The choice of nutrient was related to the metabolite pathway to be characterized due to the time limitation of the hyperpolarization compared to the length of the compounds metabolisation. Indeed, whether this technique allows studying fast metabolic processes and low concentrated products compared to standard NMR experiments, its use is restricted to the observation of events happening in the second range due to the fast decay of the hyperpolarized signal through cross-relaxation. Recently, Carravetta *et al.* described long-lived states (LLS), involving pairs of inequivalent spins that can be used to store the hyperpolarized magnetization, exhibiting lifetimes up to 37 times T_1[30], possibly extending the d-DNP NMR measurement window to the minute range. Several improvements, among which LLS could be tested and included in our proposed setup, may lead further to the possible characterization of metabolic changes in macrophages upon activation to a M1 phenotype.

Material and Methods

Macrophages biochemistry

Macrophages were grown in 175 cm^2 flask containing 100 mL growing medium (DMEM, FCS 10%, Glutamine, bicarbonate, 25 mM Glucose) at 37°C under 5 % CO_2. The growing medium was exchanged every 24 hours following the hereinafter-described protocol. The macrophages growing medium was remove by pipetting and 25 mL of PBS (phosphate buffered saline) was added to the flask. PBS was then exchanged to 25 mL PBS/EDTA solution and incubated for 5 minutes at room temperature (RT). Macrophages were detached form the flask wall and centrifuged for 5 minutes at 200 g at RT. The cell pellet was resuspended in 10 mL growing medium containing 5 mM glucose. In order to count the cells, a macrophage sample was diluted 20 times in trypan blue solution and placed in a counting plate. The total number of cells (N_{tot})

was calculated from the average number of cells on each counting plate quadrant ($N_{average}$) using the following formula where d is the dilution factor of the cells in trypan blue and V the volume of the cell sample.

$$N_{tot} = N_{average} \cdot d \cdot V \cdot 10^4$$

A sample volume corresponding to $80 \cdot 10^6$ macrophages was centrifuged 5 min at 200 g at RT, resuspended in 100 mL growing medium containing 5 mM glucose, placed in a 175 cm^2 flask and incubated for 24 hours at 37°C under 5% CO_2. NMR sample preparation was achieved following the above-described protocol containing the following changes: after cell count, a sample volume corresponding to $80 \cdot 10^6$ macrophages was centrifuged 5 min at 200 g at RT and resuspended in 250 mL of growing medium without glucose containing 30 μL D_2O.

Dissolution DNP NMR methods

1.5 M $^{13}C_1$-pyruvate (Cambridge Isotope Labs) was dissolved in a D_2O:*U*-^{2}H-glycerol solution at 1:1 ratio containing 25 mM TEMPOL (Sigma-Aldrich). 867 mM [*U*-^{2}H,-^{13}C]-glucose (Cambridge Isotope Labs) was dissolved in an H_2O:D_2O:*U*-^{2}H-glycerol solution at a 1:4:5 ratio containing 50 mM TEMPOL (Sigma-Aldrich). 10 and 5 10 μL flash-frozen sample beads were mixed with the same number of 10 μL flash-frozen beads of 3 M ascorbate in D_2O for the [*U*-^{2}H,-^{13}C]-glucose and $^{13}C_1$-pyruvate sample respectively and placed inside the sample holder. Microwave frequency modulation was set at $\Delta f_{mod} = 10$ kHz and $\Delta f_{\mu w} = 50$ MHz with a negative microwave irradiation of the ESR line of TEMPOL at 188.3 GHz. ^{13}C direct polarization was performed on the [*U*-^{2}H,-^{13}C]-glucose sample at 1.2 K and Hartmann Hahn Cross-polarization (CP) was performed on the $^{13}C_1$-pyruvate sample with a CP every 5 minutes and 2 ms contact pulse at 1.2 K. The sample was then quickly dissolved by superheated D_2O at a temperature of 400 K and a pressure of 10 bars. 400 μL of the dissolved sample were transferred, via a magnetic tunnel, in 4.5 seconds to a Bruker 500 MHz spectrometer pre-heated at 37°C containing a 400 μL macrophages suspension with D_2O to allow field-frequency locking prior to injection. ^{13}C NMR spectra were recorded as an array of 256 one-dimensional spectra using 10° flip angle. Each one-dimensional spectrum was recorded every 1 and 5 seconds and 46080 and 65536 data points were recorded in the time domain with a ^{13}C carrier centered at 113 and 130 ppm for the [*U*-^{2}H,-^{13}C]-glucose and $^{13}C_1$-pyruvate sample respectively. Spectra were processed with an exponential line broadening of 3 Hz, zero-filled to 65536 points and Fourier transformed in Topspin 3.2. Data

analysis was performed in Topspin 3.2. Spectra were normalized using the glucose and pyruvate signal.

1D ^{13}C NMR spectra at thermal equilibrium were recorded 30 minutes after dissolution in a total experiment time of 30 minutes. The ^{13}C carrier was centered at 113 and 130 ppm for the [*U*-^{2}H,-^{13}C]-glucose and $^{13}C_1$-pyruvate sample respectively. The interscan delay was set to 200 ms. In the direct dimension, 65536 complex points were recorded in an acquisition time of 1.33 s, multiplied with an exponential line broadening of 3 Hz, zero-filled to 65536 points and Fourier transformed in Topspin 3.2. Data analysis was performed in Topspin 3.2.

Acknowledgements

The cell biology was performed in collaboration with Mauricio Rosas-Ballina in the group of Dirk Bumann. The DNP NMR experiments were performed by Basile Vuichoud and Aurélien Bornet in the group of Geoffrey Bodenhausen at EPFL Lausanne. The NMR experiments were analyzed in collaboration with Basile Vuichoud, Aurelien Bornet and Sami Jannin.

References

[1] J. Allen, H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver, D. B. Kell, *Nat. Biotechnol.* **2003**, *21*, 692–696.

[2] M. F. Chauvin, F. Megnin-Chanet, G. Martin, J. M. Lhoste, G. Baverel, *J. Biol. Chem.* **1994**, *269*, 26025–26033.

[3] J. A. den Hollander, T. R. Brown, K. Ugurbil, R. G. Shulman, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 6096–6100.

[4] C. Ducrocq, M. Lenfant, G. H. Werner, B. Gillet, J. C. Beloeil, *Biochem. Biophys. Res. Commun.* **1987**, *147*, 519–525.

[5] A. W. Jans, D. Leibfritz, *NMR Biomed.* **1989**, *1*, 171–176.

[6] R. B. Moon, J. H. Richards, *J. Biol. Chem.* **1973**, *25*, 7276–7278.

[7] G. Navon, S. Ogawa, R. G. Shulman, T. Yamane, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 888–891.

[8] F. Seguin, A. Le Pape, *J. Immunol. Methods* **1994**, *4*, 179–187.

[9] K. Ugurbil, T. R. Brown, J. A. den Hollander, P. Glynn, R. G. Shulman, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 3742–3746.

[10] B. M. Goodson, *J. Magn. Reson.* **2002**, *155*, 157–216.

[11] K. Golman, O. Axelsson, H. Johannesson, S. Mansson, C. Olofsson, J. S. Petersson, *Magn. Reson. Med.* **2001**, *46*, 1–5.

[12] A. Abragam, M. Goldman, *Rep. Prog. Phys.* **1978**, *41*, 395–467.

[13] J. H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning, K. Golman, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10158–10163.

[14] A. Bornet, J. Milani, S. Wang, D. Mammoli, R. Buratto, N. Salvi, T. F. Segaw, V. Vitzthum, P. Mieville, S. Chinthalapalli, A. J. Perez-Linde, D. Carnevale, S. Jannin, M. Caporinia, S. Ulzega, M. Rey, G. Bodenhausen, *Chimia* **2012**, *66*, 734–740.

[15] S. Meier, M. Karlsson, P. R. Jensen, M. H. Lerche, J. O. Duus, *Mol. BioSyst.* **2011**, *7*, 2834–2836.

[16] S. Meier, P. R. Jensen, J. O. Duus, *FEBS Lett.* **2011**, *585*, 3133–3138.

[17] S. Meier, P. R. Jensen, J. O. Duus, *ChemBioChem* **2012**, *13*, 308–310.

[18] T. Harris, H. Degani, L. Frydman, *NMR Biomed.* **2013**, *26*, 1831–1843.

[19] M. E. Merritt, C. Harrison, C. Storey, F. M. Jeffrey, A. D. Sherry, C. R. Malloy, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19773–19777.

[20] B. Vuichoud, J. Milani, A. Bornet, R. Melzi, S. Jannin, G. Bodenhausen, *J. Phys. Chem. B* **2014**, *118*, 1411–1415.

[21] D. A. Ovchinnikov, *Genesis* **2008**, *46*, 447–462.

[22] S. Galvan-Pena, L. A. O'Neill, *Front. Immunol.* **2014**, *5*, 420.

[23] B. Ghesquiere, B. W. Wong, A. Kuchnio, P. Carmeliet, *Nature* **2014**, *511*, 167–176.

[24] L. A. O'Neill, *Biochem. J.* **2011**, *438*, e5–6.

[25] K. Schroder, P. J. Hertzog, T. Ravasi, D. A. Hume, *J. Leukoc. Biol.* **2004**, *75*, 163–189.

[26] G. C. Hard, *Br. J. Exp. Pathol.* **1970**, *51*, 97–105.

[27] P. Newsholme, R. Curi, S. Gordon, E. A. Newsholme, *Biochem. J.* **1986**, *239*, 121–125.

[28] S. E. Calvano, W. Xiao, D. R. Richards, R. M. Felciano, H. V. Baker, R. J. Cho, R. O. Chen, B. H. Brownstein, J. P. Cobb, S. K. Tschoeke, C. Miller-Graziano, L. L. Moldawer, M. N. Mindrinos, R. W. Davis, R. G. Tompkins, S. F. Lowry, *Nature* **2005**, *437*, 1032–1037.

[29] Y. Pocker, J. E. Meany, B. J. Nist, C. Zadorojny, *J. Phys. Chem.* **1969**, *73*, 2879–2882.

[30] M. Carravetta, O. G. Johannessen, M. H. Levitt, *Phys. Rev. Lett.* **2004**, *92*, 153003.

Abbreviations and Symbols

Å	angstrom
α, β	lower, higher energy state for spin ½
ASA	accessible surface area
ATP	adenosine triphosphate
AUC	analytical ultracentrifugation
B	magnetic field
Bam	β-barrel assembly machinery
BSA	Bovine Serum Albumin
CD	circular dichroism
CP	cross polarization
C_p	heat capacity
CSA	chemical shift anisotropy
CSP	chemical shift perturbation
δ	chemical shift
d_1	interscan delay
D_{is}	dipolar-coupling between residues i and s
Da	Dalton, $g.mol^{-1}$
DMEM	Dulbecco's Modified Eagle's medium
DNP	dynamic nuclear polarization
DSS	2-dimethyl-2-silapentane-5-sulfonic acid
E. coli	*Escherichia coli*
EDTA	ethylene-diamine-tetraacetic acid
ESR	electron spin resonance
FCS	fetal calf serum
FID	free induction decay
FT	Fourier transformation
γ	gyromagnetic ratio
G	Gibbs free energy
Gdm/HCl	guanidine hydrochloride
ΔH	enthalpy of transition

ħ	Planck constant, $1.054571726 \cdot 10^{-34}\ J \cdot s$
HMQC	heteronuclear multiple quantum coherence
Hsp	heat shock protein
HSQC	heteronuclear single quantum coherence
I, S	nuclear spin
IM	inner membrane
INEPT	insensitive nuclei enhancement by polarization transfer
IFN-γ	interferon gamma
IPTG	isopropyl β-D-1-thiogalactopyranoside
${}^{n}J_{IS}$	scalar coupling constant between I and S separated by n covalent bonds
$J(\omega)$	spectral density
k	chemical exchange rate
k_B	Boltzmann constant, $1.381 \cdot 10^{-23}\ J \cdot K^{-1}$
K_D	dissociation constant
k_{ex}	kinetic exchange rate constant
LLS	long-lived states
LPS	lipopolysaccharides
μ	magnetic moment
m	spin quantum number
M	molar (mol / L)
M_x, M_y, M_z	x, y, z component of the magnetization M
MRI	magnetic resonance imaging
MWCO	molecular weight cut-off
NADH	reduced nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
n_α, n_β	population of the α and β state
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
NUS	non-uniform sampling
OD	optical density
OM	outer membrane
OMP	outer membrane protein
PBS	phosphate buffered saline

PDB	RCSB Protein Data Bank, www.rcsb.org/pdb
PHIP	para-hydrogen induced polarization
POTRA	polypeptide transport associated
PPIase	peptidyl-prolyl isomerase
ppm	parts per million
PRE	paramagnetic relaxation enhancement
θ	ellipticity
R	gas constant, $8.314\ J \cdot K^{-1} \cdot mol^{-1}$
R_1, R_2	longitudinal and transversal relaxation time
R_I, R_S	self-relaxation constant of spin I, spin S
RT	room temperature
σ	shield
σ_{IS}	cross relaxation rate constant between spins I and S
SAXS	small-angle X-ray scattering
S. cerevisiae	*Saccharomyces cerevisiae*
Sec	secretory-translocase
SF	scaling factor
S. flexneri	*Shigella flexneri*
Skp	seventeen-kilodalton protein
SPR	surface plasmon resonance
SurA	survival factor A
t_1, t_2	chemical shift evolution times
T_1, T_2	longitudinal and transversal relaxation
τ_c	molecular rotational correlation time
τ_m	mixing time
T_m	melting temperature
TamA	translocation and assembly module A
TCA	tricarboxylic acid
TEMPOL	4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl
TROSY	transverse relaxation-optimized spectroscopy
ω	resonance frequency
W	transition rate constant

Acknowledgments

This work was carried out in the laboratory of Prof. Dr. Sebastian Hiller in the focal area of Structural Biology and Biophysics at the Biozentrum, University of Basel, in the time from September 2011 until September 2015.

I would like to thank Prof. Dr. Sebastian Hiller for offering me the possibility to do my thesis in his laboratory. Thanks to his support, guidance and availability I could perform this work in the best conditions and could learn a lot scientifically and personally. His motivation and investment in all my different projects were a key factor for its success as well as the infrastructure he provided in concert with the Biozentrum. I also wish to thank all present and past group members for their support and friendship all along my PhD, particularly Dr. Björn Burmann for his help and guidance in all my different projects.

Thanks to Dr. David Kentner from the laboratory of Prof. Dr. Bumann, for conducting the research on *S. flexneri* and preparing the NMR samples as well as to Dr. Mauricio Rosas-Ballina from the group of Prof. Dr. Bumann for the work on macrophages and Basile Vuichoud, Aurélien Bornet and Sami Jannin from the group of Prof. Dr. Goeffrey Bodenhausen, for the d-DNP NMR measurements.

I would like to thank Prof. Dr. Henning Stahlberg for offering me the possibility to work in his laboratory under the supervision of Dr. Fabian Kebbel and, Prof. Dr. Martin Spiess for being members of my PhD committee meeting and for kindly accepting to be members of my final PhD committee. I am also very thankful to the Fellowships for Excellence Foundation (formerly called Werner Siemens Fellowship) for funding my salary and other expenses during the first three years of my PhD and for great study trips.

Thanks to all the members of the 3rd floor for bringing a great atmosphere and for their help, teaching me biophysical techniques and to Dr. Tim Sharpe for his grateful help in the Biophysics facility. Finally I would like to thank my friends, for having been there all along my PhD, my parents supporting and pushing me during all my study time, my brother for his support and friendship and my boyfriend, Stéphane who gave me the strength to start and achieve my thesis, in the best conditions.