Magic-angle spinning NMR of bacteriophage viruses

Amir Goldbourt

School of Chemistry, Tel Aviv University, Ramat Aviv 6997801, Tel Aviv, Israel

amirgo@tauex.tau.ac.il

Abstract

Bacteriophages are viruses that infect bacteria. In their simplest form they are highly symmetric biomolecular assemblies that consist of an inner genomic core wrapped by a protein coat. More complex bacteriophages have capsid shells that include several proteins, some have short or long protein tails, as well as additional fibrous tail tube protein attachments. Magic-angle spinning solid-state NMR provides an opportunity to study these high-molecular-weight (tens of MegaDaltons) phage systems in great detail. This review focuses on several filamentous and icosahedral phage of various complexities, showing the hierarchy of information available by NMR – protein and DNA chemical shifts; secondary, tertiary and quaternary structures, hydration, protein-DNA interactions, and capsid dynamics.

Keywords: Magic-angle spinning; solid-state NMR; bacteriophages; filamentous phage

Introduction

Magic-angle spinning (MAS) NMR has emerged in the past two decades as a leading tool for the characterization of proteins, including their aggregates, membrane-bound, surface-bound, in-cell, embedded in biominerals, complexed with nucleic acids, and as large molecular assemblies.[1–5] A great detail of information can be extracted from such studies, including secondary, tertiary and quaternary structures, local dynamics on several time-scales, interactions between surfaces, hydration states, ligand binding phenomenon, and more. All these properties of biomolecules are key to understanding their structure and function, to rationalizing sources of disease and to exploring pathways for molecular-based treatment. The applicability of NMR to study complex biomolecules depends greatly on sample preparation, sample homogeneity, and stability. Another key element is achieving sufficient sensitivity and resolution. Towards the realization of both aspects the last decade has brought about high fields, ultra-fast spinning capabilities, non-uniform sampling and processing techniques, and dynamic nuclear polarization. All these advances have pushed forward the applicability of NMR to study proteins of increased primary sequence lengths, higher molecular-weights, and biological assemblies of rising complexity.

MAS NMR studies of viruses have been applied to study various types of viruses,[6] including HIV virus[7–9] and the measles virus.[10] A significant body of structural information on phage viruses exists from X-ray crystallography, fiber diffraction techniques and electron microscopy.[11,12] Some structures have been solved at atomic resolution – e.g. the crystallographic structure of the RNA calcium binding capsid of phage PRR1[13] belonging to family Leviviridae, the cryo electron microscopy of the filamentous bacteriophage IKe, [14] and that of the Zika virus.[15] Some structures have been elucidated at lower resolution but provide a clear view on the

phage topology and architecture, and allow modeling of the different protein components (e.g. the 7.6Å cryoEM model of *Bacillus subtilis* tailed bacteriophage SPP1[16]). Naturally such techniques cannot detect mobile portions of phage, normally they cannot provide sufficient information on the genome – RNA or DNA, and on its interactions with the capsid. In other cases phage samples are non-crystalline, or cannot be properly aligned.

The current review will focus on the application of MAS solid-state NMR to study bacteriophages, viruses that infect bacteria, which consist of a genome wrapped by a protein shell. The genome could be RNA or DNA, single or double stranded. Mostly those nucleocapsids are non-enveloped, however, their capsid may contain a single coat protein in the simplest form or a complex set of proteins that make up a capsid (mostly icosahedral), a tail, and possibly additional tail fibers that recognize and bind the target cell.

Filamentous bacteriophages

Filamentous phages of the family *inoviridae* have a circular single-stranded DNA encased in a highly symmetric capsid made of mostly a single protein (the major coat protein).[17] There are few additional minor coat proteins existing in small copy numbers at both edges of these micron-long structures that ensure the viability of the infection and assembly processes. One of these minor coat proteins, gene-3 protein, attaches to a specific pili organelle of the host cell facilitating the infection. The genome encodes a total of 11 genes, but not all are structural. For example, gene-5 protein (gVp) is expressed in the bacteria and its role is to cap the DNA during the rolling circle mechanism thus preventing further DNA replication. This gVp-ssDNA complex then reaches the cell membrane where it is being replaced by gene-8 protein – the major coat protein. A schematic of the infection process, and the possible role of the various coat proteins, are shown in Fig. 1.

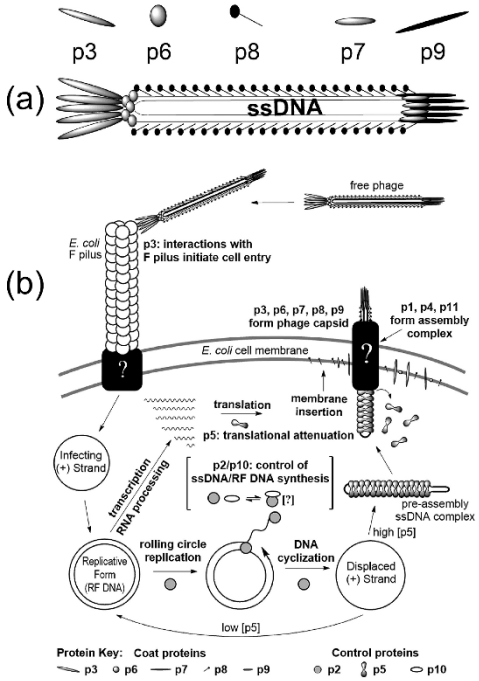

Fig. 1. Filamentous phage architecture (a) and life cycle (b). Reproduced from Smeal et al.[18]

Early detailed structural information on filamentous phage existed from fiber diffraction, aligned NMR, and other biophysical techniques such as electronic spectroscopy and mutational studies.[17,19,20] In the last decades various filamentous phage have been studied by magic-angle spinning (MAS) NMR providing new insights and new structures. Phage sample preparation and NMR methodologies, described below, are in general common, but are tailored and matched to the particular system under study, to the information that we aim to extract, and to the available hardware.

Filamentous phage sample preparation

Filamentous phages do not lyse the host cells and thus when the efficiency of infection is high, they can be prepared in many cases in high yields. Each phage has a specific host bearing a suitable pili organelle and therefore the host strains have to be initially obtained and cultured. A detailed description with complete protocols was published elsewhere;[21] here a brief description of the main steps in sample preparation and isotope labeling will be given.

A single colony of a suitable host is grown in a small tube and infected with the proper phage. The small-scale culture is then transferred to a flask containing a minimal medium labeled with the isotopes of choice. For a fully ^{13}C-^{15}N labeled phage, $^{13}C_6$-glucose and $^{15}NH_4Cl$ are added to the M9 medium. Other labeling schemes can be obtained by supplementing the media with partially labeled glycerol[22,23], glucose[24,25], or any other precursor of choice[26]. Amino-acid specific labeling or blanking can also be obtained in a similar way[27]. For phage infecting *Escherichia coli*, the metabolic pathways leading to amino acid labeling are similar to those obtained with common protein expression protocols. For Pf1, it should be taken into account that the host is *Pseudomonas aeruginosa*, a bacterium that metabolizes glucose via the Entner Doudoroff (ED) pathway. Thus, the use of different precursors can generate different labeling schemes. For example, the use of 1-^{13}C glucose results in a carbonyl-labeling scheme,[28] rather than the expected Cα and sidechain labeling.[24] A similar carbonyl-rich labeling pattern was obtained by feeding *E.coli* with pyruvate,[29] and since feeding *E.coli* with gluconate induces the ED pathway[30] instead of the Embden Meyerhof Parnas glycolysis pathway, carbonyl-labeling can also be obtained by feeding *E. coli* with 1-^{13}C-gluconate.[31] The manifestation of the ED pathway is envisioned in Figure 2 showing the labeling of ubiquitin over-expressed in E. *coli* in the presence of gluconate, and the labeling of the Pf1 capsid in the presence of glucose, where both precursors are labeled at C1.

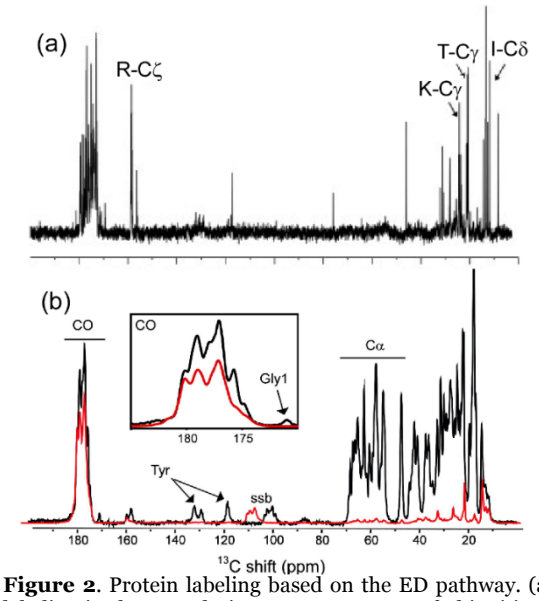

Figure 2. Protein labeling based on the ED pathway. (a) 1-^{13}C gluconate-induced ED pathway leads to mostly-carbonyl labeling in the ^{13}C solution NMR spectrum of ubiquitin. Reprinted from Refaeli and Goldbourt[31], Copyright (2012), with permission from Elsevier (b) Pf1 grows on P. Aeruginosa with a minimal media containing 1-^{13}C glucose, and the ED pathway naturally leads to the same labeling pattern. Reprinted from Goldbourt et al., Copyright (2007), with permission from Elsevier[28]

Once phage growth has maximized, the final minimal media contains in the solution both phage and bacteria, and thus by repeating centrifugation and phage PEG precipitation, the two entities can be separated. Final purification of phage samples is obtained by CsCl ultracentrifugation, and phage purity and concentration can be assessed by examining their ultra-violet spectrum.

Spectroscopic NMR Methods

There is a variety of techniques suitable for the characterization of phage, most of which are similar to those used for crystalline proteins, membrane proteins, amyloids, and other biological assemblies. Others are tailored to the particular system. Chemical shift assignment of the coat protein, which is a prerequisite for structure determination, relies on common two- and three-dimensional techniques.[32,33] For studies of filamentous phage thus far, we utilized ^{13}C and ^{15}N based experiments, and avoided ^{1}H detection. The latter is increasingly used in particular with the advent of probes spinning at 60-120 kHz.[34,35] While at 40-60 kHz deuteration is a prerequisite, spinning at over 100 kHz facilitates experiments at fully protonated proteins. Those will be discussed in the context of spherical phage systems.

There are several basic elements, which constitute the building blocks of all common pulse sequences. (i) The signal is generated by excitation of protons, and is delivered to either ^{13}C or ^{15}N via a cross-polarization (CP) block, in which the rare nucleus (^{13}C or ^{15}N) and the protons are irradiated at the Hartmann-Hahn MAS matching conditions[36], $\nu_{1H}\pm\nu_{1X}=n\nu_R$. Here ν_{1X} is the radio-frequency (rf) power level on nucleus 'x', ν_R is the spinning speed, and n is an integer, usually ±1 or ±2. The condition of summation of the two rf fields is only applicable at sufficiently fast spinning frequencies. Inverse CP is used in sequences that require transfer of polarization back to protons, either for the purpose of mixing (e.g. XHHX[37]) or for proton detection. (ii) Homonuclear transfer schemes via proton-driven spin diffusion (PDSD) while the rare nucleus is in the

longitudinal mode and protons are irradiated at certain resonance conditions (e.g. free spin diffusion, dipolar assisted rotational resonance (DARR),[38] and others[39]), via direct dipolar-based polarization transfer such as rotor frequency driven recoupling (RFDR),[40] dipolar recoupling enhanced by amplitude modulation (DREAM),[41] or symmetry-based transfer mechanisms such as combined $R2_n^\nu$–driven sequence (CORD),[42] and via direct scalar couplings (e.g. INADEQUATE[43], TOBSY[44]). (iii) Heteronuclear polarization transfer, usually between nitrogen and carbon. This transfer can be achieved selectively (that is, specific N-Cα or N-CO) by double CP[45,46] (DCP) or non-selectively by transferred echo double resonance (TEDOR)[47] and for even a long-range transfer by third-spin assisted recoupling (TSAR, or PAIN-CP).[48] (iv) Anisotropic recoupling techniques, which are mainly used to study dynamics by probing the reduction in the size of dipolar and CSA interaction with respect to some static limit values. There is a variety of recoupling techniques; for the dipolar interaction the main methods are Rotational echo double resonance (REDOR)[49] for rare spins, dipolar chemical shift correlation (DIPSHIFT) for H-X bonds,[50] and R-based symmetry sequences. Lee-Goldburg (LG) CP build-up curves[51] can be analyzed to provide H-X dipolar couplings as well. For CSA recoupling there are also various efficient symmetry-recoupling sequences,[52] and CSA can also be obtained in other ways.[53,54] (v) ^{1}H decoupling during the various periods of rare-spin free precession or polarization transfer.[55]

The four blocks (i-iv) described above are combined in different ways to generate a large variety of sequences for the determination of chemical shifts, structure and dynamics. For example, the block CP-t1-DCP-t2-PDSD-t3 will generate a 3D N-C-C experiment if the initial CP is to ^{15}N, and the selection of the rf carrier position during DCP will determine if the experiment is sequential (NCOCX) or not (NCACX); a block of the form CP-t1-DCP-t2-DCP-t3 will generate a sequential C-N-C experiment (CONCA or CANCO); a block of the form CP-DIPrec(t1)-t2-DCP-t3 will measure the H-N dipolar order parameters for each peak in a N-C isotropic space, reporting on backbone dynamics following the model-free approach. Modification of the dipolar-based mixing times will affect the distance through which correlations can be obtained and hence can be used to provide structure distance restraints. Increasing the dimensionality (or the number of transfers) is also straightforward, however, it is limited by the sensitivity and by the total experimental time. Both limitations are now becoming feasible with the advent of dynamic nuclear polarization (DNP) and non-uniform sampling (NUS) techniques. A recent example demonstrates nicely how NUS can be utilized to study filamentous phage.[56]

Pf1 phage

Pf1 filamentous phage infects *P. aeruginosa* strain K (PAK) via type IV pili. It has a circular single stranded (ss) DNA with 7349 bases[57] wrapped by approximately the same number of copies of its 46-residue-long major coat protein. Its protein sequence is as follows:

GVIDTSAVESAITDGQGDMKAIGGYIVGALVILAVAGLIYSMLRKA

The length of Pf1 phage is approximately 2 μm and its diameter is between 6-7 nm. It belongs to class-II symmetry[58] as its capsid is arranged as a one-start helix, with the exact symmetry depending on the temperature, above or below 10 °C. From a refined model based on fiber diffraction[59] and align NMR data[60], the rotation per subunit is 65.915 degrees, the rise 3.05Å,[61] and the coat protein adopts a mostly helical structure that is slightly bent.

Close to complete chemical shift NMR assignments of the coat protein for Pf1 at its high temperature form have been obtained by analyzing 2D C-C DARR and two 3D sequential and intra-residue N-C-C experiments at a field of 17.6 T (750 MHz) using only fully-labeled samples.[62] These shifts are consistent with a mostly all-

helical coat protein having a non-structured N-terminus. For the low-temperature form, chemical shift deviations have been associated with intersubunit interactions, reporting on structural rearrangement of the hydrophobic core.[63] Due to its highly symmetric arrangement, Pf1 has also been utilized to develop and examine new assignment techniques based on DNP. Despite the relatively broad lines, an approach termed sequential side-chain–side-chain (S^3) had been utilized[64], in which the sequence blocks were combined to obtain either residue *i* to *i+1* sequential transfer, or *i* to *i-1* sequential transfer: The block CP-t1-DARR-DCP-t2-DCP-DARR-t3 generates CX_i-(CO)N_{i+1}-(CA)CX_{i+1} correlations or CX_i-(CA)N_i-(CO)CX_{i-1} correlations, depending on the carrier position during DCP. As shown in Figure 3a, Despite the large amount of transfers, DNP allows sufficient remaining sensitivity, de-novo assignment of the phage, and the ability to compare complete assignments at 100 °K and 273 °K. DNP was also utilized, in conjunction with conventional experiments, to provide chemical shifts for the ssDNA[65], which is unique in Pf1[66,67] most probably resulting from the fact that unlike other phages (see below) the ratio of nucleotides to subunits is one. An interesting outcome of the unique DNA arrangement is shown in Figure 3b; the single tyrosine-40 residue in the coat protein of Pf1 splits to many different shifts most probably due to close interactions with the four DNA bases (and two opposite strands – up and down) that generate different chemical environments.

The knowledge on chemical shifts allowed also to study additional properties of the phage. Dipolar order parameters, obtained by analysis of LGCP ^{1}H-^{15}N build-up curves,[68] reported on a highly rigid backbone with the exception of the single N-terminal glycine. As expected, sidechains facing the exterior of the phage undergo large amplitude motions but interestingly, sidechains facing the interior ssDNA also experience such motion suggesting a mobile protein-DNA interface. Indeed detailed studies of hydration water show that this interface is lubricated and that water mediate protein-DNA interactions in Pf1.[69,70]

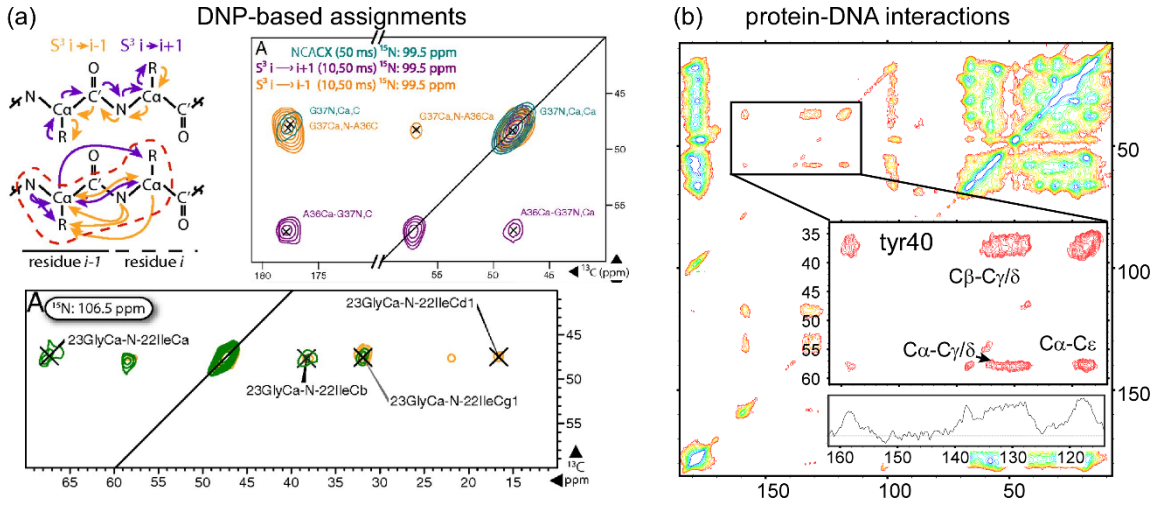

Figure 3. Pf1 phage. (a) DNP-based sidechain-sidechain correlation spectroscopy showing the pathway, forward and backward sequential transfer (G37-A36) at the top spectra and long-range correlations to the sidechains (G23-I22) in the bottom. Reproduced from Sergeyev et al.[64] (b) A broad multiplet appearance of the tyrosine-40 sidechain suggesting stacking with DNA bases in Pf1.

M13 and fd phage

M13 and fd filamentous phage (they are termed, together with f1 phage, Ff family) infect strains of *E. Coli* bearing F pili. Their circular ssDNA have 6407 and 6408 bases, respectively[71] with very minor changes, less than 3%. The genome is wrapped by approximately 2750 copies of a 50-residue-long major coat protein. The

nucleotide to subunit ratio is non-integer, ~2.3-2.4.[72] Their coat protein sequences differ in just a single amino acid (in bold) and thus in a single charge:

M13: AEGDDPAKAAF**N**SLQASATEYIGYAWAMVVVIVGATIGIKLFKKFTSKAS

fd: AEGDDPAKAAF**D**SLQASATEYIGYAWAMVVVIVGATIGIKLFKKFTSKAS

The length of M13 and fd phage is approximately 1 μm and the diameter is between 6-7 nm. They belong to class-I symmetry, for which the capsid is arranged as a five-start helix. We obtained the structure of M13 from MAS ssNMR experiments, combined with CS-Rosetta calculations.[73] The rotation per pentamer is 36.4 degrees, the rise 16.6Å, and the coat protein adopts a mostly helical bent structure (see Figure 4). NMR chemical shifts suggest that fd adopts an almost identical structure with very minor changes on the exterior.[74] The different charge reduced the effective linear charge density from -0.4 $Å^{-1}$ in M13 to -0.6 $Å^{-1}$ in fd.[75]

Chemical shifts of fd and M13 have been obtained by combining dipolar and scalar-based experiments on fully and sparsely (glycerol) labeled samples.[74,76] The scalar-based refocused INADEQUATE experiment[77] was highly useful for identifying mobile fragments and distinguishing unequivocally single-bond contacts.

M13 capsid structure determination (pdb id 2MJZ) relied mostly on the acquisition of dipolar-based (CORDxy4 experiments[42]) carbon correlations from sparsely labeled phage samples (prepared from 2-glycerol and 1,3-glycerol), and distinguishing inter-subunit contacts from intra-subunit contacts. The latter could be identified due to the helical character of the coat protein, which restricts visible contacts to residues separated in the sequence by up to four. In the first stage, a sufficient amount of non-ambiguous contacts could be identified to feed the fold-and-dock Rosetta protocol.[78] The model was then formed by assuming an initial extended coat protein arranged in a pentamer architecture (following fiber diffraction data). Seven pentamers were positioned with varying diameters, relative angles (a single tilt angle between consecutive pentamers), and relative inter-pentamer distance (again, a similar distance between all pentamers). An initial model was obtained and then used to obtain additional non-ambiguous contacts. Up to that point all distance restraints assumed an inter-nuclear distance of up to 7Å. Those were fed alongside ambiguous distance restraints to generate a higher-resolution structure. The final step was to distinguish shorter-range distances up to 5Å from longer-range distances by using the tryptophan sidechain as a scale bar, that is, spectra acquired at short mixing times that lack tryptophan contacts between the backbone and the six-membered ring were limited to a short-distance restraint. This final step allowed the determination of the final structure.

The highly resolved, highly sensitive spectra of fd and M13 phage allowed us to further explore the interactions of the fd capsid with the ssDNA. Since aromatic residues and DNA bases may overlap at the regions around 110-160 ppm, a phage sample was produced that lacks ^{13}C and ^{15}N labels on tryptophan, tyrosine, and phenylalanine (YFW^{-}). Protein-DNA contacts could then be revealed by two means: (i) assigning the DNA resonances and detecting ^{13}C-^{13}C ribose-protein and base-protein correlations; (ii) performing a PHHC correlation experiment (H→^{31}P CP block-t1-^{31}P→H CP block-HH mixing-H→^{13}C CP block-t2). Analysis of the two spectra defined the interface between the protein and the ssDNA (see Figure 4e), which spans the interior residues in the C-terminus up to residue Ile32. Beyond that residue no contacts have been observed, verifying also the structural model that defined the rise between pentamers since the C-terminus of a subunit in the upper pentamer fits exactly at that position preventing further contacts with the DNA.[79]

Evidence for the increased mobility of the protein-DNA interface has been obtained by comparing spectra of silver-titrated phage and intact phage.[80] Since silver intercalates between the DNA strands of fd[81] it strengthens the contacts between the strands and presumably loosens the DNA-protein contacts. One result is that the macroscopic properties of the particle as a whole change thus forming liquid crystals with

cholesteric pitches that depend on the silver constrations.[82] The other result was that in the spectral comparison no real shift changes were observed, but residues in the interior of the phage had decreased intensities suggesting increased mobility.

The y21m mutant of fd phage

The y21m mutant of fd has been developed initially to allow proper alignment of phage particles for the purpose of structure determination (Figure 4a,b). The overall structural architecture is similar to fd and M13, however, some key differences can be pointed out. It forms liquid crystals with a significantly different pitch,[83] it has a larger persistence length (9.9 μm vs 2.8 μm in fd *wt*), and as shown in Figure 4 below, it adopts a different structure from the wild-type – in particular with different symmetry parameters.[75] From a refined model based on fiber diffraction (2C0W[84]) and align NMR data(1NH4),[85] the rotation per subunit in this model (2C0X) is 36.0 degrees, and the rise 16.15Å. [85]

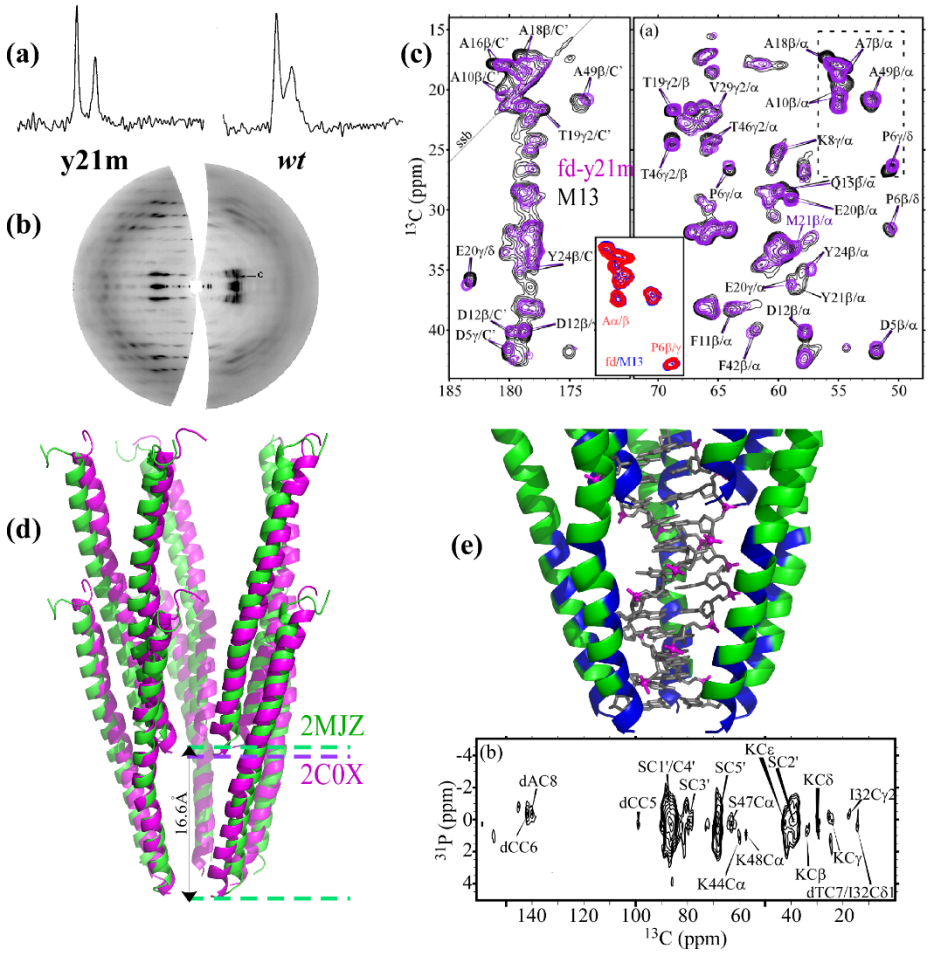

Figure 4. Ff phage: (a,b) A comparison of *wt* and y21m mutant fd phage showing the improved alignment properties of the y21m mutant. (a) ^{15}N 1D aligned NMR spectrum (^{15}N-Leu labeled). Adapted from Tan et al.[86] Copyright (1999), with permission from Elsevier. (b) Fiber diffraction patterns. Reprinted with permission from Welsh et al.[87] Copyright (1996) American Chemical Society. (c) MAS NMR 2D ^{13}C-^{13}C spectral comparison of M13 with fd-y21m showing that structural variations (shown in (d)) are not resulting from the different techniques for structure determination. On the other hand, the alanine Cα-Cβ correlations of M13 and *wt*-fd are shown (in the inset) to be indistinguishable. (d) Structure alignment of two pentamers (out of seven in a minimal representation of the phage capsid) from the structures of M13 (pdb 2MJZ) and fd-y21m (pdb 2C0X). The different rise of the two structural models is indicated by the dashed lines; the rise for the mutant is smaller in 0.45Å per pentamer. (c) and (d) Adapted with permission[75]. Copyright (2017) American Chemical Society. (e) A plot of the *wt*-fd protein-DNA interface (blue) as extracted from ^{13}C-^{13}C DARR and ^{31}P-^{13}C PHHC correlation spectra. A DNA

molecule was schematically inserted within the inner part of the phage and has the phosphate in magenta. Bottom: The PHHC spectrum of YFW- labeled fd phage. Adapted with permission. [79] Copyright (2014) American Chemical Society.

A quantitative measure of capsid mobility was obtained by CSA recoupling experiments. Specifically, adopting the RNCSA recoupling sequence,[52] which is an RN_n^{ν}-symmetry based CSA recoupling scheme,[88] a set of three recoupling experiments provided insight into backbone rigidity and motion in the capsid of M13. Backbone ^{15}N CSA values were obtained by acquiring two 3D $N_{CSA}(R14_2^5)$-N_{iso}-C_{iso} experiments differing in the detected carbon (Cα and C'),[89] and ^{13}C CSA values were obtained with a 3D $C_{CSA}(R10_1^3)$-C'_{iso}-Cx_{iso} experiment (C3CSA).[90] The CSA values are in agreement with a highly rigid helical backbone; ^{13}C CSA values differ by less than 1% then those measured for the helical part of the highly rigid GB1 protein (83.5 ppm),[53] and ^{15}N values (103.1 ppm) are lower by 3% that those of the helical part of GB1 (106.2 ppm). It was also observed that the N-terminus undergoes large amplitude motions with cone angles estimated to be ~20-50 degrees. Interestingly unlike Pf1, here motions were observed up to the 5th or 6th residue, the entire non-helical N-terminal part of the capsid. Although not all sidechains have been analyzed since only carboxyl groups could be detected in C3CSA and ^{15}N appears only in very few sidechains, there were enough to observe dynamics in the exterior of the phage, and to deduce that the tryptophan is highly rigid, basically immobile, thus its use as a scale-bar for structure calculations is justified.

Icosahedral bacteriophages

Solid state NMR studies have been utilized to study several more complex phage families. Single-stranded RNA phages belonging to the family *Leviviridae* have a simple icosahedral head made of 178 copies of their coat protein and another copy of a maturation ('A') protein. Examples are the *E.coli* phages MS2 and GA (genus *Levivirus*), Qβ and SP (genus *Allolevivirus*), and the *Acinetobacter* phage AP205.[91] Similarly to filamentous phage, they infect bacteria bearing pili however they attach to the side of the pili. More complex bacteriophages have tails and tail fibers. The most common dsDNA phage families are *Myoviridae* that have contractile tails (e.g. T4), the long non-contractile tailed *Siphoviridae* (e.g. *Bacillus* phage SPP1), and the short-tailed *Podoviridae* (e.g. T7). Those phage infect bacteria by attachment to the lipopolysaccharides.[92]

AP205 virus-like-particles

Acinetobacter AP205 is a ssRNA (4268 nucleotides) icosahedral phage with a head diameter of approximately 29 nm. The coat protein sequence is 131-residue long, and the basic structural unit is a dimer with an α/β double sandwich topology. The empty capsid (virus-like-particle, VLP) of AP205, generated from self-assembly of over-expressed coat protein subunits, was studied by MAS NMR techniques based on the combination of ultra-fast MAS, high field, and ^{1}H-detection, demonstrating how the significant advance in solid-state NMR technology facilitates a fast and reliable characterization of large molecular assemblies.[93] Furthermore, ^{1}H-^{1}H distance restraints were sufficient to obtain a well converged model of the basic dimer unit within the entire VLP.[94] A key observation by NMR was the existence of an intersubunit disulfide bond stabilizing the dimer and a shift of an N-terminal β-hairpin common to other ssRNA phages, to a combination of an N-terminal beta strand with a C-terminal beta strand. This circular shift also renders the AP205 dimer more tolerant to protein fusion for biotechnological purposes. An overlay of the VLP capsid from NMR, and that of the unassembled dimer obtained by X-ray crystallography, show an excellent agreement of the secondary structure elements (backbone RMSD of 2.35Å, see Figure 5a) with the exception of the 'FG' loop,

suggesting that upon assembly this loop readjusts itself to form the higher symmetry structure. Indeed when fitting the X-ray structure into the cryoEM map, this loop has to be readjusted.[95] Moreover, a single set of chemical shifts was obtained for this loop, suggesting it adopts a single conformation in the assembled VLP. Figure 5 shows the structural hierarchy obtained by the synergy of NMR, X-ray, and the 6.0Å cryoEM map.

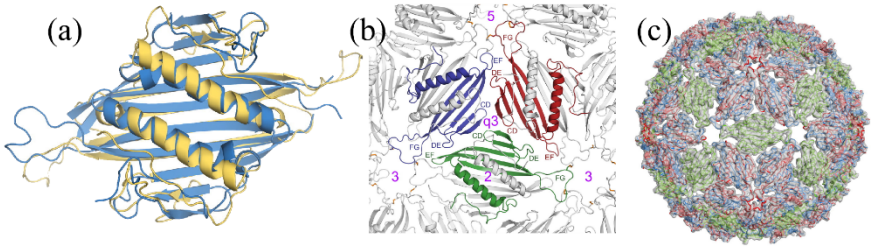

Figure 5. AP205 phage. (a) Overlay of the basic dimer unit from the NMR structure of the capsid (yellow) and from the X-ray structure of the dimer crystal (blue). (b) The organization of the dimers to form the icosahedral structure shown in (c). Adapted from figures 3 and 4 of Shishovs et al.,[95] Copyright (2016), with permission from Elsevier.

SPP1 phage tail assembly

The *Bacillus subtilis* SPP1 bacteriophage has a linear dsDNA genome with a length of 44007 base pairs, an icosahedral head with a diameter of 45 nm, and a 140 nm long tail. The Tail tube protein (TTP) gp17.1 contains 177 residues and it forms a hollow tube made of hexameric rings rotated one with respect to the other by 21° and have a rise of 40Å. In the wild-type there is also an occurrence of a longer TTP, gp17.1* (one in every three gp17.1 units), which is generated by a frameshift, but viable phage also exist without it.[96]

Solid-state NMR spectra of assembled gp17.1 tail tubes, fully deuterated and back-exchanged, have demonstrated the ability to assign large proteins with a set of three 4D ^{1}H-detected experiments (CACONH, COCANH, CBCANH) collected with non-uniformly sampled protocols, accompanied by two 3D experiments (CONH, CANH) collected with uniform sampling and traditional Fourier transform. These 3D experiments allowed navigating the 4D spectra and identifying real peaks from artifacts generated by the reconstruction algorithms.[97] In this manner 91% of backbone residues have been assigned, allowing to generate a clear view on the secondary structure of SPP1.

T7 bacteriophage

The *E. coli* T7 bacteriophage has a linear dsDNA with 39935 base pairs embedded in an icosahedral head. It contains 415 copies (60 hexamers and 11 pentamers) of two capsid proteins; gp10A with 345 amino acids and gp10B with 398 amino acids. As in SPP1, also here gp10B is generated by a translation frame-shift and exists in low amounts, less the 10%. According to cryoEM studies of a mutant containing a gp10A-only head, the average diameter of the mature head is 56.4 nm and gp10A adopts a complex fold with different domains and secondary structural elements.[98]

Solid-state NMR spectra of T7 bacteriophage focused on its DNA. The high molecular weight of its dsDNA allowed an opportunity to assess the feasibility of NMR to provide structural information on such tightly-packed B-form DNA molecules, to compare shifts between database DNA values and the coiled DNA of T7, and to compare dsDNA and ssDNA shifts (obtained for filamentous phage). Although site-specific nucleotide assignment was not sought after, and is probably not realistic for such systems, nucleotide-specific chemical shifts were obtained by recording 2D CC and NC spectra, the latter exemplified in Figure 6.[99] Almost all ribose

shifts were reported in a nucleotide-specific manner, and the highly resolved and narrow lines suggest a highly ordered DNA molecule. Overall, a comparison of the shifts in T7 DNA to reported values, suggest that the shifts were in agreement with a C2'-endo sugar pucker conformation, an *anti* orientation of the glyosidic bond, and suggest the existence of hydrogen-bonds typical of Watson-Crick base pairing.

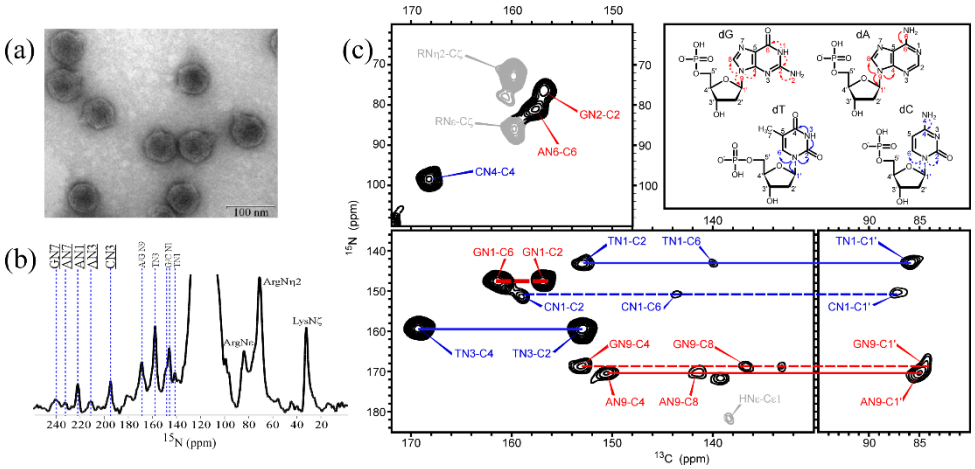

Figure 6. T7 phage. (a) EM image of T7 bacteriophage particles prior to precipitation for MAS NMR studies. (b) ^{15}N 1D CPMAS spectrum with main ^{15}N signals (taken from the thesis of G. Abramov[100]). (c) Assignment of nucleotides via heteronuclear ^{15}N-^{13}C TEDOR and the corresponding pathways in the different deoxy-nucleotides. (a,c) Reprinted by permission from Springer Nature, Journal of biomolecular NMR,[99] Copyright 2014.

Conclusions

Magic-angle spinning NMR of bacteriophages has expanded in the last decade from the initial studies of Pf1 to a variety of phage families representing different morphologies and complexities, and to other viruses with different shapes and forms. Not only have we learned a great deal on their structure and dynamics, but they also provided an opportunity for the development of new NMR techniques, and for testing new hardware and new methodologies due to the high sensitivity afforded by mainly filamentous phage viruses. Most probably, these systems will be highly useful for the upcoming revolutions of dynamic nuclear polarization and non-uniform sampling techniques, and to the approach of integrative structural biology, as was briefly shown in this review. The structure and dynamics of the genome of phage and other viruses is also of interest, in part due to their unique organization, and in part since they provide a platform for advance in the applications of solid-state NMR to study complex DNA and RNA molecules.

Acknowledgements

The study of bacteriophages fd, M13, fd-y21, and T7 has been funded over the years by the Israel Science Foundation.

Biographical Sketch

Amir Goldbourt, b. 1970, M.Sc. 1996, Tel Aviv University; PhD, 2003, under the supervision of Prof. S. Vega, Weizmann Institute of Science. In 2004 joined as postdoc in the field of biomolecular solid-state NMR at the group of Prof. Ann McDermott at Columbia University. At 2007 joined the school of Chemistry in Tel Aviv University at the Faculty of Exact Sciences as a Senior Lecturer. Since 2015 he is an associated professor.

In 2011 received the Regitze R. Vold Memorial Prize at the 7th alpine conference on Solid-state NMR. His research focuses on biomolecular MAS NMR of viruses, on structural and molecular aspects of mental illness, and on the development of quadrupolar NMR techniques.

References

1 P. C. A. van der Wel, *Emerg. Top. Life Sci.*, 2018, **2**, 57–67.

2 A. Goldbourt, *Curr. Opin. Biotechnol.*, 2013, **24**, 705–715.

3 V. S. Mandala, J. K. Williams and M. Hong, *Annu. Rev. Biophys.*, 2018, **47**, 201–222.

4 B. H. Meier, R. Riek and A. Böckmann, *Trends Biochem. Sci.*, 2017, **42**, 777–787.

5 G. Goobes, *Isr. J. Chem.*, 2014, **54**, 113–124.

6 C. M. Quinn, M. Lu, C. L. Suiter, G. Hou, H. Zhang and T. Polenova, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2015, **86–87**, 21–40.

7 Y. Han, J. Ahn, J. Concel, I.-J. L. Byeon, A. M. Gronenborn, J. Yang and T. Polenova, *J. Am. Chem. Soc.*, 2010, **132**, 1976–87.

8 Y. Han, G. Hou, C. L. Suiter, J. Ahn, I.-J. L. Byeon, A. S. Lipton, S. Burton, I. Hung, P. L. Gor'kov, Z. Gan, W. Brey, D. Rice, A. M. Gronenborn and T. Polenova, *J. Am. Chem. Soc.*, 2013, **135**, 17793–803.

9 M. J. Bayro, B. Chen, W.-M. Yau and R. Tycko, *J. Mol. Biol.*, 2014, **426**, 1109–1127.

10 E. Barbet-Massin, M. Felletti, R. Schneider, S. Jehle, G. Communie, N. Martinez, M. R. Jensen, R. W. H. Ruigrok, L. Emsley, A. Lesage, M. Blackledge and G. Pintacuda, *Biophys. J.*, 2014, **107**, 941–6.

11 D. A. Marvin, *Curr. Opin. Struct. Biol.*, 1998, **8**, 150–158.

12 M. G. Rossmann, *Crystallogr. Rev.*, 2015, **21**, 57–102.

13 M. Persson, K. Tars and L. Liljas, *J. Mol. Biol.*, 2008, **383**, 914–922.

14 J. Xu, N. Dayan, A. Goldbourt and Y. Xiang, *Proc. Natl. Acad. Sci.*, 2019, **116**, 5493–5498.

15 D. Sirohi, Z. Chen, L. Sun, T. Klose, T. C. Pierson, M. G. Rossmann and R. J. Kuhn, *Science (80-.).*, 2016, **352**, 467–470.

16 Y. Chaban, R. Lurz, S. Brasilès, C. Cornilleau, M. Karreman, S. Zinn-Justin, P. Tavares and E. V. Orlova, *Proc. Natl. Acad. Sci.*, 2015, **112**, 7009–7014.

17 L. A. Day, in *Encyclopedia of Virology*, eds. G. Allan and R. Webster, Elsevier, 3rd edn., 2008, vol. 3, pp. 117–124.

18 S. W. Smeal, M. A. Schmitt, R. R. Pereira, A. Prasad and J. D. Fisk, *Virology*, 2017, **500**, 275–284.

19 S. J. Opella, A. C. Zeri and S. H. Park, *Annu. Rev. Phys. Chem.*, 2008, **59**, 635–57.

20 D. A. Marvin, M. F. Symmons and S. K. Straus, *Prog. Biophys. Mol. Biol.*, 2014, **114**, 80–122.

21 O. Morag, N. G. Sgourakis, G. Abramov and A. Goldbourt, in *Methods in Molecular Biology*, 2018, vol. 1688, pp. 67–97.

22 D. M. LeMaster and D. M. Kushlan, *J. Am. Chem. Soc.*, 1996, **118**, 9255–9264.

23 F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein and H. Oschkinat, *Nature*, 2002, **420**, 98–102.

24 P. Lundström, K. Teilum, T. Carstensen, I. Bezsonova, S. Wiesner, D. F. Hansen, T. L. Religa, M. Akke and L. E. Kay, *J. Biomol. NMR*, 2007, **38**, 199–212.

25 A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker and A. Lange, *Nature*, 2012, **486**, 276–279.

26 C. Hoogstraten and J. J. Jr, *Concepts Magn. Reson. Part A*, 2008, **32A**, 34–55.

27 D. Lacabanne, B. H. Meier and A. Böckmann, *J. Biomol. NMR*, 2018, **71**, 141–150.

28 A. Goldbourt, L. A. Day and A. E. McDermott, *J. Magn. Reson.*, 2007, **189**, 157–65.

29 P. Lundström, D. F. Hansen and L. E. Kay, *J. Biomol. NMR*, 2008, **42**, 35–47.

30 R. C. Eisenberg and W. J. Dobrogosz, *J. Bacteriol.*, 1967, **93**, 941–949.

31 B. Refaeli and A. Goldbourt, *Biochem. Biophys. Res. Commun.*, 2012, **427**, 154–158.

32 V. A. Higman, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2018, **106–107**, 37–65.

33 A. Goldbourt, in *Encyclopedia of Analytical Chemistry*, John Wiley & Sons, Ltd, Chichester, UK, 2009, pp. 1–27.

34 A. Marchetti, S. Jehle, M. Felletti, M. J. Knight, Y. Wang, Z.-Q. Xu, A. Y. Park, G. Otting, A. Lesage, L. Emsley, N. E. Dixon and G. Pintacuda, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 10756–9.

35 S. K. Vasa, P. Rovó and R. Linser, *Acc. Chem. Res.*, 2018, **51**, 1386–1395.

36 E. . Stejskal, J. Schaefer and J. . Waugh, *J. Magn. Reson.*, 1977, **28**, 105–112.

37 D. Massiot, B. Alonso, F. Fayon, F. Fredoueil and B. Bujoli, *Solid State Sci.*, 2001, **3**, 11–16.

38 K. Takegoshi, S. Nakamura and T. Terao, *Chem. Phys. Lett.*, 2001, **344**, 631–637.

39 V. S. Mithu, S. Bakthavatsalam and P. K. Madhu, *PLoS One*, 2013, **8**, e50504.

40 J. H. Ok, R. G. S. Spencer, A. E. Bennett and R. G. Griffin, *Chem. Phys. Lett.*, 1992, **197**, 389–395.

41 T. Westfeld, R. Verel, M. Ernst, A. Böckmann and B. H. Meier, *J. Biomol. NMR*, 2012, **53**, 103–112.

42 G. Hou, S. Yan, J. Trébosc, J.-P. Amoureux and T. Polenova, *J. Magn. Reson.*, 2013, **232**, 18–30.

43 A. Lesage, C. Auger, S. Caldarelli and L. Emsley, *J. Am. Chem. Soc.*, 1997, **119**, 7867–7868.

44 M. Baldus and B. H. Meier, *J. Magn. Reson. Ser. A*, 1996, **121**, 65–69.

45 J. Schaefer, R. . McKay and E. . Stejskal, *J. Magn. Reson.*, 1979, **34**, 443–447.

46 M. Baldus, A. T. Petkova, J. Herzfeld and R. G. Griffin, *Mol. Phys.*, 1998, **95**, 1197–1207.

47 A. W. Hing, S. Vega and J. Schaefer, *J. Magn. Reson.*, 1992, **96**, 205–209.

48 J. R. Lewandowski, G. De Paepe and R. G. Griffin, *J. Am. Chem. Soc.*, 2007, **129**, 728–729.

49 T. Gullion and J. Schaefer, *J. Magn. Reson.*, 1989, **81**, 196–200.

50 M. Hong, J. D. Gross and R. G. Griffin, *J. Phys. Chem. B*, 1997, **101**, 5869–5874.

51 M. Hong, X. Yao, K. Jakes and D. Huster, *J. Phys. Chem. B*, 2002, **106**, 7355–7364.

52 G. Hou, I.-J. L. Byeon, J. Ahn, A. M. Gronenborn and T. Polenova, *J. Chem. Phys.*, 2012, **137**, 134201.

53 B. J. Wylie, L. J. Sperling, H. L. Frericks, G. J. Shah, W. T. Franks and C. M. Rienstra, *J. Am. Chem. Soc.*, 2007, **129**, 5318–5319.

54 I. Hung, Y. Ge, X. Liu, M. Liu, C. Li and Z. Gan, *Solid State Nucl. Magn. Reson.*, 2015, **72**, 96–103.

55 A. Equbal, M. Bjerring, P. K. Madhu and N. C. Nielsen, *J. Chem. Phys.*, 2015, **142**, 184201.

56 G. Porat and A. Goldbourt, *Isr. J. Chem.*, 2019, doi:10.1002/ijch.201900058.

57 D. F. Hill, N. J. Short, R. N. Perham and G. B. Petersen, *J. Mol. Biol.*, 1991, **218**, 349–364.

58 D. L. Caspar and L. Makowski, *J. Mol. Biol.*, 1981, **145**, 611–617.

59 A. Gonzalez, C. Nave and D. A. Marvin, *Acta Crystallogr. D. Biol. Crystallogr.*, 1995, **51**, 792–804.

60 D. S. Thiriot, A. A. Nevzorov and S. J. Opella, *Protein Sci.*, 2005, **14**, 1064–1070.

61 S. K. Straus, W. R. P. Scott, C. D. Schwieters and D. A. Marvin, *Eur. Biophys. J.*, 2011, **40**, 221–234.

62 A. Goldbourt, B. J. Gross, L. A. Day and A. E. McDermott, *J. Am. Chem. Soc.*, 2007, **129**, 2338–44.

63 A. Goldbourt, L. A. Day and A. E. McDermott, *J. Biol. Chem.*, 2010, **285**, 37051–9.

64 I. V. Sergeyev, B. Itin, R. Rogawski, L. A. Day and A. E. McDermott, *Proc. Natl. Acad. Sci.*, 2017, **114**, 5171–5176.

65 I. V Sergeyev, L. A. Day, A. Goldbourt and A. E. McDermott, *J. Am. Chem. Soc.*, 2011, **133**, 20208–20217.

66 M. Tsuboi, M. Tsunoda, S. A. Overman, J. M. Benevides and G. J. Thomas, *Biochemistry*, 2010, **49**, 1737–43.

67 D. Liu and L. Day, *Science (80-.).*, 1994, **265**, 671–674.

68 J. L. Lorieau, L. a Day and A. E. McDermott, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 10366–71.

69 I. V Sergeyev, S. Bahri, L. A. Day and A. E. McDermott, *J. Chem. Phys.*, 2014, **141**, 22D533.

70 R. N. Purusottam, R. K. Rai and N. Sinha, *J. Phys. Chem. B*, 2013, **117**, 2837–2840.

71 P. van Wezenbeek, T. J. M. Hulsebos and J. G. G. Schoenmakers, *Gene*, 1980, **11**, 129–148.

72 L. Day, C. J. Marzec, S. A. Reisberg and A. Casadevall, *Annu. Rev. Biophys. Biomol. Struct.*, 1988, **17**, 509–539.

73 O. Morag, N. G. Sgourakis, D. Baker and A. Goldbourt, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 971–976.

74 O. Morag, G. Abramov and A. Goldbourt, *J. Phys. Chem. B*, 2011, **115**, 15370–15379.

75 G. Abramov, R. Shaharabani, O. Morag, R. Avinery, A. Haimovich, I. Oz, R. Beck and A. Goldbourt, *Biomacromolecules*, 2017, **18**, 2258–2266.

76 G. Abramov, O. Morag and A. Goldbourt, *J. Phys. Chem. B*, 2011, **115**, 9671–9680.

77 S. Cadars, J. Sein, L. Duma, A. Lesage, T. N. Pham, J. H. Baltisberger, S. P. Brown and L. Emsley, *J. Magn. Reson.*, 2007, **188**, 24–34.

78 R. Das, I. André, Y. Shen, Y. Wu, A. Lemak, S. Bansal, C. H. Arrowsmith, T. Szyperski and D. Baker, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 18978–18983.

79 O. Morag, G. Abramov and A. Goldbourt, *J. Am. Chem. Soc.*, 2014, **136**, 2292–2301.

80 G. Abramov, O. Morag and A. Goldbourt, *J. Magn. Reson.*, 2015, **253**, 80–90.

81 A. Casadevall and L. A. Day, *Nucleic Acids Res.*, 1982, **10**, 2467–2481.

82 S. Tomar, M. M. Green and L. A. Day, *J. Am. Chem. Soc.*, 2007, **129**, 3367–3375.

83 E. Barry, D. Beller and Z. Dogic, *Soft Matter*, 2009, **5**, 2563–2570.

84 D. A. Marvin, L. C. Welsh, M. F. Symmons, W. R. P. Scott and S. K. Straus, *J. Mol. Biol.*, 2006, **355**, 294–309.

85 A. C. Zeri, M. F. Mesleh, A. A. Nevzorov and S. J. Opella, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 6458–6463.

86 W. M. Tan, R. Jelinek, S. J. Opella, P. Malik, T. D. Terry and R. N. Perham, *J. Mol. Biol.*, 1999, **286**, 787–796.

87 L. C. Welsh, M. F. Symmons, C. Nave, R. N. Perham, E. A. Marseglia and D. A. Marvin, *Macromolecules*, 1996, **29**, 7075–7083.

88 M. H. Levitt, D. M. Grant and R. K. Harris, *Encycl. Magn. Reson.*, 2002, 9, 165–196.

89 T. Aharoni and A. Goldbourt, *Chem. - A Eur. J.*, 2018, **24**, 8737–8741.

90 T. Aharoni and A. Goldbourt, *J. Biomol. NMR*, 2018, **72**, 55–67.

91 R. Olsthoorn and J. van Duin, in *eLS*, John Wiley & Sons, Ltd, Chichester, UK, 2011.

92 J. Maniloff, in *eLS*, John Wiley & Sons, Ltd, Chichester, UK, 2012.

93 E. Barbet-Massin, A. J. Pell, J. S. Retel, L. B. Andreas, K. Jaudzems, W. T. Franks, A. J. Nieuwkoop, M. Hiller, V. Higman, P. Guerry, A. Bertarello, M. J. Knight, M. Felletti, T. Le Marchand, S. Kotelovica, I. Akopjana, K. Tars, M. Stoppini, V. Bellotti, M. Bolognesi, S. Ricagno, J. J. Chou, R. G. Griffin, H. Oschkinat, A. Lesage, L. Emsley, T. Herrmann and G. Pintacuda, *J. Am. Chem. Soc.*, 2014, **136**, 12489–12497.

94 L. B. Andreas, K. Jaudzems, J. Stanek, D. Lalli, A. Bertarello, T. Le Marchand, D. Cala-De Paepe, S. Kotelovica, I. Akopjana, B. Knott, S. Wegner, F. Engelke, A. Lesage, L. Emsley, K. Tars, T. Herrmann and G. Pintacuda, *Proc. Natl. Acad. Sci.*, 2016, **113**, 9187–9192.

95 M. Shishovs, J. Rumnieks, C. Diebolder, K. Jaudzems, L. B. Andreas, J. Stanek, A. Kazaks, S. Kotelovica, I. Akopjana, G. Pintacuda, R. I. Koning and K. Tars, *J. Mol. Biol.*, 2016, **428**, 4267–4279.

96 I. Auzat, A. Dröge, F. Weise, R. Lurz and P. Tavares, *Mol. Microbiol.*, 2008, **70**, 557–569.

97 M. Zinke, P. Fricke, C. Samson, S. Hwang, J. S. Wall, S. Lange, S. Zinn-Justin and A. Lange, *Angew. Chemie Int. Ed.*, 2017, **56**, 9497–9501.

98 F. Guo, Z. Liu, P.-A. Fang, Q. Zhang, E. T. Wright, W. Wu, C. Zhang, F. Vago, Y. Ren, J. Jakana, W. Chiu, P. Serwer and W. Jiang, *Proc. Natl. Acad. Sci.*, 2014, **111**, E4606–E4614.

99 G. Abramov and A. Goldbourt, *J. Biomol. NMR*, 2014, **59**, 219–30.

100 G. Abramov, Tel Aviv University, 2015.