DOI: 10.1070/RCR4581

Structure and function analysis of protein – nucleic acid complexes

Svetlana A. Kuznetsova, Tatiana S. Oretskaya

Department of Chemistry, M.V.Lomonosov Moscow State University
Leninskie Gory 1, str. 3, 119991 Moscow, Russian Federation

The review summarizes published data on the results and achievements in the field of structure and function analysis of protein – nucleic acid complexes by means of main physical and biochemical methods, including X-ray diffraction, nuclear magnetic resonance spectroscopy, electron and atomic force microscopy, small-angle X-ray and neutron scattering, footprinting and cross-linking. Special attention is given to combined approaches. The advantages and limitations of each method are considered, and the prospects of their application for wide-scale structural studies *in vivo* are discussed.
The bibliography includes 145 references.

Contents

I. Introduction 445
II. Physical methods of structure and function analysis of protein – nucleic acid complexes 446
III. Biochemical methods of structure and function analysis of protein – nucleic acid complexes 455
IV. Combined approaches and prospects of their application 459
V. Conclusion 460

I. Introduction

Specific protein – nucleic acid (NA) interactions play a vital role in most processes in all living organisms. Such important cellular processes, as replication, transcription, translation, repair and transport of nucleic acids from the nucleus to the cytoplasm, involve specific protein – nucleic acid complexes. There are both stable complexes existing for a long period of time (chromatin, ribosomes, viral particles) and short-lived complexes (complexes of nucleic acids with enzymes and regulatory proteins).

Investigation of the structural basis of the specificity, the mechanism of action and specific features of the functioning of protein – nucleic acid complexes is a key step towards the understanding of their molecular function and role in evolution, which is a challenge in structural biology and bioorganic chemistry. The structure and function analysis provides a basis for obtaining information on cellular processes and developing new approaches to their targeted regulation. The accumulation of structural information about the location and binding sites of NA-recognizing proteins and nucleic acid metabolism enzymes, whose mutations cause various diseases, is an important step in the understanding of molecular mechanisms of the development of pathological processes, the identification of new nosological forms, the elaboration of their diagnostic criteria and the improvement of the therapy. A knowledge of the structure and the structure – function relationship for specific protein – nucleic acid complexes underlies the design of a new generation of antitumour, antibacterial and antiviral drugs, as well as the diagnosis of the diseases.

Data on the structures of protein – nucleic acid complexes can be obtained by means of X-ray diffraction, nuclear magnetic resonance spectroscopy and electron microscopy coupled with chemical and immunochemical methods allowing one to assess the spatial configuration of the complexes and the mutual arrangement of individual components. The chemical modification with bifunctional reagents, monoclonal and polyclonal antibodies, spectroscopic and computational methods are widely used in such investigations.[1]

A large number of original papers and reviews concerning many aspects of structural studies of protein – nucleic acid complexes (see, for example, reviews[2–4]) have been published in the past decade. However, most of these publications are devoted to particular methods, methodological approaches and experimental techniques. For instance, the application of paramagnetic enhancement relaxation, as measured by NMR spectroscopy, has been considered a review;[5] another review[6] is devoted to the approach based on the analysis of residual dipole – dipole interactions (also called dipolar couplings); the application of atomic force microscopy is discussed in a review;[7] one

S.A. Kuznetsova. Doctor of Chemical Sciences, Leading Researcher at the Department of Chemistry, MSU.
Telephone: +7(495)939 – 3153, e-mail: svetlana@belozersky.msu.ru
Current research interests: protein – nucleic acid interactions, post-synthetic modification and DNA repair, cross-linking of proteins to nucleic acids.
T.S. Oretskaya. Doctor of Chemical Sciences, Professor at the same Department.
Telephone: +7(495)939 – 5411, e-mail: oretskaya@belozersky.msu.ru
Current research interests: synthesis of modified nucleosides and oligonucleotides, protein – nucleic acid interactions, design of specific reagents for affinity modification of proteins.

Received 29 June 2015
Translation: T.N.Safonova

© 2016 Russian Academy of Sciences and Turpion Ltd.

more review[8] summarizes data on the use of photocross-linking to investigate different steps of the mechanism of nucleotide excision repair.

Some publications discuss the application of a particular method to structural studies of one class of biopolymers (for example, a review[9] summarizes data on the use of NMR spectroscopy for the structure and function analysis of protein–DNA complexes; the corresponding data for protein–RNA complexes have been surveyed[10]). Some other publications are limited to one particular class of biopolymers having a certain specificity (for example, the use of a combined approach based on cross-linking and immunoprecipitation to study RNA complexes with RNA-binding proteins, whose mutations cause central nervous system disorders;[11] the application of NMR spectroscopy to the structural analysis of RNA-binding proteins involved in the regulation of mRNA metabolism[12]).

This review summarizes the results and advances in the field of structure and function analysis of protein–nucleic acid complexes by means of all main physical and biochemical methods, including X-ray diffraction, nuclear magnetic resonance spectroscopy, electron and atomic force microscopy, small-angle X-ray and neutron scattering, footprinting and cross-linking, which have been published mostly in the past decade. Special attention is given to the use of combined approaches. Advantages and limitations of each method are described, and the prospects of their applications for wide-scale structural studies *in vivo* are considered.

II. Physical methods of structure and function analysis of protein–nucleic acid complexes

Universal approaches to the analysis of the structure and specific features of functioning of protein–NA complexes with a different degree of complexity can be arbitrarily divided into physical and chemical. Physical methods are based on the determination of the three-dimensional structures of protein–NA complexes at atomic resolution. These methods include X-ray diffraction, nuclear magnetic resonance spectroscopy, high-resolution electron, optical and neutron microscopy, small-angle X-ray and neutron scattering. The following two physical methods play a key role in obtaining information on the structures of biological molecules at high resolution (< 2 Å): X-ray crystallography and NMR spectroscopy. Most of the atomic-resolution structures of proteins, nucleic acids and protein–nucleic acid complexes have been determined by one of these methods. Thus, as of May 2015, the Protein Data Bank (PDB) maintained at the Rutgers University Center for Structural Biology (http://rcsb.org/pdb) contained a total of 4935 structures of protein–nucleic acid complexes, 4742 of which were determined by X-ray diffraction and NMR spectroscopy.[13]

II.1. X-Ray diffraction studies of protein–nucleic acid complexes

Due to the use of synchrotron radiation sources, the design of new types of detectors and the development of new programs for the data processing and phasing, protein X-ray crystallography has gained most popularity. More than 90% of the total number of the structures of protein–nucleic acid complexes were determined by this method. The Nobel Prize in Chemistry for 2009 was awarded jointly to T. Steitz, V. Ramakrishnan and A. Yonath for studies of the structure and function of the ribosome, which is a complex ribonucleoprotein assembly composed of ribosomal proteins, rRNA molecules and associated translation factors, using the X-ray diffraction techniques.[14]

X-Ray crystallography is based on diffraction that occurs as a result of scattering of X-ray waves by electrons of different atoms of an ordered substance when the waves pass across its three-dimensional crystal lattice. The diffraction data processing gives electron density maps and unit cell parameters of the crystalline compound. The accuracy of X-ray diffraction data depends on the resolution of the diffraction pattern, which is influenced by the size and quality of the crystals, the intensity of the X-ray source, the efficiency of detector systems, as well as the power of computational techniques and algorithms.

Synchrotron radiation sources generating a powerful X-ray beam allow the user to collect better data in a shorter time, which is of great importance for investigation of crystals of protein–nucleic acid complexes, which can be subjected to radiation damage.[15] In this case, smaller crystals can be used, which are considerably easier to grow. This is an obvious advantage since growing the crystals of protein–nucleic acid complexes remains the most difficult and least predictable part of the X-ray diffraction experiment.[16] Besides, it should be noted that synchrotron radiation sources, as opposed to conventional X-ray tubes, offer an intense continuum of radiation wavelengths, and one is able to select any desired wavelength. This makes it possible to solve the phase problem by means of specially designed algorithms based on the simultaneous use of the scattering data at different wavelengths.[17]

At present, X-ray diffraction techniques are successfully used for structural studies of complex multicomponent protein–nucleic acid complexes. For instance, McGinty *et al.*[18] studied the complex of the human PRC1 (polycomb repressive complex 1) proteins with nucleosome by the X-ray diffraction method using synchrotron radiation.

The PRC1 complex belongs to the Polycomb group of epigenetic enzymes, which suppress the expression of regulatory genes in many eukaryotes. The Ring1B and Bmi1 E3 ubiquitin ligases included in this complex, as well as the UbcH5c E2 ubiquitin-conjugating enzyme, are involved in ubiquitination (the multistep ATP-dependent process of the covalent binding of the regulatory protein ubiquitin to a target protein) of the nucleosomal histone H2A at the residue Lys119. The X-ray crystallography of the complex of the Ring1B, Bmi1 and UbcH5c proteins with nucleosome (Fig. 1, 3.28 Å resolution) revealed how the specific recognition of nucleosome occurs and elucidated the role of the E2 enzyme in this process.[18]

Structural data obtained by X-ray diffraction provide new information on the mechanisms of cellular processes involving proteins and nucleic acids and the mechanisms of their regulation. Thus, new molecular mechanisms of the control of nitrogen homeostasis in the bacteria *Bacillus subtilis* were proposed based on the results of X-ray diffraction study for complexes of DNA with the DNA-binding regulatory proteins, TnrA and GlnR, of the new structural family of dimeric proteins.[19]

Let us mention the following recent advances in X-ray crystallography of protein–nucleic acid complexes: the determination of the crystal structure of the complex of human immunodeficiency virus type 1 reverse transcriptase with DNA and α-carboxy nucleoside phosphonates

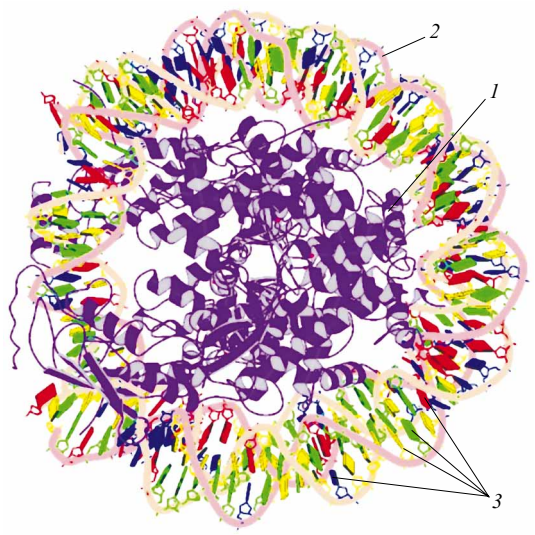


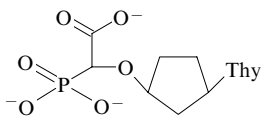
Figure 1. Crystal structure of the Ring1B–Bmi1–UbcH5c complex bound to minimal nucleosome (Brookhaven PDB code 4R8P).[18]
(*1*) Secondary structure elements of the proteins; (*2*) the sugar–phosphate backbone of DNA; *3*, heterocyclic bases of DNA.

(α-CNP), which are new synthetic mimetics of nucleoside triphosphates having a high therapeutic potential (Structures **1**–**4**);[20] the investigation of the structural aspects of hepatitis C virus RNA replication by determining the crystal structures of the complexes of hepatitis C virus RNA-dependent RNA polymerase with RNA templates, RNA primers, incoming nucleotides and catalytic metal ions during both initiation and elongation of RNA synthesis;[21] the establishment of the nature of interactions responsible for conformational flexibility and selectivity of RNA upon the formation of a specific complex with human immunoglobulin IgG1;[22] the determination of the structural basis for sequence-specific recognition of DNA by C-proteins — regulatory proteins of a new class involved in the restriction–modification system and the construction of a model for their interaction with DNA;[23, 24] the structural analysis of the interleukin-6 complex with the ligand SOMAmer, which is a nucleic acid fragment consisting of two structural domains (a G-quadruplex and a double-stranded hairpin structure);[25] some other investigations.[26–29]

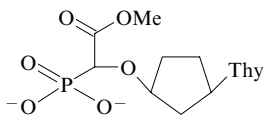
A new impulse to the development of X-ray crystallography has been given by the designation of modern powerful sources of X-ray radiation — X-ray free-electron lasers — and putting them into operation. These lasers generate extremely intense short X-ray pulses. The pulse duration can vary from several hundred femtoseconds down to a few femtoseconds. The X-ray radiation emitted by such lasers can be tuned over a wide range of wavelengths and the extremely high intensity can be achieved, which substantially exceeds that of third-generation synchrotron radiation sources. The radiation power is such that nano-sized crystals of biological objects can be used[30] or even single copies of large macromolecular complexes can be investigated without resorting to their crystallization,[31] which opens up new prospects for the application of X-ray diffraction in structural studies of complex biological objects.[32–34] Thus, the use of femtosecond exposure time opens up new opportunities for the molecular imaging of biochemical reactions at room temperature in solution. In the diffract-and-destroy mode, an extremely intense laser pulse virtually evaporates the crystal; however, the diffraction pattern of the object can be obtained before radiation damage destroys the sample.[33]

Meanwhile, X-ray crystallography is not always straightforward because growing crystals of biopolymers and their complexes can be a serious bottleneck and, as a rule, this method does not provide an insight into the dynamics of interactions.[35, 36] The crystallization of protein–nucleic acid complexes is a more difficult problem compared to the protein crystallization. Crystals of some complexes can be grown only after the removal of certain conformationally labile fragments. Thus, one of the obstacles to the formation of crystals of protein–nucleic acid complexes is flexibility of the nucleic acid molecule, mainly of its part directly involved in the interaction with the protein.[10] Besides, it was noted that protein–nucleic acid complexes often form more fragile and X-ray radiation-sensitive crystals compared to protein crystals; the resolution achieved for such structures is usually not high.[37] It should also be taken into account that the whole complex or its components can undergo rearrangements under crystallization conditions. In addition, the level of radiation damage of individual components of protein–nucleic acid complexes is higher when using third-generation intense synchrotron radiation sources.[15]

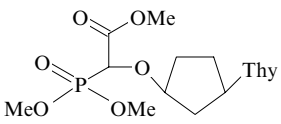
Structures 1–4



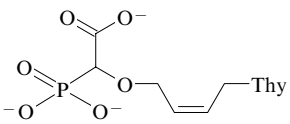
1 (α-Carboxythymidine phosphonate, α-CTP)



2 (α-CTP ester)



3 (α-CTP ester)



4 (Acyclic analogue of α-CTP)

Thy is thymine

II.2. Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance spectroscopy is a good alternative to X-ray diffraction as applied to complexes of nucleic acids with enzymes and regulatory proteins. This method yields high-resolution information on their structures, dynamic characteristics and the nature of specific interactions in solution with no restrictions on the mutual arrangement of atoms imposed by the crystal packing. In the past two decades, NMR has become one of the most powerful instrumental techniques in molecular and structural biology. As of May 2015, the PDB database contained 225 structures of protein–nucleic acid complexes determined by NMR spectroscopy (http://rcsb.org/pdb).[13] An advantage of this method is that it allows researchers to study biomolecules under near physiological conditions. The structure and function analysis by means of NMR

spectroscopy provides information on the three-dimensional structures and intramolecular dynamics (conformational mobility) of the object to be studied, the binding sites of particular biomolecules and the character of their interactions, including hydrogen bonds, π–π interactions, electrostatic effects and charge redistribution.[2, 9, 12]

Recent methodological advances in NMR spectroscopy open up new opportunities for detailed investigation of protein–nucleic acid complexes, including complexes with a high molecular mass (about 50 kDa) and allows one to obtain general structural information for molecules with a mass of several hundred kDa.[38–40] Let us mention such advances as heteronuclear multidimensional NMR spectroscopy, the approach based on the analysis of residual dipolar couplings (RDCs), the use of paramagnetic relaxation enhancement (PRE), transverse relaxation optimized spectroscopy (TROSY) for measurements of large biomolecules; solid-state NMR spectroscopy.

II.2.a. Heteronuclear multidimensional NMR methods

Heteronuclear multidimensional NMR methods are based on the magnetization transfer from protons to heavy nuclei (^{13}C, ^{15}N and ^{31}P) *via* spin–spin couplings. Investigations are often perfomed by employing a combination of classical methods (COSY is correlation spectroscopy; TOCSY is a complementary technique for the detection of all protons belonging to one spin system, for example, to one amino acid residue of the protein; NOESY is nuclear Overhauser effect spectroscopy) with heteronuclear correlation methods based on correlations between carbon or nitrogen atoms and their attached protons of ^{13}C- and ^{15}N-enriched proteins or nucleic acids (HSQC is heteronuclear single-quantum correlation spectroscopy, HMBC is heteronuclear multiple bond correlation spectroscopy). The HSQC technique is used to find correlations between resonances of ^{13}C or ^{15}N nuclei and their attached protons. The HMBC techniques gives correlations between heavy ^{13}C (^{15}N, ^{31}P) nuclei and protons that are two to four bonds away. This approach provides information not only on protons but also on ^{13}C, ^{15}N and ^{31}P nuclei. These additional data are used for the determination of the three-dimensional structure of the object and its dynamic characteristics. Heteronuclear NMR spectroscopy offers the opportunity to study the structures and dynamic characteristics of complexes with a molecular mass of up to 40–50 kDa.[41]

Cieniková *et al.*[42] applied heteronuclear NMR spectroscopy to investigate the structural and molecular basis for the interaction between the N-terminal RNA recognition motif of the human regulatory factor hnRNP C (heterogeneous nuclear ribonucleoprotein C) and U-rich penta- and heptaribonucleotide sequences of RNA. The hnRNP C protein is the main component of the protein core of the nuclear ribonucleoprotein particle, hnRNP, composed of the newly synthesized RNA and a complex of proteins that serve to condense and package the growing RNA transcript. This protein contains the N-terminal RNA recognition motif (RRM) that is known to bind uridine-rich RNA sequences (Fig. 2). Cieniková and co-workers measured 2D and 3D NMR spectra of samples using TOCSY, HSQC, NOESY, *etc.* technqiues. Samples contained the ^{15}N- or ^{15}N/^{13}C-labelled protein and ^{13}C-enriched RNA fragments. The study showed that RRM contains five RNA-binding motifs that recognize oligouridine tracts of RNA with a gradient from the 5′- to 3′-terminus. The

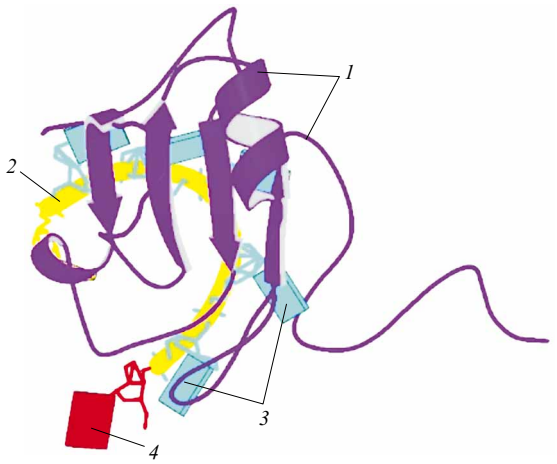


Figure 2. Structure of the complex of the N-terminal RNA recognition motif of the RNA regulatory factor hnRNP C with the heptaribonucleotide 5′-AUUUUUC-3′ (Brookhaven PDB code 2MXY).[42]
(*1*) Secondary structure elements of the proteins; (*2*) the sugar–phosphate backbone of the heptaribonucleotide; *3*, nucleosides; *4*, two 5′-terminal nucleotides that are in contact with the secondary structure elements of the protein.

authors used a new approach which combines the results of the structural investigation and the thermodynamic description of RNA binding to model the saturation of cellular oligouridine tracts by the hnRNP C protein. This model was in complete agreement with the experimentally observed distribution of the hnRNP C full-length protein on short oligouridine tracts of RNA transcripts. Therefore, it was found that the RNA recognition motif (RRM) is the primary RNA-binding domain of the regulatory factor hnRNP C tetramer.

Heddi *et al.*[43] applied heteronuclear NMR techniques to study the structure of the complex of 18-amino acid peptide of the human RNA helicase RHAU with the parallel-stranded DNA G-quadruplex. It is known that the RHAU helicase of the DEAH (Asp-Glu-Ala-His) helicase family binds and unwinds G-quadruplexes in RNA and DNA.

The complex composed of a ^{13}C,^{15}N-labelled peptide and a G-quadruplex was studied by 2D and 3D NMR spectroscopy. The structure of the specific complex of the 18-amino acid peptide (the RHAU fragment containing the G-quadruplex-binding domain) with the G-quadruplex was determined based on the analysis of the NMR spectra. A knowledge of the structure of this complex gave the authors an insight into the specific recognition of parallel-stranded DNA G-quadruplexes (G-quadruplexes, in which all the strands are oriented in the same direction) by the RHAU helicase. It was found that the peptide covers the terminal guanine base tetrad (G-tetrad) and clamps the quadruplex core using electrostatic interactions between three positively charged amino acid residues (Lys8, Arg10 and Lys19) and negatively charged phosphate groups. This type of interactions is characteristic of most of the ligands that specifically interact with G-quadruplexes.

The same approach was used to determine the structure of the complex of the *Escherichia coli* antitoxin MazE, the N-terminal DNA-binding domain of which is involved in the regulation of its own promoter region, with a palin-

dromic DNA sequence. This study gave an insight into the mechanism of regulation of transcription of the *mazEF* gene encoding the *E. coli* MazF and MazE toxin–antitoxin modules.[44]

II.2.b. Approach based on the analysis of residual dipolar couplings

Since recently, the analysis of residual dipolar couplings (RDCs) between magnetic nuclei in systems aligned in magnetic field has been extensively used to study biological macromolecules.[45, 46]

In solution, the rotational motion of molecules is fully random, resulting in that dipolar interactions between magnetic nuclei within these molecules are averaged out. Upon dissolution of molecules in a lyotropic liquid crystal system, the molecular motion becomes anisotropic due to the interaction with magnetically oriented molecular assemblies. This leads to an incomplete averaging of dipole–dipole interactions (residual dipolar couplings) and manifests in the NMR spectra as changes in the signal splitting compared to an isotropic medium. The molecular structure can be determined from a set of measured residual dipolar coupling constants (RDCCs) between several pairs of magnetic nuclei (for example, between ^{15}N and ^{1}H nuclei of the amide groups of the protein). This approach allows for the determination of the structural parameters based not on the local information (for example, the mutual arrangement of atoms estimated from nuclear Overhauser effects) but on the orientation of molecular fragments with respect to the external magnetic field. An analysis of RDCCs provides information on the orientation of biomolecules and their structural domains in macromolecular complexes and makes it possible to independently determine the structures of these compounds in solution.[6, 47, 48]

For instance, Williams Jr. *et al.*[49] analyzed RDCCs in order to establish the structure of the macromolecular complex composed of two transcription factors, Oct1 and Sox2, and a DNA fragment containing the recognition sequence for the Sox2 factor.

It is known that transcription factors of the Oct and Sox families control various aspects of the development of the nervous system and human embryogenesis by targeting specific DNA sequences. These factors indirectly interact with each other in the complex with DNA and exert an synergistic effect on the transcription activation. The Oct proteins are composed of two DNA-binding domains (POU_S and POU_{HD}) separated by a flexible linker. The binding of the Sox2 transcription factor to the DNA sequence (Hoxb1 regulatory element) gives rise to the Oct1·Sox2·Hoxb1-DNA complex composed of two proteins and the Hoxb1-containing DNA fragment. The formation of this ternary complex increases the site-specific affinity of Oct1 for DNA by locking the POU_S domain in the orientation required for the efficient binding.

The 3D NMR spectra of samples containing the isotope-labelled Oct1 and Sox2 proteins, with different combinations of isotopes (^{15}N, $^{15}N/^{13}C$, $^{15}N/^{13}C/^{2}D$), and a 19-mer DNA fragment were measured using the TROSY and other techniques. The analysis of the NMR spectra of the Oct1($POU_S + POU_{HD}$)·Sox2·Hoxb1-DNA complex (42 kDa), including the analysis of RDCCs, demonstrated that the Sox2 transcription factor directly binds to the POU_S DNA-binding domain of the factor Oct1. It was hypothesized that this is one of the steps in the regulation of DNA transcription.[49]

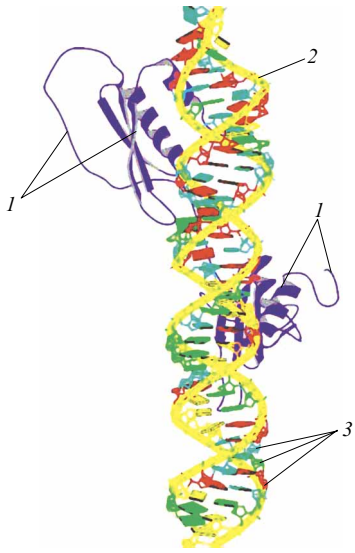


Figure 3. Structure of the complex of ADAR2 dsRBM12 with GluR-2 R/G (Brookhaven PDB code 2L3J).[50]
(*1*) Secondary structure elements of the proteins; (*2*) the sugar–phosphate backbone of RNA; (*3*), heterocyclic bases of RNA.

This approach was also used to study the sequence-specific recognition of double-stranded RNAs by adenosine deaminases.[50] The structure of the complex of dsRBMs (RNA-binding domain of adenosine deaminase ADAR2) with pre-mRNA encoding the R/G editing site of GluR-2 was determined (Fig. 3). It was shown that the sequence-specific recognition by the protein occurs at the minor groove of double-stranded RNA. The structural study revealed a molecular mechanism of recognition of the RNA nucleotide sequence by adenosine deaminase, which plays a key role in the editing (deamination of adenosine), and provided an explanation of how numerous dsRBM-containing proteins are involved in the post-translational regulation of gene expression.[50]

II.2.c. Paramagnetic relaxation enhancement

Paramagnetic compounds containing unpaired electrons can have a considerable effect on the magnetic environment of nuclei, thus shifting NMR signals and changing the nuclear relaxation rate. This is due to the fact that the magnetic moment of the unpaired electron is $\sim 10^3$-fold larger than the magnetic moment of the nucleus.[51] The paramagnetic relaxation enhancement (PRE) observed in the presence of paramagnetic compounds underlies the technique capable of supplying valuable information on the dynamic structure of biological macromolecules, which cannot be obtained by conventional NMR and ESR methods. This method is based on the analysis of electron–nucleus interactions between an uncompensated electron spin (unpaired electrons) and resonating nuclei. The signal broadening values (or relaxation rates) due to nuclear paramagnetic relaxation provide information on the structures of molecules coordinated to a paramagnetic centre.[52] In studies of the structures of biological macromolecules, the strict dependence of the PRE linewidth on the distance between the nucleus and the unpaired electron (the paramagnetic centre) is used to evaluate this distance.

Stable nitroxide radicals of the imidazoline series and transition metal complexes, including rare-earth metal (lanthanide) complexes, are most commonly used as paramagnetic compounds (spin labels) to investigate biomolecules by NMR methods.[52] Rare-earth metal complexes have some advantages over other transition metal complexes, allowing the measurement of well-resolved NMR spectra and analysis of larger macromolecules.[53, 54]

Spin-probe and spin-label techniques can be employed to study the surface topography of biopolymers, conformational transitions, the structure and dynamics of active sites, as well as to obtain unique information enabling one to detect and visualize transient (low-population) states of macromolecules and their complexes.[5, 55–58]

For instance, Iwahara and Clore[55] applied this technique to characterize the search process whereby the HOXD9 human sequence-specific transcription factor recognizes and binds to the specific nucleotide sequence. For this purpose, the authors investigated the interaction of the $^{2}H/^{15}N$-labelled HOXD9 homeodomain with a 24-bp DNA duplex containing simultaneously the specific binding site and modified thymidine residues (d)T-EDTA-Mn^{2+}(Ca^{2+}) (one to four residues per DNA-duplex molecule). The analysis of the NMR spectra of HOXD9 homeodomain–DNA complexes (by means of the 2D $^{1}H-^{15}N$ HMBC technique) revealed the formation of a transient intermediate — the ternary complex of free DNA with the DNA–protein complex. A knowledge of the mechanism of formation of the specific HOXD9–DNA complex (including the protein search for the target site), involving the formation of a ternary complex as one of the steps, provides an explanation for the fact that, in the target search process, the rate of protein translocation on DNA is substantially (approximately three orders of magnitude) higher than the dissociation rate. Iwahara and Clore[55] presented the first direct evidence for the formation of intermediate complexes of the protein with nonspecific DNA sequences as a result of the following two processes: the intramolecular sliding of the protein along the DNA and the intermolecular translocation of the protein between DNA molecules.

Cai *et al.*[59] used this approach to determine the structure of the complex of the ARID domain (a conserved domain that specifically binds to AT-rich DNA sequences) of the Mrf2 protein with the specific DNA sequence. The human transcription cofactor Mrf2, which belongs to a new class of ARID proteins, plays a critical role in lipogenesis.

The complex composed of the ^{15}N-enriched ARID domain of Mrf2 and the 14-mer AT-rich DNA duplex containing 3-(2-iodoacetamido)-2,2,5,5-tetramethylpyrrolidin-1-oxyl at the 3′(5′)-end was studied by the 2D HSQC technique. The analysis of the NMR spectra demonstrated that the Mrf2 protein contacts with DNA mainly using two flexible loops, L1 and L2. Thus, positively charged amino acid residues of the loop L1 contact with the sugar–phosphate backbone of the DNA, while amino acid residues of the loop L2 and some amino acid residues of the adjacent loops H1 and H5 make contacts with AT base pairs in the major groove of the DNA. Cai *et al.*[59] demonstrated that, despite the structural diversity of DNA-binding proteins containing ARID domains, they form similar specific contacts with DNA. Thus, the Mrf2 and Dri (Dead ringer) proteins make similar contacts with DNA.

Mackereth *et al.*[60] applied the paramagnetic relaxation enhancement-based NMR technique to study the structure of the complex of the splicing factor U2AF65 with the polypyrimidine tract RNA. The authors studied molecular mechanisms of the key step of the pre-mRNA splicing — the recognition of the 3′-splice-site-associated polypyrimidine tract RNA by the large subunit of the human splicing factor U2AF65. It was shown that, in the presence and in the absence of the polypyrimidine tract, the RNA recognition domains of U2AF65 adopt different conformations. The equilibrium between the open and closed conformations functions as a molecular rheostat that quantitatively correlates the natural variations in the polypyrimidine tract nucleotide composition, length and functional strength to the efficiency to recruit the splicing factor to the intron during spliceosome assembly.

II.2.d. Solid-state NMR spectroscopy

Solid-state NMR methods are characterized by low molecular mobility and require rapid rotation of the sample (tens of thousands of revolutions per second) at the so-called magic angle (54°44′) with respect to the external magnetic field (magic angle spinning, MAS).[41] This rotation leads to the averaging of dipole–dipole interactions between magnetic nuclei, which substantially complicate the spectra of even simple compounds. The time averaging results in that other anisotropic characteristics of the solid sample become equal to zero or negligibly small.[61] Due to rapid rotation of the sample at the magic angle, the chemical shifts and dipole–dipole interactions are similar to those observed in solution.[41]

Solid-state NMR spectroscopy yields high-resolution spectra for solid and viscous samples.[62] This method has a considerable potential. Thus, there are no sample size limits for solid-state NMR methods. The ^{13}C, ^{15}N, ^{19}F and ^{31}P NMR spectroscopy is most often used in these experiments.

For example, solid-state NMR spectroscopy was employed[62, 63] to determine the structure of the 26-mer fragment of the small non-coding nucleolar RNA from *Pyrococcus Furiosus* containing conserved sequences [C (UGAUGA) and D (CUGA)] in the complex with the ribosomal protein L7Ae (21 kDa). Small nucleolar RNA are known to play a crucial role in biogenesis of eukaryotic ribosomal RNA (rRNA). In a complex with a specific set of proteins, they are involved in the post-translational 2′-O-methylation of rRNA nucleotides.

The NMR experiments were performed on selectively ^{13}C- and ^{15}N-labelled RNA sequences, which were produced by *in vitro* transcription using T7 polymerase, by employing mixtures of ^{15}N-, ^{13}C- or $^{13}C,^{15}N$-labelled nucleoside triphosphates. A microcrystalline sample of the RNA–L7Ae complex prepared for the analysis contained 4 mg of the RNA and 6 mg of the protein.

A series of experiments showed that the RNA fragment contains the so-called kink-turn (or k-turn) — a secondary structure motif that induces a sharp kink in the RNA helix. The geometry of this motif established by NMR spectroscopy is in good agreement with the parameters determined by X-ray diffraction for similar complexes containing RNA from *Archaeoglobus fulgidus* and *Solpholobus solfataricus*. In particular, it was found that, in the complex with the L7Ae protein, the kink angle of the RNA helix is approximately 23°, which corresponds to this angle for all the characterized complexes of similar RNA with the L7Ae protein. Carlomagno an co-workers[62, 63] believed that this approach will be useful for investigations of the structures

of RNA and their complexes with proteins, which are difficult to crystallize.

In recent years, experimental techniques have been extensively developed in order to overcome limitations associated with an increase in the transverse relaxation rate with increasing molecular mass of the target (*e.g.*, transverse relaxation optimized spectroscopy, TROSY). These techniques yield high-resolution spectra of biological macromolecules with a molecular mass of up to 100 kDa and even higher and enable researchers to study supramolecular complexes with a molecular mass of up to 1 MDa.[10,64–66] This approach holds promise also for the structural analysis of protein–nucleic acid complexes.

Thus, Sahu *et al.*[67] used the experimental TROSY technique to investigate the translocation of the HOXD9 sequence-specific human transcription factor between DNA molecules and, in particular, to evaluate the activation energy of this process. The translocation plays an important role in the targeting of the HOXD9 transcription factor and involves the direct transfer of the protein molecule from one binding site to another bypassing the intermediate dissociation step, in which the protein is not complexed with DNA in solution (see also Section II.2.c and the study [55]). For this purpose, the interaction of the $^{2}H/^{15}N$-labelled HOXD9 homeodomain with the 24-bp DNA duplex containing the homeodomain-binding site was studied. The analysis of the 2D NMR spectra established that the energy barrier for translocation is lower than that for the dissociation of the specific HOXD9–DNA complex, which accounts for the absence of the intermediate dissociation step in the translocation of the HOXD9 protein.

Sugitani *et al.*[68] employed the NMR method using the TROSY technique to perform a comparative study of the XPA_{98-219} DNA-binding domain of the XPA protein, which was discovered more than 20 years ago, and the proposed new XPA_{98-239} construct (17 kDa fragment of the full-length human XPA protein).

The XPA (xeroderma pigmentosum complementation group A) protein is one of the key factors in the nucleotide excision repair of DNA. Mutations in the XPA gene region encoding the DNA-binding domain cause the genetic disorder — xeroderma pigmentosum complementation group A (XPA) characterized by UV-induced skin changes and high risk of developing skin cancer.[69] Although the XPA_{98-219} DNA-binding domain is well-characterized, the molecular aspects of its specific binding to DNA are poorly understood because this domain binds to DNA less efficiently than the full-length protein.

The analysis of the 2D NMR spectra of the XPA_{98-219} and XPA_{98-239} domains in the absence and in the presence of DNA substrates (by means of $^{15}N-^{1}H$ TROSY HSQC spectroscopy) demonstrated that these domains have the identical topology and globular structure. Besides, the XPA_{98-239} domain adopts a stable conformation, does not aggregate and binds to DNA as efficiently as the full-length protein. Taking into account the fact that mutations in the amino acid residues in the F219–T239 sequence are associated with the appearance of some groups of xeroderma pigmentosum, *i.e.*, this region is also functionally significant, Sugitani *et al.*[68] suggested that XPA_{98-239} is a more suitable model for investigation of structural and biochemical aspects of the interaction of XPA with the DNA target compared to XPA_{98-219}.

Walavalkar *et al.*[70] used NMR spectroscopy to study the search process for mCpG/TpG mismatches by the MBD4 protein (protein 4 that binds to methylcytosine-containing DNA fragments). Unlike other members of the MBD protein family that bind to methylcytosine-containing DNA sequences, the MBD4 protein acts in the DNA mismatch repair system as DNA glycosylase specifically targeting mCpG/TpG mismatches. The latter arise from deamination of methylcytosine. This protein is composed of two domains: the N-terminal MDB domain ($MDB4_{MDB}$) and the C-terminal glycosylated domain ($MDB4_{GD}$) separated by a long linker. The authors studied the binding of the N-terminal $MDB4_{MDB}$ domain to DNA. The investigations, including those with the use of 2D TROSY techniques (samples contained the $^{13}C/^{15}N$- and $^{13}C/^{15}N/^{2}H$-enriched protein and 17–30-mer DNA sequences), showed that the $MDB4_{MDB}$ domain binds more efficiently to mCpG sites in DNA than to mismatches, unmethylated and hydroxymethylated DNA. It was demonstrated that $MDB4_{MDB}$ slowly migrates between the binding sites located on different DNA molecules and rapidly moves by hops between two sites on one DNA molecule *via* the so-called hopping mechanism. These results confirm the model, according to which the role of the N-terminal $MDB4_{MDB}$ domain is to target the intact MBD4 protein to mCpG sites and to enhance the efficiency of the DNA scanning by the protein, which, in turn, enhances the efficiency of mismatch repair by the C-terminal ($MDB4_{GD}$) domain of the protein possessing DNA glycosylase activity.

Experimental TROSY techniques were used also in other studies, for example, in order to determine the structure of the macromolecular complex composed of two transcription factors, Oct1 and Sox2, and the DNA fragment, which also contains the Sox2 recognition sites.[49]

Nuclear magnetic resonance spectroscopy is successfully employed for the ligand screening and the search for their optimal structures. Koharudin *et al.*[71] used NMR spectroscopy to find the crystallization conditions for the protein–DNA complex. The optimal structure of the DNA ligand consisting of a double-stranded fragment and a GGA-containing single-stranded fragment was found for the CsrA protein by NMR screening. The NMR measurements showed that the protein binds two DNA oligomers. These data were taken into account to perform the crystallization of the complex of the CsrA protein with DNA.

However, although new methodological approaches substantially extended the scope of the method and allowed researchers to investigate a wide range of molecular complexes, it has a number of limitations. Thus, NMR spectroscopy in liquid phase is applicable only to relatively small complexes, which are soluble, homogeneous, do not aggregate and remain stable over the period of the experiment. This method cannot be used to investigate some multisubunit complexes, or if considerable changes in the structure of biopolymers occur during the interaction process. The calculation of the structures of the complexes using NMR spectroscopy is a multistep procedure as compared with calculations based on X-ray diffraction data because it is necessary to carry out numerous experiments in order to assign the signals in the NMR spectra prior to the extraction of structural information from the spectra and the data interpretation. Besides, NMR spectroscopy requires considerable amounts (milligram quantities) of compounds. Therefore, the limitations of the method, such as low

sensitivity, molecular mass limits, time-consuming measurements, the necessity of using considerable amounts of stable isotope-enriched biopolymers, expensive equipment and costly experiments, are obstacles to the broad application of the method. Solid-state NMR spectroscopy has a considerable potential for investigation of the structures of biological macromolecules and their complexes, allowing for structural studies of solid and viscous samples as an obvious advantage. The methodology of this approach is being actively developed but, on the whole, except for the absence of molecular size and molecular mass limits, it has the same limitations as NMR spectroscopy in liquid phase.

II.3. Electron microscopy

High-resolution electron microscopy is currently one of the main tools for the investigation of surface topography and imaging of biological materials at the nanometre scale. As of May 2015, the PDB contained 185 structures of protein–nucleic acid complexes determined by electron microscopy (http://www.rcsb.org/pdb).

An electron microscope is a microscope that uses a beam of electrons with an energy from 100–300 keV to 5 MeV instead of light beams, providing a resolution of about 1 nm (in special cases up to 0.1 nm, *i.e.*, a near atomic resolution). Currently, transmission electron microscopes (TEM) and scanning electron microscopes (SEM) are widely used instruments. Transmission electron microscopes create series of planar images of a micro-object. Scanning electron microscopes produce images of a sample by scanning it with an electron microprobe. The main advantages of scanning electron microscopy are a large depth of field coupled with a wide range of continuous magnification (from tens to tens of thousands times) and high resolving power, owing to which SEM generates clear micrographs yielding a characteristic three-dimensional appearance for objects with complex topography.[72] Conventional methods of electron microscopy require the development of special conditions for the sample preparation: chemical fixation and staining using special contrasting agents, for example, heavy metal salts.

Cryo-electron microscopy is a technique, which enables studying unfixed samples, avoiding risks of artifacts caused by fixation. This technique is a form of transmission electron microscopy, where a flash-frozen thin layer ($\sim$100 nm) of a sample is spread on a microscopy grid and is investigated in a vacuum chamber of the microscope at -160 °C. Such samples give high-contrast images without additional treatment.

Cryo-electron microscopy has gained popularity in cellular and molecular biology. This method is widely used in structure–function studies of ribosomes and for investigation of the molecular mechanism of protein biosynthesis. One goal of these studies is to design a new class of antibiotics targeted to bacterial ribosomes. The design of new types of desired antibiotics is one of challenges in modern biomedicine.

Consider the study of Greber *et al.*[73] as one of recent advances in this area. The authors determined the structure of the 50S ribosomal subunit from the archaeon *Methanothermobacter thermautotrophicus* in complex with archaeal translation initiation factor 6 (IF6) by cryo-electron microscopy at 6.6 Å resolution. The molecular binding of IF6 on the ribosome was found to be conserved between the archaeon *M. thermautotrophicus* and eukaryotes. This shed new light on the evolution of the translation in these single-celled organisms that lack a nucleus.

Greber *et al.*[74] determined the three-dimensional structure of the 39S subunit of the porcine mitochondrial ribosome by cryo-electron microscopy at 4.9 Å resolution. A comparison of these results with the data from mass spectrometry and chemical cross-linking gave a detailed insight into the architecture of the exit tunnel, through which the growing polypeptide chain exits the ribosome. This site of the mitochondrial ribosome differs substantially from that in the bacterial analogue ensuring the synthesis and integration of growing hydrophobic polypeptides, which form functional sites of complexes of the electron transport chain (respiratory chain), into the inner membrane.

Amunts *et al.*[75] determined the structure of the yeast mitochondrial large ribosomal subunit by cryo-electron microscopy at 3.2 Å resolution, which enabled a nearly complete atomic model to be built, including the location and packing of 39 ribosomal proteins, 13 of which are unique, as well as the folding of mitoribosomal RNA.[75] The structure of the ribosomal tunnel, unique elements of the E site and the putative site for ribosome integration into the mitochondrial membrane was established.

The structure of the intact human mitoribosome was determined at 3.5 Å resolution.[76]

The structural analysis of the entire nuclear receptor in complex with a inverted repeat of DNA was carried out for the first time by cryo-electron microscopy.[77] Nuclear receptors of thyroid and steroid hormones, which were found in mammals and insects, are among the major targets for therapeutic agents because they regulate gene transcription by binding to DNA and specific ligands. A heterodimeric complex of the insect nuclear receptor USP with the activated ecdysone receptor (EcR, a phytoecdysteroid) binds to the promoter region of the target gene, thus activating its transcription. The structural analysis of the USP–EcR nuclear receptor from the tobacco budworm *Heliothis virescens* in complex with an inverted repeat of the palindromic DNA sequence revealed that, despite the symmetry of palindromic DNA, the complex adopts an asymmetric architecture, in which the ligand-binding domains are positioned 5′-off-centred. A comparison of this structure with the structural characteristics of nuclear receptors in complexes with direct repeats of DNA suggests that DNA is the major allosteric driver in positioning the ligand-binding domains, which serve as the main binding site for transcriptional regulators.

In recent years, considerable progress has been made in developing instrumentation for electron microscopy and improving mathematical methods for image recognition and computational image analysis. This made it possible to perform the cryo-electron microscopy reconstruction at a resolution comparable to that achieved by X-ray crystallography. A breakthrough in this field is the study,[78] in which the three-dimensional structure of the ribosomal complex was determined by cryo-electron microscopy at a resolution (2.65–2.9 Å) higher than that of the structure established earlier by X-ray crystallography (Fig. 4).

An image at such a high resolution was obtained using a transmission electron microscope equipped with a corrector for spherical and chromatic aberrations, which cause serious distortions in the TEM images. The structure of the complex composed of the *E. coli* 70S ribosome, elongation

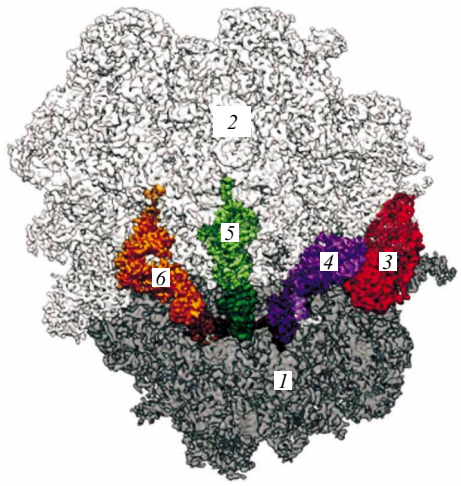


Figure 4. Cryo-electron microscopy map for *E. coli* 70S ribosome in complex with elongation factor EF-TU at 2.9 Å resolution (code EMD-2847 in Protein Data Bank in Europe, http://www.ebi.ac.uk/pdbe/entry/emdb/EMD-2847).[78]
The map shows: 30S ribosomal subunit (*1*), 50S ribosomal subunit (*2*), elongation factor EF-TU (*3*), Phe–tRNAPhe (*4*), fMet-tRNAfMet in the P site (*5*) and deacetylated tRNAfMet in the E site (*6*).

factor Tu (EF-TU), aminoacyl-tRNA and the antibiotic kirromycin contained, for the first time, all modified bases of ribosomal RNA. The role of modified bases in the fine tuning of ribosome functions and the mechanism of action of antibiotics was revealed.

Therefore, apart from the above-mentioned advantages, electron microscopy enables the determination of the real arrangement of components of the complexes. Meanwhile, it should be taken into account that atomic-resolution electron microscopes are very expensive unique instruments, which require high vacuum, high-precision electron optical lenses, complex electronics and specialized computer software for image analysis and processing, special housing and maintenance to isolate the microscopes from vibrations and noises and laborious sample preparation. Besides, in high-voltage electron microscopes, the sample can be damaged by high-energy electrons.[72]

II.4. Atomic force microscopy

Atomic force microscopy (AFM) is one of the most promising research tools that enables the analysis of the structure at the atomic level. This method may be classified as one of scanning probe microscopy techniques based on the measurement of van der Waals forces arising from interactions between the probe and the sample surface. These forces include two opposing components having different signs: a long-range attraction and a short-range repulsion. The AFM probe consists of a nano-sized sharp tip attached at the end of a flexible cantilever. The AFM method enables the measurement of the surface topography of the sample at a resolution from tens of Ångströms down to atomic and even subatomic level.[79, 80]

Atomic force microscopy allows for the visualization of individual macromolecular complexes in real time in air and liquid media, in particular, under near-physiological conditions, the production of high-resolution images of complexes (fractions of a nanometre in the plane of the sample and hundredths of a nanometre perpendicular to the sample surface), which is a serious advantage of this method. It can be used to study individual proteins and DNA molecules, as well as protein–protein and protein–nucleic acid complexes, the characteristic features of their formation and specific interaction.[81–85]

Wang *et al.*[86] studied the interaction of restriction endonuclease BspMI (*Bacillus* species M) with DNA by AFM. It was found that the specific enzyme binding to DNA is induced by looping out of the latter, the number of the loops formed being increased with increasing enzyme concentration. Nonspecific BspMI binding to DNA was found under conditions when all recognition sites were bound to the enzyme. The number of nonspecific complexes was >8% of the total number of BspMI–DNA complexes. A further increase in the enzyme concentration and in the incubation time give rise to additional DNA loops. The authors attributed this fact to enzyme aggregation. The schematic view of the interaction between BspMI and DNA is shown in Fig. 5.

Hamon *et al.*[87] used the AFM method to visualize DNA in complexes with single-stranded DNA-binding (SSB) proteins that bind with a high affinity to single-stranded DNA sequences. This binding is a transient state in many DNA metabolic processes, such as repair, replication and recombination. The improvement of the procedure for the absorption of the complexes on a substrate (mica) enabled the study of DNA binding to various SSB proteins (*E. coli* SSB, gp32 and yRPA). The authors arrived at the conclusion that AFM coupled with biochemical methods can be used to investigate the dynamics of the DNA processes, involving the formation of intermediates.

The AFM method was successfully employed to visualize the site-specific DNA binding to nicking endonuclease Nt.BspD6I (*Bacillus* species D6), which generates a DNA

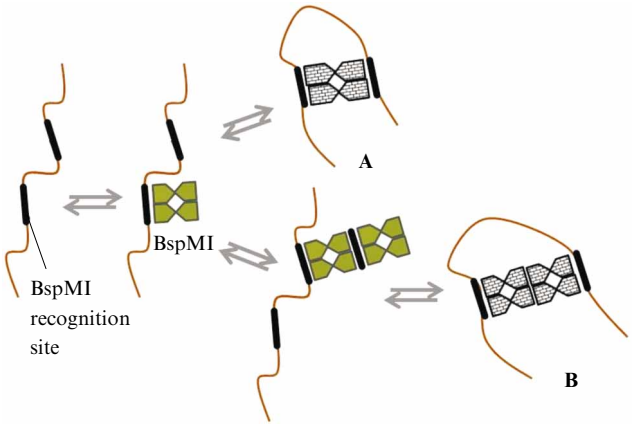


Figure 5. Schematic view of the specific interaction between restriction endonuclease BspMI and DNA.[86]
First, two subunits of the BspMI tetramer bind to the recognition site in DNA (the BspMI tetramer is represented as four green subunits, the DNA-binding surface of BspMI is marked by flat edges). Then the other two subunits of the tetramer bind to the second recognition site, which causes conformational changes in all four subunits (these subunits are dashed), and DNA is looped out (**A**). The BspMI tetramer bound to DNA aggregates with the second protein tetramer, which, in turn, also binds to DNA (**B**).

single-strand break. This endonuclease has unique functions and is commonly used in a number of biotechnological methods. The conditions and the DNA : protein ratio were chosen, which made it possible to visualize the resulting DNA–protein complex at high resolution.[88]

It should be noted that AFM has a number of limitations. Thus, it is impossible to generate a full image of the sample in real time (the area of the sample surface under study is probed not all at once), the sample surface can be deformed by the scanning probe and the images can be distorted by external factors (acoustic noise and mechanical vibrations). These issues reduce advantages of this method. Besides, for such samples as protein–nucleic acid complexes, it is necessary to choose proper conditions for the sample preparation taking into account, apart from specific interactions, interactions with the substrate.[89–92]

Hence, a promising way is to integrate an atomic force microscope into hybrid devices, which combine two or three complementary techniques in one instrument. In this way, a comprehensive description of molecular processes can be obtained since one experiment can give simultaneously morphological, electrochemical, mechanical and kinetic information. The combination of AFM with spectroscopic, optical, mechanical or electrochemical methods is considered in the review.[93] In more recent studies, investigation techniques using combined approaches were optimized. Thus, Ebenstein *et al.*[94] showed that the utilization of semiconductor quantum dots, serving as contrast agents, facilitates the imaging of the samples in experiments using a combination of AFM and fluorescence microscopy. The authors demonstrated the advantages of this combined approach by imaging complexes of RNA polymerases on bacteriophage T7 genomic DNA. Sanchez *et al.*[95] suggested that DNA–protein complexes can be studied by a combination of AFM and fluorescence microscopy using fluorescent-labelled polystyrene nanospheres for the alignment of optical and topographic images, which allows the precise localization of different fluorescent labels within complex molecular assemblies.[95]

II.5. Small-angle X-ray and neutron scattering

Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) are commonly used in the structure and function analysis of biopolymers. These diffraction techniques rely on the collection of SAXS or SANS patterns at very small scattering angles in the vicinity of the primary beam at inhomogeneities of the substance. In small-angle scattering experiments, with X-rays or neutrons having a wavelength of a few Ångströms (0.1–1 nm), information about molecules 1–1000 nm in size can be obtained. A key feature of small-angle scattering is that it can be used to analyze disordered systems. This method is often the only way to obtain direct structural information on systems with random or partially ordered arrangement of density inhomogeneities. It enables the study of the molecular-mass and geometric characteristics, as well as the three-dimensional structures of biological macromolecules and their complexes in solution.[96]

Modern small-angle scattering instruments and techniques allow researchers to address an array of biological problems, including folding of biomolecules, location of disordered structural elements, conformational transitions, flexibility and twisting, self-assembly and heteromolecular assembly. As a rule, the small-angle scattering is used in combination with other structural methods, such as X-ray crystallography, NMR spectroscopy, electron microscopy and footprinting, resulting in higher accuracy of structural analysis and better efficiency of structural analysis of multidomain and multicomponent proteins and protein–nucleic acid complexes (see Section IV).

Sanchez *et al.*[97] employed small-angle X-ray scattering to investigate the binding between the catalytic domain of the ComD protein, the ComE protein and the promoter of the comcde operon in the bacterium *Streptococcus pneumoniae*. It is known that the ComD–ComE two-component system controls the competence state of *S. pneumoniae via* the regulation of phosphorylation of the ComE protein. Thus, activated histidine kinase ComD phosphorylates the ComE protein, which, starting from a certain concentration, binds to the promoter of the comcde operon. This leads to activation of the latter (a positive autoregulatory loop), activation of the promoter of the comab operon (extracellular protein export system), activation of the comx operon, which induces late competence genes responsible for the binding and uptake of transforming DNA and all other late transformation steps.[98] The ComE protein can exist both as a monomer and a dimer. Earlier, it has been shown that the non-phosphorylatable ComE(D) (58A) mutant is monomeric in solution, whereas the ComE(D) (58E) active mimic mutant dimerizes *via* REC domains. Histidine kinase ComD is a dimer that adapts two monomers of ComE, one on each side, placing the aspartate residue D58 of ComE in front of histidine H248 of ComD in position that is convenient for intermolecular transfer of the phosphate group. The modelling of ComE(D) (58A) and ComE(D) (58E) in complexes with comcde using small-angle scattering data showed that ComE(D) (58E) exists in the complex as a compact dimer (Fig. 6), whereas ComE(D) (58A) can change its conformation. Different conformational states of ComE induce different bending angles of the comcde promoter, which provides a mechanistic scenario for the activation: the phosphorylation of ComE enhances

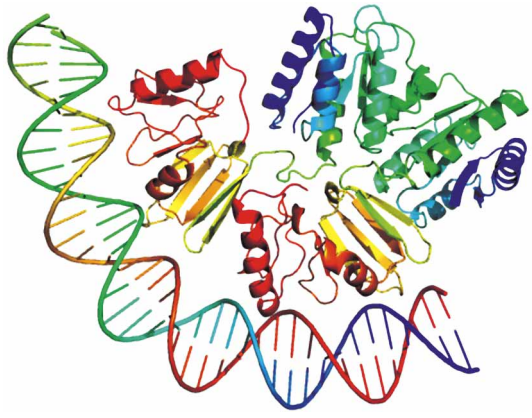


Figure 6. Structure of the ComE–comcde complex determined by small-angle X-ray scattering (code SASDAB7 in the Small Angle Scattering Biological Data Bank, http://www.sasbdb.org/aboutSASBDB/).[97]
The comcde promoter is represented as a DNA double helix. The compact ComE dimer (secondary structure elements of this dimer are shown) binds to the promoter region comcde and induces its bending.

bending of the promoter region comcde and the release of this bending strain *via* the disruption of the dimer may signal the shut-off of the competence state of the bacterium *S. pneumoniae*.

Small-angle neutron scattering has some key features that distinguish it from small-angle X-ray scattering. Thus, thermal neutrons penetrate matter far better than electrons, the scattering depends on the isotopic composition of the sample and its magnetic properties, due to which this technique is a valuable tool for investigating the structures of biological samples.[99] It was noted that a combined structural approach, in which both small-angle X-ray and neutron scattering data are used, resolves remaining ambiguities in the quaternary structure and improves the molecular modelling of the quaternary structure of protein–nucleic acid complexes.[100]

Two reviews[101,102] summarize data on computational tools and mathematical methods in small-angle scattering data analysis for proteins, nucleic acids and protein–nucleic acid complexes, which are used for the molecular modelling of the samples. The improved approach to mathematical methods for processing and analyzing small-angle scattering data for proteins, nucleic acids and protein–nucleic acid complexes, which takes into account the scattering contribution from excess electron density around the surface of the protein–NA complex, was proposed by Ravikumar *et al.*[103] Emerging applications of small-angle scattering techniques in structural biology were considered by Chaudhuri.[104]

The improvement of experimental methods and considerable progress in computer modelling and methods for mathematical processing of experimental data led to a rapid increase in the amount of experimental small-angle scattering data and the growing number of publications. This generates a need for a global repository in order to compile, systematize and store small-angle scattering data. The Small-Angle Scattering Biological Data Bank (SASBDB) designed by analogy with the PDB allows investigators to freely access experimental small-angle scattering data and associated models that are deposited together with the relevant experimental conditions and technical characteristics of instrumentation.[105] As of May 2015, the SASBDB contained 113 structures of proteins, 3 structures of nucleic acids and 8 structures of heterocomplexes (www.sasbdb.org).

III. Biochemical methods of structure and function analysis of protein–nucleic acid complexes

The above-considered physical methods of structure and function analysis of protein–nucleic acid complexes provide valuable information on the structures of a wide range of biopolymers. Meanwhile, the limitations in solubility, amount, size, homogeneity and conformational flexibility of functional complexes of biopolymers, including complex macromolecular assemblies involved in most of cellular processes, promote the development of biochemical methods. In some cases, the probing of the structures of biopolymers by biochemical methods may be the only source of information for compounds that are difficult to study by other methods. The development of chemical approaches enabling one to prepare a wide range of cross-linking reagents (including mono-, bi- and multifunctional reagents, zero-length cross-linking reagents),[106] the application of state-of-the-art methods of sequencing, mass spectrometry and immunoprecipitation, the use of modern instrumentation, including femtosecond lasers, the design of powerful computational methods for the generation of full models give a new impetus to the development of conventional biochemical methods and their successful application, including *in vivo* experiments, for investigation of different structural aspects of cellular processes. It should be emphasized that biochemical methods of structure and function analysis of protein–nucleic acid complexes perfectly complement physical structural approaches, by compensating the lack of high-resolution data and bypassing the limitations, and allow researchers to identify structures that have been remained unknown.

III.1. Cross-linking method

The cross-linking of proteins to nucleic acids is based on an interaction between reactive groups of proteins and nucleic acids giving rise to a covalent bond. The interaction is formed within a specific complex by means of chemical reagents or through exposure to radiation with a specified wavelength. There are the following types of reactive groups: natural functional groups of biomolecules and chemically active groups, which are specially introduced into biomolecules and which have no dramatic effect on specific recognition, binding and functional activity of the biopolymers under study. The structural data obtained from the analysis of a covalently bound complex give an insight into its structural organization, arrangement and architecture of binding sites and catalytic sites, enable the identification of individual groups of biopolymers closely spaced in different stages of the protein–nucleic acid interaction, allow the preservation of relatively unstable complexes and transient states and provide a knowledge of the dynamics of molecular processes. This approach is most useful in studies of multisubunit complexes and supramolecular structures. It allows researchers to locate individual components within multicomponent a complex.

Numerous methods for the covalent attachment of proteins to nucleic acids are currently available. For example, nucleic acid analogues, which contain various reactive groups in heterocyclic bases or a carbohydrate residue, attached to the terminal or internucleotide phosphate group are used for this purpose. The range of reagents available for cross-linking of proteins to nucleic acids, cross-linking techniques and the chemistry of these processes are described in detail in the books.[106,107] The present review is focused on modern aspects of applications and developments of the cross-linking method in structural studies of protein–nucleic acid complexes, including its integration in comprehensive *in vivo* genome-wide studies as a part of interdisciplinary approaches.

Although many classical techniques are available for the efficient cross-linking of proteins to nucleic acids,[106] new original approaches are still emerging. For instance, Heinze *et al.*[108] developed a new cross-linking strategy to attach the MutS protein, which is involved in a mismatch repair system, on DNA. The MutS protein binds to the DNA sequences containing a mismatch and bends it by 45–60° to form a so-called clamp-like structure. The process is accompanied by ATP-induced conformational changes in MutS giving rise to the DNA–MutS–MutL ternary complex. This complex is a key active intermediate that couples mismatch recognition and discrimination between the tem-

plate DNA strand and the newly synthesized DNA strand. The authors developed a strategy for the covalent trapping of MutS on DNA during the search for and recognition of mismatches. This strategy is based on the covalent binding of the cysteine-containing MutS protein generated by mutagenesis to the disulfide group $RSS(CH_2)_3$ [$R = (CH_2)_3OH$] linked to the 3′-terminal phosphate of the DNA ligand. The mutant protein was constructed so that the disulfide group of the modified DNA ligand was in close vicinity of the cysteine residue. This approach allows the stabilization of different transient MutS–DNA complexes and enables their structural and functional analysis. It was shown [108, 109] that, regardless of the way of stabilization of the MutS–DNA complex by means of modification in DNA, which is attached to the terminal phosphate group or at position 2′ of the carbohydrate moiety, the protein retains its activity in the complex and the latter can be used to study subsequent steps of the repair process.

Lavrik and co-workers [110] used photocross-linking to study the interaction of key nucleotide excision repair factors, XPA (xeroderma pigmentosum group A protein) and RPA (replication protein A), with damaged DNA in the steps of primary damage recognition and the formation of the pre-incision complex. In order to prepare covalently bound protein–nucleic acid complexes, 5-iodo-2′-deoxyuridine 5′-monophosphate residues were introduced into damaged DNA. The binding loci of RPA and XPA on the damaged DNA duplex were experimentally found. The authors obtained for the first time direct evidence for the localization of XPA on the 5′-side from the lesion, while RPA binds to the undamaged strand of DNA duplex opposite the lesion. It was shown that the binding loci of XPA and RPA on the damaged DNA duplex are similar to those on the partially unwound damaged DNA duplex. The review by Rechkunova *et al.*[8] summarizes data on the use of photocross-linking to study individual steps of the mechanism of nucleotide excision repair.

Hybrid approaches combining such methods as mass spectrometry, immunoprecipitation and sequencing are successfully used in the structural analysis of protein–nucleic acid complexes. For instance, Urlaub *et al.* [111] described the methodology for the analysis of protein–RNA contacts in purified ribonucleoprotein complexes. This methodology comprises cross-linking of proteins to RNA, hydrolysis of RNA and protein fragments of a covalently bound complex, isolation of the resulting oligonucleotide–peptide conjugate, MALDI (matrix-assisted laser desorption/ionization) mass spectrometric analysis of the isolated conjugate and subsequent sequencing. The utility of this methodology was demonstrated by the analysis of UV cross-linked protein–RNA spliceosomal complexes — multicomponent ribonucleotide–peptide complexes that catalyze splicing.

Lauber *et al.*[112] used cross-linking coupled with high-resolution mass spectrometry to study the localization of the ribosomal protein S1, which is an effector of prokaryotic translation, on mRNA and investigate the dynamics of this process. The experimental results support a previous hypothesis that S1 acts as the mRNA catching arm of the prokaryotic ribosome. It should be noted that attempts to study the S1 protein by X-ray crystallography failed.

Rogacheva *et al.*[113] used a combination of cross-linking and high-resolution mass spectrometry to analyze contacts between DNA phosphate groups and the active site residues of the excision repair enzymes, *E. coli* formamidopyrimidine DNA glycosylase (Fpg) and its eukaryotic analogue — human 8-oxoguanine DNA glycosylase (hOGG1). These enzymes are bifunctional N-glycosylases/AP-lyases that catalyze excision of 8-oxoguanine residues and subsequent cleavage of the sugar–phosphate backbone of DNA.[114] They play a key role in protecting the genetic apparatus of the cell against the mutagenic effect of 8-oxoguanine — one of the most common oxidative DNA lesions. The active sites of the enzymes were probed using a series of reactive substrate analogues containing O-ethyl-substituted pyrophosphate groups, resulting in the identification of specific contacts between amino acids and phosphate groups of DNA within specific enzyme–substrate complexes (Fig. 7).

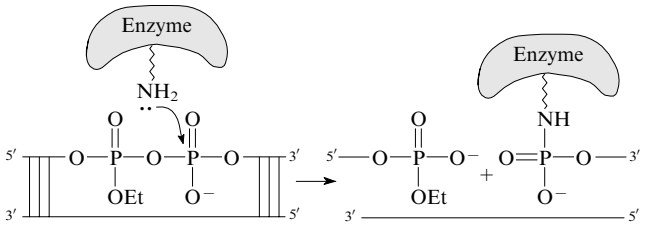


Figure 7. Formation of a covalent bond between an amino acid of a protein and a phosphate group of DNA within a specific enzyme–substrate complex.[113]

The study by means of mass spectrometry using a set of mutant forms revealed that Lys56 Fpg and Lys249 hOGG1 are cross-linked to the phosphate group of DNA located 3′ to the 8-oxoguanine residue at the site catalytically cleaved by both enzymes. On the basis of the analysis of the experimental data, models were constructed for the interaction of Fpg and hOGG1 with phosphate groups of damaged DNA within the specific enzyme–substrate complex in solution.[113] The authors suggested that this approach can be employed to study prokaryotic and eukaryotic homologues of 8-oxoguanine DNA glycosylases, as well as other repair enzymes.

Combined methods of investigation of RNA-binding proteins in complexes with RNA based on cross-linking coupled with mass spectrometry and subsequent sequencing are considered in the review.[4]

Granneman *et al.*[115] studied the architecture of small nucleolar ribonucleoprotein U3 (snoRNP), which plays an essential role in ribosome biogenesis, by means of UV cross-linking of the Nop1, Nop56, Nop58 and Rrp9 proteins involved in snoRNP to small nucleolar RNA. Then RNA fragments cross-linked to the proteins were sequenced using complementary DNA (cDNA) libraries prepared by primer extension assays. The cross-linking was performed both *in vitro* on purified complexes and *in vivo* in living cells. The authors identified the binding sites for each of these proteins on the U3 RNA, which provided an insight into the architecture of the U3 snoRNP. The snoRNP proteins were also cross-linked to U3 pre-rRNA. The preferential association of these proteins at sites of box C/D pre-rRNA was revealed, which confirms that these proteins can be cross-linked to unprocessed RNA ligands. The results of the cited study demonstrate that this method is reliable and applicable to analysis of RNA–protein interactions.

The development of the combined UV cross-linking–immunoprecipitation method can also be employed in structural studies of protein–nucleic acid complexes *in vivo*.[116] The authors mentioned that UV light as a zero-length cross-linking reagent causes less distortion in the structure of the complex under study compared to chemical reagents. The application of a laser as an UV light source, which generates pulses in nano-, pico- or femtosecond ranges, makes it possible to decrease substantially the duration of irradiation and increase the efficiency of cross-linking compared to classical UV lamps. Therefore, this approach is suitable for kinetic studies. A combination of UV cross-linking with immunoprecipitation is successfully used in structural studies of protein–DNA complexes, including investigation of the chromatin structure, localization of DNA-binding proteins on a genome-wide scale, investigation of transcription and DNA replication.[116]

The cross-linking method has been successfully used for a long time for the stabilization of protein–DNA complexes with the aim of their subsequent crystallization. This approach can substantially facilitate the crystal growth. The disulfide bond formation is one of the most common cross-links. This bond is formed between a cysteine residue of the protein introduced by mutagenesis and a thiol or disulfide group of a modified nucleic acid. An example is the crystal structure of human immunodeficiency virus 1 reverse transcriptase with inverse transcriptase in complex with DNA determined in 1998. Prior to X-ray crystallography, the complex was stabilized by engineering a disulfide cross-link between the cysteine side chain of the protein and thiol-containing DNA.[117] It should be emphasized that the covalent complex retained the native conformation and specific activity. Later, this approach was used to determine the crystal structure of DNA glycosylase in complexes with DNA.[118] A research team headed by Professor Chuan He at the University of Chicago [119] implemented stabilization by means of disulfide cross-linking strategy and succeeded in obtaining crystals and solving the three-dimensional structure of the *E. coli* repair protein AlkB and its human homologue ABH2, which are involved in the oxidative demethylation of 1-methyladenine and 3-methylcytosine, in complexes with DNA.

III.2. Chemical and enzymatic footprinting

Chemical and enzymatic footprinting methods are widely used for revealing elements of the spatial organization of biopolymers and their complexes, probing interactions between biomolecules and ligands and analyzing protein–nucleic acid interactions and induced conformational changes. These methods provide the so-called single-nucleotide resolution, which nicely complements the structural information available from X-ray crystallography, NMR spectroscopy or cryo-electron microscopy. The footprinting data also complement the dynamic information obtained by optical methods based on fluorescence resonance energy transfer (FRET).[120] The footprinting method does not require large amounts of the compound and has no technical limitations, such as the preparation of diffraction-quality crystals (for X-ray crystallography) or purified highly soluble samples (for NMR spectroscopy).

The footprinting has been used as a tool for studying biopolymers for a rather long time.[121, 122] However, impressive advances have been made in this area in recent years. Thus, next-generation sequencing (NGS) methods [123] are suitable for parallel sequencing of several genome regions and can generate hundreds of megabase to gigabases of sequencing data in a single instrument run, thus providing new powerful tools for characterizing cellular processes *in vivo*.[124]

Ingolia *et al.*[125] developed a technique for investigation of *in vivo* translations at a single-codon resolution based on a novel methodology for sequencing RNA fragments, which are bound to ribosomes by footprinting. This method produces global snapshots of all the ribosomes active in a cell at a particular moment. It includes the sequencing of cDNA libraries prepared by reverse transcriptase, which correspond to mRNA fragments produced by footprinting. The authors used this approach to obtain a snapshot of the steady-state linear distribution of ribosomes on RNA transcripts in cells of *Saccharomyces cerevisiae*. Later, Ingolia *et al.*[126] developed a protocol for genome-wide monitoring of translation (ribosome profiling) allowing the mapping of exact positions of ribosomes on nuclease-protected mRNA fragments (Fig. 8).[126] Besides, footprints can reveal the exact regions of the transcriptome that are translated, as well as the sites of translation initiation.

Lareau *et al.*[127] implemented the ribosome profiling approach to visualize conformational changes in the ribosome during elongation at a single-codon resolution. This provides new opportunities to study the kinetics of translation and to identify the factors that affect each step in the elongation cycle.

Therefore, the ribosome profiling approach has revolutionized the ability to monitor the internal state of cells: the translational control can play an important role in modulating gene expression. Studies by means of this approach provide new knowledge on the mechanisms of cellular protein synthesis and new structural data on protein–nucleic acid complexes involved in this process.[3]

Kvaratskhelia and Grice [128] used an approach combining footprinting and mass spectrometry. The essence of this approach is that amino acid residues in a free protein and in the protein involved in a specific complex with a nucleic acid are subjected to selective chemical modification followed by mass spectrometric analysis of fragments of modified proteins produced by proteolytic cleavage, resulting in the identification of protein–nucleic acid contacts. To determine the contacts between human immunodeficiency virus type 1 reverse transcriptase and RNA, the authors used biotinylated *N*-hydroxysuccinimide, which is capable of modifying amino groups of lysine residues in both protein subunits. The approach is attractive for analyzing large nucleoprotein complexes, which are difficult to study by X-ray crystallography and NMR.

Zhao *et al.*[129] employed a similar technique to study specific contacts between human immunodeficiency virus type 1 integrase and viral DNA. The N-terminal Lys14 was found to play an important role in viral DNA binding. In addition, a DNA-induced conformational change in integrase was revealed. Based on these data, the authors identified the DNA-binding channel in the active full-length multimeric protein — integrase. This experimental approach can be used for mapping specific protein–NA interactions within functional nucleoprotein complexes.

Silverman *et al.*[130] described a new universal high-throughput approach, which allows the identification of specific RNA-binding sites of numerous individual RNA-binding proteins in eukaryotic transcriptomes in a single

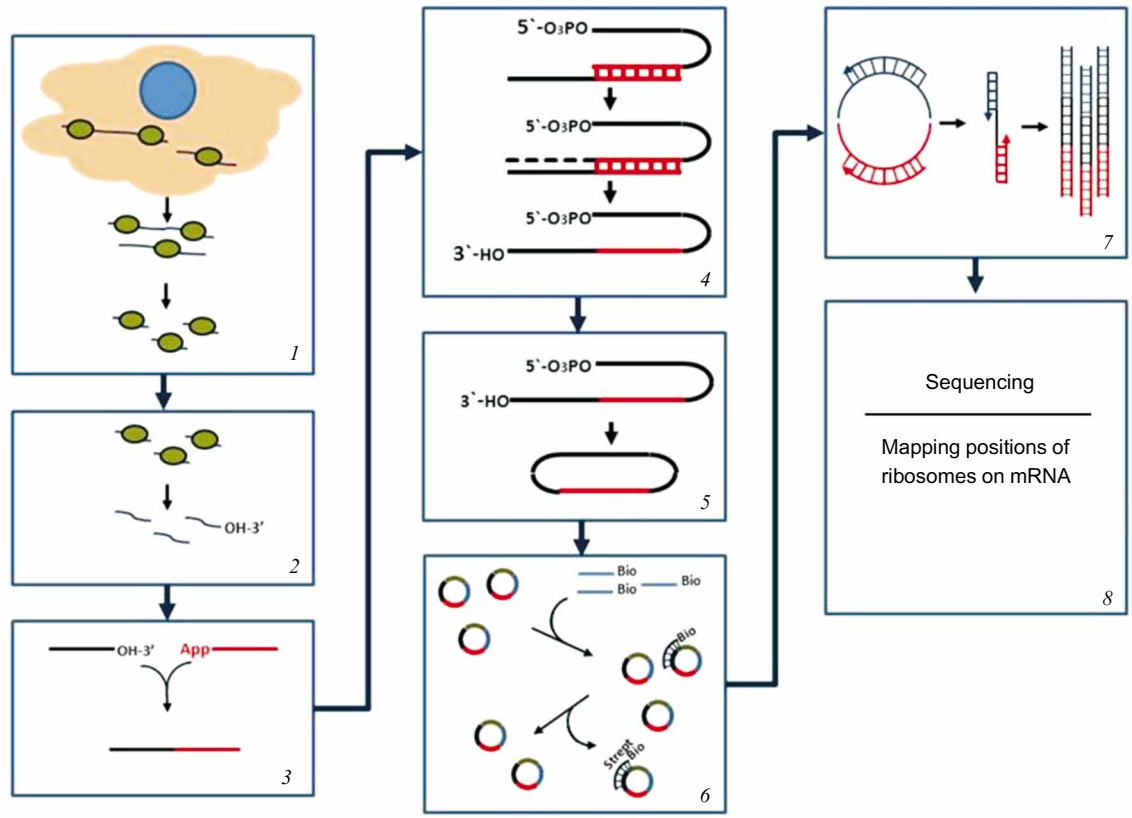


Figure 8. Overview of the ribosome profiling protocol.[126]
(*1*) Ribonuclease treatment of cell lysates in order to digest all mRNAs that are not protected by ribosomes; (*2*) extraction of mRNA fragments bound to ribosomes; (*3*–*8*) steps of sequencing of mRNA fragments using cDNA libraries generated by primer extension assays; (*6*) removal of DNA copies of rRNA fragments (rRNA fragments are produced in step *1* upon ribonuclease treatment of ribosomal RNA) from the sample. Step (*8*) includes the analysis of the results, the localization of individual ribosomes and the mapping of their distribution along mRNA. Bio is a biotin molecule attached to a oligonucleotide probe; Strept/Bio is the streptavidin–biotin complex.

experiment. This approach termed PIP-seq (protein interaction profile sequencing) is based on the high-throughput sequencing of protein-bound RNA fragments, which are produced upon treatment of cells with UV light or formaldehyde followed by nuclease hydrolysis with RNases specific for single- or double-stranded RNA. It allows the identification of specific protein-binding sites within both unprocessed and mature RNAs and gives the full footprint of protein–RNA interaction sites for all samples and organisms.

Singh *et al.*[131] developed an approach to map RNA–protein complexes using immunoprecipitation (RNA: protein immunoprecipitation in tandem) termed RIPiT-seq. This approach includes two sequential purifications of ribonucleoprotein complexes, which are generated by footprinting, by means of immunoprecipitation. It produces highly specific RNA fragments of both individual RNA-binding proteins and multisubunit complexes and has the capacity to differentiate RNA fragments bound to proteins with overlapping protein composition. This approach is suitable for analysis of complexes, in which it is difficult to perform UV cross-linking to produce RNA fragments,[132] as well as of complexes of proteins that bind double-stranded RNA.[133]

Among recent achievements of this method, let us mention the iCLIP (individual-nucleotide resolution UV cross-linking and immunoprecipitation) approach, which combines UV cross-linking, including the application of photoactivated ribonucleoside analogues, with immunoprecipitation and high-throughput sequencing of cDNA libraries generated by primer extension assays.[134, 135] High resolution and specificity of this method are achieved by the intramolecular cDNA circularization step.[136] This approach was employed to study heterogeneous nuclear RNA-binding proteins hnRNP that act in coordination to regulate alternative splicing of human pre-mRNA,[137] including the hnRNP C (see Ref. 134) and hnRNP L proteins,[138] and to elucidate their role in the regulation of splicing. The results demonstrated an unprecedented degree of complexity of interactions between the hnRNP proteins and their targets. Huelga *et al.*[137] suggested that this fact likely confers robustness to cells.

Liu *et al.*[139] investigated the mechanism of biological regulation of specific recognition by RNA-binding proteins of their targets. The authors employed the PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation) method, which includes UV cross-linking using photoactivated ribonucleoside analogues, and the MeRIP immunoprecipitation method. They demonstrated that, in human cells, N6-methyladenosine (m6A), which is the most prevalent modification in eukaryotic mRNA, acts as a switch that controls the

accessibility of binding sites in RNA to RNA-binding proteins by altering the local structure of mRNA and long non-coding RNA (lncRNA). This structural change facilitates the binding of heterogeneous nuclear ribonucleoprotein C (hnRNP C) involved in pre-mRNA processing. The results of the cited study demonstrated the regulatory role of m6A-switches in gene expression and RNA maturation.

Modic *et al.*[11] summarized the results of investigations, in which the CLIP approach combined with other methods was used to study RNA in complexes with RNA-binding proteins, whose mutations cause central nervous system disorders. RNA-Binding proteins play a key role in post-translational regulation of gene expression. The post-translational processing, RNA stability, RNA localization and translation — are only some of the processes that are regulated by multiple RNA-binding proteins present within the cell. It is known that mutations in different RNA-binding proteins can lead to a number of central nervous system diseases, including neurodegenerative disorders that cause progressive neuronal cell death. These disorders include frontotemporal lobar degeneration, amyotrophic lateral sclerosis and Martin–Bell syndrome (fragile X syndrome), which is one of causes of mental retardation. Therefore, the analysis of the structural information on the localization of RNA-binding proteins and the protein-binding sites in RNA sequences is an important step in investigations of molecular mechanisms of the development of this pathology. Advances in molecular biology coupled with achievements in experimental and clinical medicine hold promise for the development, with time, of methods of pathogenetic therapy for treating neurodegenerative disorders, which will change the pessimistic attitude about such patients.

IV. Combined approaches and prospects of their application

The structure and function analysis is a powerful tool for elucidating molecular mechanisms of biological processes underlying biological functions of the body. Recent advances, which have been made due to the development of experimental methods, acquisition of new equipment and modernization of instrumentation, the improvement of methods for analysis and processing of experimental data, enable researchers to perform structural analysis of multicomponent biological structures with ever increasing complexity in the cellular environment. Nevertheless, integrated multidisciplinary approaches are required to overcome limitations of individual methods and take advantages of each method.[140] Although X-ray crystallography remains the method of choice for structural analysis of large complexes, the crystallization of flexible dynamic systems is often a difficult problem and typically does not provide information on the conformational dynamics of the structure in solution. Nuclear magnetic spectroscopy is a well-suited technique to investigate the dynamics of compounds at picosecond to second time scales, as well as to map large systems at single-nucleotide resolution; however, the sensitivity of this method decreases with increasing molecular mass of the object. Small-angle scattering methods provide low-resolution structural information on the dynamics and conformational equilibria in solution and nicely complement the data from high-resolution methods. In high molecular mass systems, where NMR data are sparse, small-angle scattering data provide additional structural information and can differentiate between models derived from NMR data.[100, 141]

The combination of NMR spectroscopy with small-angle scattering methods for the structural analysis of biopolymers allows researchers to obtain more informative and accurate results. Tidow *et al.*[142] used this combined approach to study the structure of the functionally active p53 protein, which acts as a suppressor of human malignant tumours, and the structure of its specific complex with DNA. The quaternary structure of the functionally active p53 protein contains intrinsically disordered regions, which hinders X-ray crystallography. The authors determined the quaternary structure of p53 in solution using a combination of small-angle X-ray scattering and NMR spectroscopy. The mathematical processing of the small-angle scattering data demonstrated that the functionally active p53 protein is a tetramer having an elongated cross-shaped structure with a pair of loosely coupled dimers at the ends, which are accessible for binding to DNA and partner proteins. The monomers of the protein in this open conformation interact with a specific DNA sequence to form a compact complex, the structure of which is consistent with the structure of the p53 monomers in complex with DNA determined earlier by X-ray crystallography and NMR spectroscopy.

Besides, the combination of NMR spectroscopy and small-angle X-ray and neutron scattering methods was used in the studies.[100, 141] For instance, Hennig *et al.*[100] investigated the structure of the specific ternary complex composed of RNA and two RNA recognition protein domains. The authors employed small-angle X-ray scattering data to refine the NMR data and obtained a more detailed view of interactions between macromolecular components of the complex.

An example of a successful application of the combined approach is the study performed by van Nuland and coworkers.[44] The authors used NMR spectroscopy, small-angle X-ray scattering and calorimetry to study the structural and dynamic characteristics of the binding of the MazE antitoxin to the palindromic DNA sequence. Toxin–antitoxin (TA) modules are pairs of genes involved in bacterial regulation of the cell response to environmental stresses. The mazEF module encodes the MazF toxin and its cognate MazE antitoxin. The highly dynamic MazE has the N-terminal DNA-binding domain through which it can negatively regulate its own promoter. The authors determined the structure of MazE in complex with DNA by NMR spectroscopy and proposed a model for the transcriptional regulation. Studies by complementary techniques demonstrated that the long intrinsically disordered C-terminus in MazE, which is essential for MazF neutralization, does not affect the interaction between the antitoxin and its operator. The authors suggested that the C-terminus of MazE plays an important role in the MazF binding, by increasing the MazE affinity for the palindromic operator site.

Nowak *et al.*[143] used a combination of X-ray crystallography and small-angle scattering to model full-length monomeric reverse transcriptase of xenotropic murine leukaemia virus — a close relative of Moloney murine leukaemia virus. Thus, the crystal structure of a fragment of reverse transcriptase of this virus in complex with the RNA/DNA hybrid substrate was determined by X-ray crystallography, while small-angle scattering data were

required for the construction of the full-length model. It was demonstrated that the mobile RNase H domain of monomeric reverse transcriptase of xenotropic murine leukaemia virus becomes ordered in the presence of the substrate, which is a key difference between monomeric and dimeric transcriptases.

The combined structural approach allowed Sattler and co-workers [144] to shed light on the molecular mechanism of regulation of genetic equality between males and females. The so-called dosage compensation mechanism promotes the double gene expression of the single X-chromosome in mammalian (including human) males, while the translation of the male-specific lethal protein msl2 is inhibited in females. This is due to the recruitment of the female-specific sex-lethal protein Sxl to the Unr region of the msl2 RNA template. The structure of the ternary Sxl–Unr–msl2 ribonucleoprotein complex in the fruit fly drosophila *Drosophila melanogaster* was studied by X-ray crystallography, NMR spectroscopy and small-angle X-ray and neutron scattering methods. The authors found novel ternary contacts between components of the complex, as well as non-canonical contacts between RNA and the α1 helix of the RNA recognition motif 1 of the Sxl protein. These results suggest that the inhibition of msl2 translation, necessary for female viability, is triggered by specific cooperative molecular interactions that lock a ribonucleoprotein switch.[144]

Advantages of the combination of the AFM method with spectroscopic, optical, mechanical or electrochemical methods are considered in the review.[93] As mentioned above, the use of semiconductor quantum dots [94] and fluorescent-labelled compounds [95] facilitates the visualization of objects in studies by the AFM method coupled with fluorescence microscopy.

The above-mentioned studies, which include cross-linking as an important step prior to crystallization of the complexes required for X-ray crystallography,[118, 119] are examples of the successful combination of biochemical methods with high-resolution structural methods. This approach allows the stabilization of protein–nucleic acid complexes and is often the only opportunity to grow crystals suitable for X-ray diffraction.

Examples of the successful application of the combined approaches are also considered in other sections of the present review.

Meanwhile, it is not always possible to integrate information obtained by different experimental methods or their combinations in order to build three-dimensional models of biological objects. Webb *et al.*[145] describe a protocol for the integration of all structural data obtained for biological objects using state-of-the-art experimental methods, such as X-ray crystallography, NMR spectroscopy, electron microscopy, small-angle scattering, footprinting, cross-linking and so on. The Integrative Modeling Platform software necessary for the implementation of this protocol is described. This protocol can be used to generate models, which are consistent with all available data sets and, consequently, they are more precise, complete and accurate.

V. Conclusion

The development of methods of structural biology in the past decade resulted in the determination of the structures of more than 75 000 individual proteins, nucleic acids and protein–nucleic acid complexes, with a total of about 100 000 structures over the entire period of experimental studies beginning in the second half of the 1970s. The number of structures is growing very rapidly, in a nearly exponential fashion. This is mostly due to the following advances:

— the development of powerful synchrotron radiation sources and femtosecond lasers, which allow researchers to study nanocrystals of biomolecules and even single copies of large macromolecular complexes bypassing their crystallization;

— the development of experimental approaches and techniques, including the sample preparation and data collection, which makes it possible to measure solid-state NMR spectra and analyze objects with a molecular mass of 100 kDa and higher;

— the progress in electron microscopy and improvement of mathematical methods for image recognition and computational image analysis, which makes it possible to produce electron microscopy images at a resolution comparable to that achieved by X-ray crystallography;

— the use of nanomaterials in structural studies, as well as for the data collection and processing of experimental results, which leads to a multiple increase in the performance of experimental data transmission, processing and storage systems, the creation a of new architectures of high-performance devices having capabilities of computing systems with elements of artificial intelligence that behave like living organism.

New knowledge of the molecular mechanisms of the functioning of living cells, the appearance and development of pathological processes can be derived from structural studies of single macromolecules possessing unique individual properties, because measurements using samples consisting of a large number of molecules provide an averaged pattern.

Meanwhile, there are no individual methods, which can be used to comprehensively characterize a particular object in space and time. Each method has advantages and limitations and provides an insight only into some fragments of the versatile picture created by nature. The use of complementary methods, which can be employed to study the structure and dynamics of complex biological objects, and the combination of physical and biochemical approaches can brought us closer to understanding of the comprehensive and adequate molecular model of the operation of the living cell and allows the development of methods for regulation of cellular processes.

Therefore, the structure and function analysis of protein–nucleic acid complexes provides detailed information on these biological objects with a complex organization and is an important step towards an understanding of their molecular function and role in the evolution and development of human diseases. The results of these studies are widely used in medicine, biotechnology and nutricology. Advances in structural biology coupled with achievements in bioorganic chemistry, experimental and clinical medicine hold promise that new efficient and appropriate methods for the diagnosis and therapy of various diseases will be developed, which will lead to an increase in the life expectancy and improve the quality of life.

We thank V.I.Polshakov, the Leading Researcher at the M.V.Lomonosov Moscow State University for valuable

advice and help in the preparation of this review for publication.

This review has been written with the financial support of the Russian Science Foundation (Project No. 14-24-00061).

References

1. I N Serdyuk *Usp. Biol. Khim.* **42** 3 (2002)
2. C Dominguez, M Schubert, O Duss, S Ravindranathan, F H-T Allain *Prog. Nucl. Magn. Reson. Spectrosc.* **58** 1 (2011)
3. N T Ingolia *Nat. Rev. Genet.* **15** 205 (2014)
4. M Ascano, S Gerstberger, T Tuschl *Curr. Opin. Genet. Dev.* **23** 20 (2013)
5. G M Clore, J Iwahara *Chem. Rev.* **109** 4108 (2009)
6. M Blackledge *Prog. Nucl. Magn. Reson. Spectrosc.* **46** 23 (2005)
7. *Force Microscopy: Applications in Biology and Medicine*. (Eds B P Jena, J K H Hörber) (Hoboken, NJ: Wiley, 2006)
8. N I Rechkunova, Yu S Krasikova, E A Maltseva, O I Lavrik *Biopolym. Cell* **28** 207 (2012)
9. S Campagne, V Gervais, A Milon *J. R. Soc. Interface* **8** 1065 (2011)
10. T Carlomagno *J. Magn. Reson.* **241** 126 (2014)
11. M Modic, J Ule, C R Sibley *Mol. Cell. Neurosci.* **56** 429 (2013)
12. C D Cukier, A Ramos *Eur. Biophys. J.* **40** 1317 (2011)
13. H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, P E Bourne *Nucleic Acids Res.* **28** 235 (2000)
14. J R Williamson *Cell* **139** 1041 (2009)
15. C Bury, E F Garman, H M Ginn, R B G Ravelli, I Carmichael, G Kneale, J E McGeehan *J. Synchrotron Radiat.* **22** 213 (2015)
16. M V Koval'chuk, V O Popov *Nauka v Rossii* (3) 4 (2013)
17. C M Ogata *Nat. Struct. Mol. Biol.* **5** 638 (1998)
18. R K McGinty, R C Henrici, S Tan *Nature (London)* **514** 591 (2014)
19. M A Schumacher, N Chinnam, B Cuthbert, N K Tonthat, T Whitfill *Genes Dev.* **29** 451 (2015)
20. J Balzarini, K Das, J A Bernatchez, S E Martinez, M Ngure, S Keane, A Ford, N Maguire, N Mullins, J John, Y Kim, W Dehaen, J V Voorde, S Liekens, L Naesens, M Götte, A R Maguire, E Arnold *Proc. Natl. Acad. Sci. USA* **112** 3475 (2015)
21. T C Appleby, J K Perry, E Murakami, O Barauskas, J Feng, A Cho, D Fox III, D R Wetmore, M E McGrath, A S Ray, M J Sofia, S Swaminathan, T E Edwards *Science* **347** 771 (2015)
22. Y Nomura, S Sugiyama, T Sakamoto, S Miyakawa, H Adachi, K Takano, S Murakami, T Inoue, Y Mori, Y Nakamura, H Matsumura *Nucleic Acids Res.* **38** 7822 (2010)
23. R N A Martin, J E McGeehann, N J Ball, S D Streeter, S-J Thresh, G G Kneale *Acta Crystallogr., Sect. F* **69** 962 (2013)
24. M B Shevtsov, S D Streeter, S-J Thresh, A Swiderska, J E McGeehan, G G Kneale *Acta Crystallogr., Sect. D* **71** 398 (2015)
25. A D Gelinas, D R Davies, T E Edwards, J C Rohloff, J D Carter, C Zhang, S Gupta, Y Ishikawa, M Hirota, Y Nakaishi, T C Jarvis, N Janjic *J. Biol. Chem.* **289** 8720 (2014)
26. Y-M Chang, C-H Ho, C K-M Chen, M Maestre-Reyna, M W Chang-Chien, A H-J Wang *Nucleic Acids Res.* **42** 5314 (2014)
27. D Lu, S Fillet, C Meng, Y Alguel1, P Kloppsteck, J Bergeron, T Krell, M-T Gallegos, J Ramos, X Zhang *Genes Dev.* **24** 2556 (2010)
28. P-H Kuo, C-H Chiang, Y-T Wang, L G Doudeva, H S Yuan *Nucleic Acids Res.* **42** 4712 (2014)
29. J Lin, M G Gagnon, D Bulkley, T A Steitz *Cell* **160** 219 (2015)
30. H N Chapman, P Fromme, A Barty, T A White, R A Kirian, A Aquila, M S Hunter, J Schulz, D P Deponte, U Weierstall, R B Doak, F R N C Maia, A V Martin, I Schlichting, L Lomb, N Coppola, R L Shoeman, S W Epp, R Hartmann, D Rolles, A Rudenko, L Foucar, N Kimmel, G Weidenspointner, P Holl, M Liang, M Barthelmess, C Caleman, S Boutet, M J Bogan, J Kryzwinski, C Bostedt, S Bajt, L Gumprecht, B Rudek, B Erk, C Schmidt, A Hömke, C Reich, D Pietschner, L Strüder, G Hauser, H Gorke, J Ullrich, S Herrmann, G Schaller, F Schopper, H Soltau, K-U Kühnel, M Messerschmidt, J D Bozek, S P Hau-Riege, M Frank, C Y Hampton, R G Sierra, D Starodub, G J Williams, J Hajdu, N Timneanu, M M Seibert, J Andreasson, A Rocker, O Jönsson, M Svenda, S Stern, K Nass, R Andritschke, C-D Schröter, F Krasniqi, M Bott, K E Schmidt, X Wang, I Grotjohann, J M Holton, T R M Barends, R Neutze, S Marchesini, R Fromme, S Schorb, D Rupp, M Adolph, T Gorkhover, T Andersson, H Hirsemann, G Potdevin, H Graafsma, B Nilsson, J C H Spence *Nature (London)* **470** 73 (2011)
31. M M Seibert, T Ekeberg, F R N C Maia, M Svenda, J Andreasson, O Jönsson, D Odić, B Iwan, A Rocker, D Westphal, M Hantke, D P DePonte, A Barty, J Schulz, L Gumprecht, N Coppola, A Aquila, M Liang, T A White, A Martin, C Caleman, S Stern, C Abergel, V Seltzer, J-M Claverie, C Bostedt, J D Bozek, S Boutet, A A Miahnahri, M Messerschmidt, J Krzywinski, G Williams, K O Hodgson, M J Bogan, C Y Hampton, R G Sierra, D Starodub, I Andersson, S Bajt, M Barthelmess, J C H Spence, P Fromme, U Weierstall, R Kirian, M Hunter, R B Doak, S Marchesini, S P Hau-Riege, M Frank, R L Shoeman, L Lomb, S W Epp, R Hartmann, D Rolles, A Rudenko, C Schmidt, L Foucar, N Kimmel, P Holl, B Rudek, B Erk, A Hömke, C Reich, D Pietschner, G Weidenspointner, L Strüder, G Hauser, H Gorke, J Ullrich, I Schlichting, S Herrmann, G Schaller, F Schopper, H Soltau, K-U Kühnel, R Andritschke, C-D Schröter, F Krasniqi, M Bott, S Schorb, D Rupp, M Adolph, T Gorkhover, H Hirsemann, G Potdevin, H Graafsma, B Nilsson, H N Chapman, J Hajdu *Nature (London)* **470** 78 (2011)
32. W A Barletta, J Bisognano, J N Corlett, P Emma, Z Huang, K-J Kim, R Lindberg, J B Murphy, G R Neil, D C Nguyen, C Pelligrini, R A Rimmer, F Sannibale, G Stupakov, R P Walker, A A Zholents *Nucl. Instrum. Methods Phys. Res., Sect. A* **618** 69 (2010)
33. J C H Spence, U Weierstall, H N Chapman *Rep. Prog. Phys.* **75** 102601 (2012)
34. D O Sinitsyn, V Yu Lunin, A N Grum-Grzhimailo, E V Gryzlova, N K Balabaev, N L Lunina, T E Petrova, K B Tereshkina, E G Abdulnasyrov, A S Stepanov, Yu F Krupyanskii *Russ. J. Phys. Chem. B* **8** 457 (2014) [*Khim. Fiz.* **33** (7) 21 (2014)]
35. D D Boehr, R Nussinov, P E Wright *Nat. Chem. Biol.* **5** 789 (2009)
36. D Pakotiprapha, D Jeruzalmi *Encyclopedia of Life Sciences* (Wiley, 2009); DOI: 10.1002/9780470015902.a0002720.pub3
37. K Pröpper, K Meindl, M Sammito, B Dittrich, G M Sheldrick, E Pohl, I Usón *Acta Crystallogr., Sect. D* **70** 1743 (2014)
38. C A Theimer, N L Smith, M Khanna *Methods Mol. Biol.* **831** 197 (2012)
39. M D Sam, R T Clubb *Methods Mol. Biol.* **831** 219 (2012)
40. R Kaptein *J. Biomol. NMR* **56** 1 (2013)

41. I N Serdyuk, N Zaccai, J Zaccai *Methods in Molecular Biophysics. Structure, Dynamics, Function* (Cambridge: Cambridge University Press, 2007)
42. Z Cieniková, F F Damberger, J Hall, F H-T Allain, C Maris *J. Am. Chem. Soc.* **136** 14536 (2014)
43. B Heddi, V V Cheong, H Martadinata, A T Phan *Proc. Natl. Acad. Sci. USA* **112** 9608 (2015)
44. V Zorzini, L Buts, E Schrank, Y G J Sterckx, M Respondek, H Engelberg-Kulka, R Loris, K Zangger, N A J van Nuland *Nucleic Acids Res.* **43** 1241 (2015)
45. E de Alba, N Tjandra *Prog. Nucl. Magn. Reson. Spectrosc.* **40** 175 (2002)
46. R S Lipsitz, N Tjandra *Annu. Rev. Biophys. Biomol. Struct.* **33** 387 (2004)
47. D Fushman, R Varadan, M Assfalg, O Walker *Prog. Nucl. Magn. Reson. Spectrosc.* **44** 189 (2004)
48. J L Ortega-Roldan, M R Jensen, B Brutscher, A I Azuaga, M Blackledge, N A J van Nuland *Nucleic Acids Res.* **37** e70 (2009)
49. D C Williams Jr, M Cai, G M Clore *J. Biol. Chem.* **279** 1449 (2004)
50. R Stefl, F C Oberstrass, J L Hood, M Jourdan, M Zimmermann, L Skrisovska, C Maris, L Peng, C Hofr, R B Emeson, F H-T Allain *Cell* **143** 225 (2010)
51. C R Cantor, P R Shimmel *Biophysical Chemistry, Part 2: Techniques for the Study of Biological Structure and Function* (New York: Freeman and Co., 1980)
52. V K Voronov, I A Ushakov *Russ. Chem. Rev.* **79** 835 (2010)
53. U Brath, S I Swamy, A X Veiga, C-C Tung, F Van Petegem, M Erdélyi *J. Am. Chem. Soc.* **137** 11391 (2015)
54. W M Liu, M Overhand, M Ubbink *Coord. Chem. Rev.* **273–274** 2 (2014)
55. J Iwahara, G M Clore *Nature (London)* **440** 1227 (2006)
56. G M Clore, C Tang, J Iwahara *Curr. Opin. Struct. Biol.* **17** 603 (2007)
57. G M Clore *J. Biomol. NMR* **51** 209 (2011)
58. C P Jaroniec *Solid State Nucl. Magn. Reson.* **43–44** 1 (2012)
59. S Cai, L Zhu, Z Zhang, Y Chen *Biochemistry* **46** 4943 (2007)
60. C D Mackereth, T Madl, S Bonnal, B Simon, K Zanier, A Gasch, V Rybin, J Valcárcel, M Sattler *Nature (London)* **475** 408 (2011)
61. H Günther *NMR Spectroscopy. An Introduction* (New York: Wiley, 1980)
62. A Marchanka, B Simon, G Althoff-Ospelt, T Carlomagno *Nat. Commun.* **6** Art. No. 7024 (2015)
63. S Asami, M Rakwalska-Bange, T Carlomagno, B Reif *Angew. Chem., Int. Ed.* **52** 2345 (2013)
64. C Fernández, G Wider *Curr. Opin. Struct. Biol.* **13** 570 (2003)
65. R Rosenzweig, L E Kay *Annu. Rev. Biochem.* **83** 291 (2014)
66. J Fiaux, E B Bertelsen, A L Horwich, K Wüthrich *Nature (London)* **418** 207 (2002)
67. D Sahu, G M Clore, J Iwahara *J. Am. Chem. Soc.* **129** 13232 (2007)
68. N Sugitani, S M Shell, S E Soss, W J Chazin *J. Am. Chem. Soc.* **136** 10830 (2014)
69. J C States, E R McDuffie, S P Myrand, M McDowell, J E Cleaver *Hum. Mutat.* **12** 103 (1998)
70. N M Walavalkar, J M Cramer, W A Buchwald, J N Scarsdale, D C Williams Jr *Nucleic Acids Res.* **42** 11218 (2014)
71. L M I Koharudin, R Boelens, R Kaptein, A M Gronenborn *J. Biomol. NMR* **56** 31 (2013)
72. Yu S Nagornov, I S Yasnikov, M N Tyur'kov *Sposoby Issledovaniya Poverkhnosti Metodami Atomno-silovoi i Elektronnoi Mikroskopii* (Methods of Investigation of Surface by Atomic Force and Electron Microscopy) (Tol'yati: Tol'yatti State University, 2012)
73. B J Greber, D Boehringer, V Godinic-Mikulcic, A Crnkovic, M Ibba, I Weygand-Durasevic, N Ban *J. Mol. Biol.* **418** 145 (2012)
74. B J Greber, D Boehringer, A Leitner, P Bieri, F Voigts-Hoffmann, J P Erzberger, M Leibundgut, R Aebersold, N Ban *Nature (London)* **505** 515 (2014)
75. A Amunts, A Brown, X-c Bai, J L Llácer, T Hussain, P Emsley, F Long, G Murshudov, S H W Scheres, V Ramakrishnan *Science* **343** 1485 (2014)
76. A Amunts, A Brown, J Toots, S H W Scheres, V Ramakrishnan *Science* **348** 95 (2015)
77. M Maletta, I Orlov, P Rodlin, Y Beck, D Moras, I M L Billas, B P Klaholz *Nat. Commun.* **5** Art. No. 4139 (2014)
78. N Fischer, P Neumann, A L Konevega, L V Bock, R Ficner, M V Rodnina, H Stark *Nature (London)* **520** 567 (2015)
79. J K H Hörber, in *Force Microscopy: Applications in Biology and Medicine* (Eds B P Jena, J K H Hörber) (Hodoken, NJ: Wiley, 2006) p. 300
80. I Tessmer, P Kaur, J Lin, H Wang *J. Nanobiotechnol.* **11** 25 (2013)
81. Yu L Lyubchenko, L S Shlyakhtenko *Methods* **47** 206 (2009)
82. P J M Murphy, M Shannon, J Goertz *J. Vis. Exp.* (53) e3061 (2011)
83. M Murakami, I Narumi, K Satoh, A Furukawa, I Hayata *Biochim. Biophys. Acta* **1764** 20 (2006)
84. I Sorel, O Piétrement, L Hamon, S Baconnais, E Le Cam, D Pastré *Biochemistry* **45** 14675 (2006)
85. Y Wang, W Liao, J Cai *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi. (J. Biomed. Eng.)* **24** 1172 (2007)
86. Y Wang, S Ran, G Yang *Sci. Rep.* **4** Art. No. 5897 (2014); DOI: 10.1038/srep05897
87. L Hamon, D Pastré, P Dupaigne, C Le Breton, E Le Cam, O Piétrement *Nucleic Acids Res.* **35** e58 (2007)
88. A V Machulin, E I Deryusheva, A K Yunusova, L A Zheleznaya, I N Serdyuk *Biophysics* **57** 314 (2012) [*Biofizika* **57** 432 (2012)]
89. Yu L Lyubchenko *Micron* **42** 196 (2011)
90. D Ristic, H Sanchez, C Wyman *Methods Mol. Biol.* **783** 213 (2011)
91. L S Shlyakhtenko, A A Gall, Yu L Lyubchenko *Methods Mol. Biol.* **931** 295 (2013)
92. Yu L Lyubchenko, A A Gall, L S Shlyakhtenko *Methods Mol. Biol.* **1117** 367 (2014)
93. S Moreno Flores, J L Toca-Herrera *Nanoscale* **1** 40 (2009)
94. Y Ebenstein, N Gassman, S Kim, S Weiss *J. Mol. Recognit.* **22** 397 (2009)
95. H Sanchez, R Kanaar, C Wyman *Ultramicroscopy* **110** 844 (2010)
96. D I Svergun, L A Feigin *Rentgenovskoe i Neitronnoe Malouglovoe Rasseyanie* (Small-Angle X-Ray and Neutron Scattering) (Moscow: Nauka, 1986)
97. D Sanchez, M Boudes, H van Tilbeurgh, D Durand, S Quevillon-Cheruel *FEBS J.* **282** 1538 (2015)
98. V D Gruzina *Antibiotiki i Khimioterapiya* **48** (10) 32 (2003)
99. V I Gordelii, A I Kuklin *Preprint OIYaI R14-2002-250* (Dubna, 2002)
100. J Hennig, I Wang, M Sonntag, F Gabel, M Sattler *J. Biomol. NMR* **56** 17 (2013)
101. R P Rambo, J A Tainer *Curr. Opin. Struct. Biol.* **20** 128 (2010)
102. D Schneidman-Duhovny, S J Kim, A Sali *BMC Struct. Biol.* **12** 17 (2012)
103. K M Ravikumar, W Huang, S Yang *J. Chem. Phys.* **138** 024112 (2013)
104. B N Chaudhuri *Protein Sci.* **24** 267 (2015)
105. E Valentini, A G Kikhney, G Previtali, C M Jeffries, D I Svergun *Nucleic Acids Res.* **43** D357 (2015)
106. S S Wong, D M Jameson, *Chemistry of Protein and Nucleic Acid Cross-Linking and Conjugation* (2nd Ed.) [New York: Taylor and Francis (CRC Press), 2012]

107. G T Hermanson *Bioconjugate Techniques* (3rd Ed.) (San Diego, CA: Academic Press, 2013)
108. R J Heinze, S Sekerina, I Winkler, C Biertümpfel, T S Oretskaya, E Kubareva, P Friedhoff *Mol. BioSyst.* **8** 1861 (2012)
109. M Monakhova, A Ryazanova, A Hentschel, M Viryasov, T Oretskaya, P Friedhoff, E Kubareva *J. Chromatogr., A* **1389** 19 (2015)
110. Yu S Krasikova, N I Rechkunova, E A Maltseva, I O Petruseva, O I Lavrik *Nucleic Acids Res.* **38** 8083 (2010)
111. H Urlaub, E Kühn-Hölsken, R Lührmann *Methods Mol. Biol.* **488** 221 (2008)
112. M A Lauber, J Rappsilber, J P Reilly *Mol. Cell. Proteomics* **11** 1965 (2012)
113. M Rogacheva, A Ischenko, M Saparbaev, S Kuznetsova, V Ogryzko *J. Biol. Chem.* **281** 32353 (2006)
114. M V Rogacheva, S A Kuznetsova *Russ. Chem. Rev.* **77** 765 (2008)
115. S Granneman, G Kudla, E Petfalski, D Tollervey *Proc. Natl. Acad. Sci. USA* **106** 9613 (2009)
116. L Zhang, K Zhang, R Prändl, F Schlffl *Biochem. Biophys. Res. Commun.* **322** 705 (2004)
117. H Huang, R Chopra, G L Verdine, S C Harrison *Science* **282** 1669 (1998)
118. C He, G L Verdine *Chem. Biol.* **9** 1297 (2002)
119. C-G Yang, C Yi, E M Duguid, C T Sullivan, X Jian, P A Rice, C He *Nature (London)* **452** 961 (2008)
120. D Fourmy, S Yoshizawa *WIREs RNA* **3** 557 (2012)
121. *Footprinting of Nucleic Acid-Protein Complexes* (Ed. A Revzin) (San Diego, CA: Academic Press, 1993)
122. S S Jain, T D Tullius *Nat. Protoc.* **3** 1092 (2008)
123. M L Metzker *Nat. Rev. Genet.* **11** 31 (2010)
124. D R Morris *Genome Biol.* **10** 215 (2009)
125. N T Ingolia, S Ghaemmaghami, J R S Newman, J S Weissman *Science* **324** 218 (2009)
126. N T Ingolia, G A Brar, S Rouskin, A M McGeachy, J S Weissman *Nat. Protoc.* **7** 1534 (2012)
127. L F Lareau, D H Hite, G J Hogan, P O Brown *eLife* **3** e01257 (2014)
128. M Kvaratskhelia, S F J Le Grice *Methods Mol. Biol.* **488** 213 (2008)
129. Z Zhao, C J McKee, J J Kessl, W L Santos, J E Daigle, A Engelman, G Verdine, M Kvaratskhelia *J. Biol. Chem.* **283** 5632 (2008)
130. I M Silverman, F Li, A Alexander, L Goff, C Trapnell, J L Rinn, B D Gregory *Genome Biol.* **15** R3 (2014)
131. G Singh, E P Ricci, M J Moore *Methods* **65** 320 (2014)
132. Y W Yang, R A Flynn, Y Chen, K Qu, B Wan, K C Wang, M Lei, H Y Chang *eLife* **3** e02046 (2014)
133. E P Ricci, A Kucukural, C Cenik, B C Mercier, G Singh, E E Heyer, A Ashar-Patel, L Peng, M J Moore *Nat. Struct. Mol. Biol.* **21** 26 (2014)
134. J Konig, K Zarnack, G Rot, T Curk, M Kayikci, B Zupan, D J Turner, N M Luscombe, J Ule *J. Vis. Exp.* (50) 2638 (2011)
135. M J Moore, C Zhang, E C Gantman, A Mele, J C Darnell, R B Darnell *Nat. Protoc.* **9** 263 (2014)
136. I Huppertz, J Attig, A D'Ambrogio, L E Easton, C R Sibley, Y Sugimoto, M Tajnik, J König, J Ule *Methods* **65** 274 (2014)
137. S C Huelga, A Q Vu, J D Arnold, T Y Liang, P P Liu, B Y Yan, J P Donohue, L Shiue, S Hoon, S Brenner, M Ares Jr, G W Yeo *Cell Rep.* **1** 167 (2012)
138. O Rossbach, L-H Hung, E Khrameeva, S Schreiner, J König, T Curk, B Zupan, J Ule, M S Gelfand, A Bindereif *RNA Biol.* **11** 146 (2014)
139. N Liu, Q Dai, G Zheng, C He, M Parisien, T Pan *Nature (London)* **518** 560 (2015)
140. J Hennig, M Sattler *Protein Sci.* **23** 669 (2014)
141. A Lapinaite, B Simon, L Skjaerven, M Rakwalska-Bange, F Gabel, T Carlomagno *Nature (London)* **502** 519 (2013)
142. H Tidow, R Melero, E Mylonas, S M V Freund, J G Grossmann, J M Carazo, D I Svergun, M Valle, A R Fersht *Proc. Natl. Acad. Sci. USA* **104** 12324 (2007)
143. E Nowak, W Potrzebowski, P V Konarev, J W Rausch, M K Bona, D I Svergun, J M Bujnicki, S F J Le Grice, M Nowotny *Nucleic Acids Res.* **41** 3874 (2013)
144. J Hennig, C Militti, G M Popowicz, I Wang, M Sonntag, A Geerlof, F Gabel, F Gebauer, M Sattler *Nature (London)* **515** 287 (2014)
145. B Webb, K Lasker, J Velázquez-Muriel, D Schneidman-Duhovny, R Pellarin, M Bonomi, C Greenberg, B Raveh, E Tjioe, D Russel, A Sali *Methods Mol. Biol.* **1091** 277 (2014)