PORTLAND PRESS

Review Article

Computational modeling of RNA 3D structure based on experimental data

Almudena Ponce-Salvatierra[1,*], Astha[1,*], Katarzyna Merdas[1,*], Chandran Nithin[1], Pritha Ghosh[1], Sunandan Mukherjee[1] and Janusz M. Bujnicki[1,2]

[1]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, ul. Ks. Trojdena 4, Warsaw PL-02-109, Poland; [2]Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, Poznan PL-61-614, Poland

Correspondence: Janusz M. Bujnicki (iamb@genesilico.pl)

OPEN ACCESS

RNA molecules are master regulators of cells. They are involved in a variety of molecular processes: they transmit genetic information, sense cellular signals and communicate responses, and even catalyze chemical reactions. As in the case of proteins, RNA function is dictated by its structure and by its ability to adopt different conformations, which in turn is encoded in the sequence. Experimental determination of high-resolution RNA structures is both laborious and difficult, and therefore the majority of known RNAs remain structurally uncharacterized. To address this problem, predictive computational methods were developed based on the accumulated knowledge of RNA structures determined so far, the physical basis of the RNA folding, and taking into account evolutionary considerations, such as conservation of functionally important motifs. However, all theoretical methods suffer from various limitations, and they are generally unable to accurately predict structures for RNA sequences longer than 100-nt residues unless aided by additional experimental data. In this article, we review experimental methods that can generate data usable by computational methods, as well as computational approaches for RNA structure prediction that can utilize data from experimental analyses. We outline methods and data types that can be potentially useful for RNA 3D structure modeling but are not commonly used by the existing software, suggesting directions for future development.

Introduction

RNA molecules play fundamental roles in many biological processes, such as regulation of gene expression [1–4], RNA splicing [5–7], and protein synthesis [8–10]. The understanding of these processes improves as new structures of RNA molecules and their complexes with various molecular partners are determined, and the molecular details of interactions analyzed [11–17].

As of 20 September 2018, 1345 RNA structures (and 2284 RNA complexes with other molecules) were available in the Protein Data Bank (PDB) [18], including 807 (1680) solved by X-ray crystallography, 504 (120) by NMR spectroscopy, 29 (478) by cryo-EM, and 5 (6) by other methods. However, the experimental determination of RNA structures at high resolution, which allows the description of individual atomic features as well as atomic interactions, is time-consuming, and in many cases not possible. As a result, complete high-resolution structures are available for an extremely small fraction of all the known RNA molecules that are crucial for numerous fundamental cellular processes; for instance, only 87 out of 2791 RNA families in the Rfam database have at least one experimentally determined structure.

It will be very challenging to obtain a comprehensive overview of structures for all RNA molecules that have a biochemical function (beyond protein-coding). Such efforts have been undertaken toward elucidating protein structures, e.g., through the Protein Structural Genomics initiative and related efforts

around the world [19,20]. The flexible nature of biomacromolecules has proven to be a challenge to overcome when it comes to sample preparation for biophysical techniques that have a scope for their structural determination [21–23]. For instance, in crystallography, it is unusual to have a sample that readily forms well-diffracting crystals. Even the production of the sample *per se* is difficult, and biomacromolecules (proteins and nucleic acids likewise) often require some extent of engineering or chemical modification so that their structure can be analyzed with a given method [24–26]. In any of the methods used for structure determination of biomacromolecules the homogeneity of the sample is also not to be taken for granted and it poses additional difficulties as the number of different molecules involved in the formation of the macromolecular complex increases [23,27]. Furthermore, RNA molecules are known to be intrinsically dynamic, more than protein molecules, so even if a high-resolution structure of an RNA molecule is obtained, it may not represent the full spectrum of native solution conformations [28–30]. Thus, studies on RNA sequence–structure–function relationships present major challenges for macromolecular structural biology, requiring alternate approaches to gain structural information of macromolecular complexes.

Given the scarcity of experimentally determined high-resolution structures of RNA molecules, theoretical models of RNA 3D structure can assist in understanding and guiding the identification of functionally important regions and can provide a starting point for higher resolution descriptions [31]. However, 'purely theoretical' structural predictions usually suffer from limited accuracy, and moreover, the deviation between the theoretical model and the 'true' (unknown) structure is very difficult to assess in the absence of additional data. Fortunately, even though complete high-resolution structures of RNA molecules are scarce, often a wide range of heterogeneous information from biochemical and biophysical data is available. There exists various approaches for enzymatic or chemical probing of RNA secondary structure [32,33] and ways to obtain information about the overall shape of the molecule [34]. This low-resolution information can be supplemented with partial high-resolution data, e.g., structures determined for fragments of the system under study. Computational techniques can be used to integrate the existing data, guide the structure elucidation, and subsequently determine the mechanisms of action and interactions between the functional elements of the molecule [35–37]. This approach is commonly called a hybrid or integrative modeling.

This review provides necessary information about the use of experimental data that, in combination with computational approaches for the determination of RNA 3D structures, are commonly applied to study structural properties of RNA molecules. We review this issue from two complementary perspectives: experimental methods providing data that can be used in the RNA 3D structure modeling process, and computational methods that can use the experimental data. It is worth emphasizing that various experimental approaches discussed in this article have various interesting applications beyond the generation of data for macromolecular modeling: such applications are out of the scope of this review and are not discussed here. We focus primarily on the state-of-the-art and contemporary techniques while avoiding giving historical overviews. Likewise, in the description of computational methods, we concentrate on the utilization of information derived from experimental data, while 'purely theoretical' modeling approaches and historical perspectives are not covered. The discussion describes approaches combining experimental data (both low and high-resolution) on RNA molecules as well as their complexes, with computational modeling to obtain high-resolution models. In summary, we provide a methodological workflow to build structural models of RNA 3D structures for cases, where the use of traditional methods (X-ray crystallography, NMR, or cryo-EM) may encounter difficulties and where the application of purely theoretical modeling is unlikely to be successful.

A related important area of research, not covered by this article, is the use of computational modeling methods, in particular various types of simulations, to make predictions about various physicochemical properties of RNA molecules, which can be compared with results of experiments. The reader interested in the use of RNA modeling methods to predict functional properties of RNA molecules beyond the 3D structure is referred to recent publications on this topic, e.g., [38–40].

Experimental methods that generate data useful for RNA 3D structure modeling

Our knowledge of RNA 2D and 3D structures is primarily based on various experimental observations. On the one hand, experiments can be interpreted in terms of specific structural knowledge about a particular RNA molecule, although large sets of experimental data collected systematically for various RNAs can be used to infer principles that apply to RNA structure in general. This review focusses on the use of experimental data to infer individual RNA structures.

Methods that generate 3D structural information in the form of shapes

A number of experimental methods can be used to determine the overall 3D shapes of biological macromolecules. Here we review the main types of experimental approaches applicable to structural studies of RNAs alone or in complex with other molecules. We focus on the type of data generated by these methods, which can be further processed to obtain spatial restraints for building 3D structural models of RNA molecules.

Macromolecular X-ray crystallography

Macromolecular X-ray crystallography (MX) is the most widely used technique of structural biology because it provides highly accurate models of large macromolecular complexes [41]. In a crystallographic experiment, the crystal is exposed to an X-ray beam, and single diffraction images are collected on an area detector while rotating the crystal, which is mounted on a goniostat. The structure factors are vectorial magnitudes consisting of a module and a phase and correspond to the Fourier transform of the electron density distribution within the unit cell of the crystal. The electron density map can be calculated with the application of an inverse Fourier transform to the structure factors. While the modules can be directly obtained from the diffraction intensities measured during the experiment, the information about the phases is not recorded. The phase information has to be therefore obtained independently [42]. In a best-case scenario, a structure of a molecule related to the one under study has already been determined, and it can be used as a model for molecular replacement (MR). Roughly, the structure of the known molecule needs to be transferred into the unit cell of the unknown molecule, where rotations and translations are applied to it until a solution with the best fit between experimental diffraction data and calculated diffraction data from the model is found [43]. Alternatively, phases can be determined by Single/Multiple Isomorphous Replacement (SIR/MIR) or Single/Multiple Anomalous Dispersion (SAD/MAD) [44]. A combination of the aforementioned methods is also possible as Single/Multiple Isomorphous Replacement with Anomalous Signal (SIRAS/MIRAS) or MR with Single Anomalous Dispersion (MR-SAD). In any case, the obtained electron density map is then used to determine the 3D location of the atoms. The procedure of electron density map calculation and model building can be iterated.

One of the most significant bottlenecks of MX is the necessity to obtain well-diffracting crystals of the sample of interest. The resolution, at which crystal structures can be determined, relies highly on the size of the crystal and on how well the molecules are ordered in it. It is very difficult to obtain such crystals for RNA molecules due to their overall negatively charged surface, that often leads to non-specific interactions and causes a poor degree of internal order within the crystal [45–47].

Hydrogen atoms exhibit very low electron density; hence MX experiments cannot define their positions even when the structures are determined at very high resolution. The identification of H atoms can be addressed by neutron diffraction experiments. Although it is in principle the method of choice for the experimental determination of H atoms positions, neutron crystallography is still very challenging, e.g., the technique requires relatively large crystals, which are usually challenging to grow and there are very few facilities worldwide where the experiments can take place [48]. Despite to date no neutron structures of RNA have been determined, one potential application of this technique could be the assessment of tautomeric states of the nucleobases, which are known to play diverse roles in RNA biochemistry [49]. Currently, the positions of hydrogen atoms are rarely determined experimentally in macromolecular studies, and computational approaches are combined with chemical knowledge to predict their positions.

Small angle scattering

Small angle scattering (SAS) is the collective name given to various techniques, including **X-ray (SAXS)** and **neutron (SANS) scattering**, employed to characterize biological macromolecules, including RNAs. In each of these techniques, radiation is elastically scattered by a sample in solution, where molecules are randomly oriented [50]. The scattering intensity is integrated at each angle (typically $0.1–10°$) to produce a 1D pattern as a function of the momentum transfer parameter [51]. In SAXS, the scattering describes the distribution of electron density that interacts with X-rays, while in SANS, it describes the distribution of the nuclei of atoms that interact with the neutrons [50,52,53]. Neutrons have two advantages over X-rays: first, they cause less molecular degradation due to radiative heating, and second, they discriminate between hydrogen and deuterium, which have opposite signs and substantial difference between the scattering lengths, which means that at certain measurable deuterium to hydrogen ratios, selected components of a complex molecular system can be made invisible [54]. In both SAXS and SANS, the scattering pattern can be analyzed to provide information about the averaged particle sizes and shapes. Depending on the angular range in which a clear scattering signal can be recorded, these techniques are capable of delivering structural information of dimensions between 1 and 100 nm [55]. The effect of averaging many randomly oriented molecules is equivalent to averaging all directions of one molecule, hence for example chirality cannot be determined [56,57]. Therefore

SAXS methods require combination with other experimental approaches and/or with computational modeling to discriminate between the true structure and its mirror image.

Electron microscopy

Electron microscopy (EM) is a technique used for visualization of biomacromolecules, which relies on the contrast formed by the interaction of incoming electrons with the specimen. In single-particle EM, multiple 2D projection images of randomly oriented molecules are recorded, which can be then computationally combined into a 3D density map. It should be emphasized that electron micrographs are real-space images containing both amplitude and phase information. Therefore, EM techniques do not have the 'phase problem' that occurs in X-ray crystallography.

In negative stain EM, the sample is immobilized on the surface of a carbon film and covered with a thin layer of stain that yields high contrast images. Even though the stain limits the level of molecular detail of the reconstructed envelopes, negative stain EM has been very useful in the study of RNA assemblies, in particular when higher resolution crystal and NMR structures can be fit in the maps [58].

In cryo-EM, the sample is embedded in a thin layer of amorphous ice, which avoids the artifacts of staining. The high-energy electrons used for imaging cause, however, severe radiation damage of the individual molecules, which requires imaging with low doses of electrons. These imaging conditions yield lower contrast images, requiring the analysis of a very large number of particle images to obtain a satisfactory signal-to-noise ratio. Once a sufficient number of 2D images containing high-resolution information are captured, they are aligned and averaged into 2D classes. Particles from these classes are used to produce a 3D reconstruction. The 3D reconstruction is a Coulomb potential density map that can be interpreted similarly to electron density maps determined by X-ray crystallography. At very high resolution, cryo-EM maps can be used to determine the position of individual atoms. Currently, it is hard to say where the limits of cryo-EM are, as structures of proteins as small as hemoglobin (64 kDa) [59] have been determined, and resolutions as high as 1.8 Å have been achieved [60].

There are several examples of structures of large RNA–protein assemblies solved by cryo-EM, such as the yeast mitochondrial large ribosomal subunit [61] and the spliceosome [62]; while smaller structures, such as the Group II intron in complex with its maturase, were solved to 3.8 Å resolution [63]. Even though small RNA structures have not been determined yet with this technique alone, a combination of cryo-EM, NMR, and molecular dynamics (MD) have successfully yielded the structure of the HIV-1 RNA dimerization signal, which is only 30 kDa [64]. In comparison with other methods, cryo-EM has as a key advantage that the minimal amount of sample is required, there is no need of producing crystals and that during image classification it is possible to sort structural heterogeneity. Many 3D EM maps are however of relatively low resolution and require additional information to enable building atomic models.

Atomic force microscopy

Atomic force microscopy (AFM) is a type of scanning probe microscopy (SPM), which can be used for various applications, including imaging and force-probing biological samples. Unlike optical or electron microscopes that record images, AFM instruments collect data to generate images by physically touching the samples. In AFM, a molecule is immobilized on a solid substrate, such as mica or glass. Subsequently, the surface is raster scanned with sub-nanometer precision by an ultrasharp tip, mounted at the end of a force-sensing cantilever. It can be bent by repulsive and attractive interactions between the tip and the sample. The lever deflection is measured, enabling the detection of height changes with a nanometer spatial resolution, leading to a 3D topography image of the surface [65]. The tapping mode AFM, in which the cantilever is oscillated close to the sample surface without actually 'touching' it, found widespread application in structural biology, and in particular it has been successfully applied to image fragile biological objects, including RNA molecules [66]. Structure determination requires computational processing of AFM images. In the case of complex RNA molecules, one of the goals is to identify skeleton-like structures, from which information can be extracted about the presence of double- and single-stranded regions, their connectivity and mutual orientations [67]. Literature presents some examples of AFM application for structural characterization of RNA, including the architectural organization of the Hepatitis C virus genomic RNA [68] or the structure of the HIV-1 Rev response element (RRE) on its own and in complex with virion regulator (Rev) [69].

Methods that generate 3D structural information in the form of spatial interactions

Another class of methods for 3D structure analysis studies interactions within (or between) molecules rather than their global shapes. These techniques offer a different point of view on the molecular structure, providing spatial restraints from which 3D molecular models can be obtained. We review the main types of experimental approaches applicable to structural studies of RNA and its complexes.

Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a technique used to observe local magnetic fields around atomic nuclei and measure magnetic interactions between them. The principle behind NMR is that some nuclei isotopes (e.g., ^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{31}P) carry magnetic dipoles that can take up various orientations characterized by different energies when the magnetic field is applied. In response to the pulsed magnetic wave at specific (resonance) frequencies, these nuclei can absorb and re-emit electromagnetic radiation, and the obtained NMR signals are the consequence of the occurring energy state transitions [70]. The measured frequencies depend mainly on the composition of the nucleus and its chemical surrounding (i.e., covalent bonds and proximal atoms), which introduces small changes in the recorded frequencies, called chemical shifts. Magnetic interactions between nuclei connected with covalent bonds result in spin-spin coupling, while the correlation of nuclei through-space interactions can be detected by applying the nuclear Overhauser effect (NOE) [71,72]. The latter yields information about intermolecular distances of nearby nuclei and the molecular motion. Long-range structural information can be extracted from residual dipolar couplings (RDCs), which are caused by the presence of an aligning medium that interferes with the isotropic tumbling of a molecule and induces a certain degree of alignment of the molecule to the magnetic field [72]. Another important parameter is the rate at which an ensemble of nuclear spins return to the equilibrium state after the radiofrequency pulse is applied, which provides data on the atomic mobility [73].

1D NMR can also generate useful information about base pairing patterns. The imino proton resonances of G and U can only be observed when these are protected from proton exchange with the solvent. The latter happens if these protons are engaged in base pairing through hydrogen bonding interactions [74]. Counting the imino protons resonances allow the determination of the number of base pairs present in the molecule under study. 1D spectra of imino protons can inform about the type of base pairing depending on the region where the chemical shifts appear (12–15 ppm for canonical Watson–Crick base pairs, and upfield in the case of non-canonical base pairing). A combination of 1D and 2D experiments are central to the RNA secondary experimental structure determination [75,76].

After assigning recorded resonance frequencies to individual atoms in the analyzed sample, this information can be used to infer structural restraints. If the number and quality of the restraints are sufficient to define the conformations and interactions of essentially all residues in the RNA, they can be used to model its 3D structure. Traditionally it has been accomplished by energy minimization based on MD simulations that attempt to satisfy physical constraints as well as NMR-based restraints (or at least to minimize their violation) [77]. Typically, a simulated annealing protocol is used, followed by refinement, and the resulting lowest energy structures are selected to be compared with each other to check for convergence and estimate local uncertainty.

Solution NMR is an effective tool for RNA structure elucidation since its advantage relies on the collection of dynamic information of flexible macromolecules at atomistic or residue resolution. However, NMR probing of biomolecules has been limited by their size (usually <50 kDa) and so it is most successfully applied for determining the structure of small RNA domains, aptamers, and ribozymes [78–80]. NMR structure determination for larger molecules has been meeting a problem of broadening and decay of the NMR signal, making the NMR-based structural studies of large RNAs very challenging. The largest RNA molecule solved by NMR so far is the HIV-1 core packaging signal, which is 155 nucleotides long [80]. Besides, sample preparation for many types of NMR experiments requires isotope labeling of RNA molecules.

Recently developed solid-state NMR (ssNMR) is the spectroscopic method applicable to macromolecules of any size in a non-crystalline form and may become a solution to the obstacles in RNA structure determination. Although ssNMR has been rarely used in nucleic acids study and still requires more advancement and methodology preparation, recent reports suggest that it holds great promise for the study of RNAs and RNPs [81–84].

Electron paramagnetic resonance spectroscopy

Electron paramagnetic resonance (EPR) is a spectroscopic technique used to study chemical species with unpaired electrons. The principles behind EPR are analogous to those of NMR, with the difference in electron spins being excited instead of the spins of atomic nuclei [85]. Radicals are very reactive, and so do not normally occur in stable forms in nucleic acids or proteins. Therefore, biomacromolecules or their parts usually have to be labeled (chemically modified) with special non-reactive radical reagents, such as nitroxide, used as spin-labels, whose chemical environment can be then inferred from the EPR spectra [86]. Another approach to RNA structure determination is based on binding paramagnetic metal ions to the RNA, which enables to probe metal-ion binding sites. Although NMR and EPR share some similarities, the properties of the electron spin utilized in the second method allows for obtaining constraints of greater distances. Continuous-wave EPR measurements are based on the dipole–dipole coupling and provide the distance restraints of approximately 20 Å, while relaxation enhancement—up to 30–40 Å. Finally, pulsed electron double resonance (PELDOR) spectroscopy, also referred to as double electron-electron resonance (DEER),

enables to measure long-range distances (∼80 Å) between nitroxides and it can be used to reveal conformational changes in nucleic acid molecules [86,87]. Moreover, spin labels function as probes sensitive to the nucleotide dynamics. Hence the EPR spectroscopy can be successfully applied for studying RNA interactions with various ligands. EPR-based methods have been used to study structural and dynamic features of RNA, to explore different modes of RNA–ligand interaction, to obtain long-range structural restraints, and to probe metal-ion binding sites [88,89], also in combination with NMR [90].

Förster resonance energy transfer

Förster resonance energy transfer (FRET) is a technique that measures non-radiative energy transfer between two light-sensitive molecules (chromophores). In typical applications relevant to structural biology, both the donor and the acceptor chromophores are fluorescent, and the term 'fluorescence resonance energy transfer' is often used instead. Intrinsic fluorescence of RNA is relatively low, thus for FRET analysis, the potentially interacting molecules or their parts have to be labeled at specific positions with donor and the acceptor fluorophores. The energy transfer occurs by means of intermolecular long-range dipole–dipole coupling, which is distance dependent [91]. FRET is highly efficient when the emission and absorbance spectra of the fluorophores overlap and if they are positioned within the Förster radius from each other (the distance at which half the excitation energy of the donor is transferred to the acceptor, typically 3–6 nm) [92]. The FRET efficiency can be determined from the acceptor emission intensity, from the photobleaching rates of the donor in the presence and absence of an acceptor, or from the change in the fluorescence lifetime of the donor. Spatial measurements provided by the technique rely upon the spectral properties of fluorophores incorporated to studied molecule and range between 10 and 80 Å [93]. Thus, FRET can be used to identify molecular proximity at distances relevant to macromolecular structures, making it a useful tool to quantitate molecular interactions and conformational changes. In particular, single-molecule FRET (smFRET) has become a technique widely applied in studying intra- and intermolecular interactions of biomacromolecules, especially RNA [91]. The method allows resolving individual dynamics and subpopulations, which is especially useful in the investigation of dynamic processes such as folding or systems existing in multiple conformations, e.g., riboswitches [94]. The advantage over ensemble methods is that instead of averaging the results, smFRET provides information about conformations of individual molecules.

Chemical cross-linking

Chemical cross-linking (XL) is a technique for analyzing macromolecular interactions, which has been widely used in studies of proteins and nucleic acids. It involves the formation of chemical bonds between proximal elements of the macromolecule analyzed. One variant of this approach uses photo-activatable reactive chemical groups incorporated into the molecule before the experiment. For studies on RNA structure and interactions, photosensitive nucleotides have been used, such as 4-thiouridine (4SU). Upon irradiation with far-UV (<300 nm) or near-UV (320–365 nm) light, photo-activated compounds generate zero-length cross-links between residues that interact with each other or at least form transient contacts in the folded RNA structure. An alternative is to employ bifunctional cross-linking reagents, such as cisplatin (CPT), nitrogen mustard (NM), chlorambucil (CHB), cyanuric acid (sTT), bikethoxal (BKT), or phenyl-diglyoxal (PDG) that can react with specific nucleophilic groups found on the surface of protein and nucleic acid substrates. These reagents bridge across chemical groups present in the native molecule, if they are positioned within the cross-linking range specified by the size of the cross-linker [95]. As the cross-linking reaction is carried out in solution where molecular structures and interactions are dynamic, XL can provide a snapshot from an ensemble of various interactions and conformations that exist simultaneously. Intramolecular cross-links provide information on spatial proximities within the molecule analyzed, while intermolecular cross-links provide information on contacts between the interacting partners. Information from XL can be used as distance restraints between proximal residues.

The sites of cross-linking can be identified by alkaline hydrolysis, primer extension, PAGE, or RNA sequencing techniques [96]. Cell-based approaches, such as Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP), allow for the identification of the cross-linking sites with a single-nucleotide resolution. 4SU incorporation, which is responsible for the cross-link formation, can be mapped by sequencing RNA molecules converted into a cDNA library [97]. Because of the intrinsic character of the chemical probing, in which the formation of a stable covalent modification leaves a unique mass signature, XL has also been coupled with mass spectrometry analysis (MS). The XL-MS approach allows for detecting the cross-linked moieties (oligonucleotides, peptides, and individual residues resulting from the hydrolytic or gas-phase fragmentation of the probed material)

that were chemically bound to each other. Conjugates analyzed by XL-MS with bifunctional cross-linkers can be analyzed with Links and MS2Links [98], e.g., through the ms3d.org portal [99]. These programs can handle RNA interactions, as well as DNA, proteins, and any of their hetero-conjugate combinations. XL-MS has been applied for studying RNA–protein interactions, in particular for identifying RNA-binding sites in RNA-binding proteins [100,101]. The application of bifunctional cross-linking reagents has been tested for their ability to provide vital spatial constraints on well-characterized RNAs [95]. See also 'RNA 3D structure modeling approaches that use local structural features and pairwise distance restraints (including contact information)' section for applications in modeling.

Ligation-based techniques

Ligation-based techniques exemplified by RNA proximity ligation (RPL) are a family of methods for studying interactions within or between RNA molecules by reconnecting the ribonucleotide backbone. Typically, RNA molecules under study are initially fragmented to introduce breaks in the RNA chain and to generate chemically reactive 5′ and 3′ termini. Then, chemical ligation of spatially proximal termini of RNA fragments is induced, followed by sequencing to identify ligation junctions in chimeric reads. Some variants of this approach include PARIS [102], LIGR-seq [102], and SPLASH [103], add an inter- or intramolecular cross-linking step to hold the interacting fragments together, and these methods present different strategies of fragmenting and enriching cross-linked RNA molecules. Structural information obtained from ligation-based techniques is similar to that from XL-based methods.

Probing methods that generate structural information in the form of local features

Enzymatic or chemical structure probing methods are often used in footprinting experiments, to obtain information about local structural features of ribonucleotide residues, such as solvent accessibility, paired/unpaired state, or rigidity/flexibility. Such 1D features do not indicate which ribonucleotide residues are mutually interacting, and hence they cannot be used to determine the secondary or tertiary structure. However, they can be used to inform computational methods for higher order structure prediction, e.g., to find the best match between experimental results and possible patterns of interactions, especially when correlations between structural features of different residues are considered.

Probing techniques involve enzymatic cleavage and/or chemical modification of the RNA, followed by the detection of modified residues and/or sites of cleavage via reverse-transcribing the modified RNA to cDNA. The cDNA synthesis starts with a primer (bound to a site within the RNA molecule or to a special adapter sequence added at the 3′ end) and the reverse transcriptase halts at cleavage sites as well as may halt at modifications, leading to cDNA truncations (RT-stop). Some techniques employ the ability of the RT to misincorporate non-complementary residues for the modification sites (RT-map). Sequenced cDNA reads are computationally aligned to map the positions affected by probing and to calculate average reactivities for all residues in the RNA under study. Reactivities can be then used to represent the relative strength or prevalence of particular structural features, e.g., as restraints or constraints in various types of computational structure prediction. Primer extension by the reverse transcriptase is problematic for small RNA molecules that may not allow for the stable annealing of short primers and for RNA molecules with very stable secondary structures that may cause the RT to stop regardless of any chemical modifications. In these cases, the implementation of MS analysis allows one to dispense with the cDNA synthesis and to directly characterize the chemical adducts that exhibit a characteristic change in molecular mass [104]. The most common types of probing methods are described in more detail below.

Enzymatic probes

Enzymatic probes are nucleases that discriminate between paired and unpaired regions [105]. Nucleases such as S1 and P1 and RNase T2 cleave phosphodiester bonds in unpaired regions of RNA (ssRNAs), in a sequence-independent manner. RNases such as T1, U2, and A are sequence-dependent, and they cleave ssRNAs only after specific nucleotides. Conversely, RNases such as V1 preferentially cleaves phosphodiester bonds in double-stranded regions of RNA (dsRNA). The use of data from ssRNA and/or dsRNA-specific cleavage allows measuring the degree to which each ribonucleotide residue is in a single- or double-stranded conformation, which in turn can be applied to obtain information about the secondary structure [106,107]. It should be emphasized that RNases are relatively bulky reagents and they can cleave RNA only in sites that are exposed to the solvent and provide sufficient space for protein binding to the nucleic acid substrate.

Chemical probes

Chemical probes may be used as footprinting reagents to interrogate various structural features, including base pairing, local nucleotide dynamics, and solvent accessibility. Chemical reagents typically used for RNA structure

probing can be divided into ones that modify bases and are sequence-specific, and those that act on backbone in a sequence-nonspecific manner.

Base-specific probes that modify the Watson–Crick (WC) edge of specific ribonucleotides include dimethyl sulphate (DMS), 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT), and 3-ethoxy-a-ketobutyraldehyde (kethoxal). They can be used to discriminate between bases that have the WC edge involved in pairing (and hence protected from modification) from those that have the WC edge exposed and are therefore more reactive. It should be emphasized that having the WC edge involved in pairing does not necessarily indicate the involvement of a given residue in secondary structure, as the WC edge can be used for the formation of various non-canonical interactions with bases, backbone, and with ligands bound to RNA.

Light-activated structural examination of RNA (LASER) technique is used to obtain solvent accessibility information of purine nucleobases in RNA molecules [108]. This technique involves the use of UV light to activate nicotinoyl azide (NAz) molecules into reactive electrophilic aroyl nitrenium ions, which form adducts with solvent-exposed purine nucleobases. This technique can be applied *in vitro* to monitor RNA structural changes, e.g., resulting from ligand binding, and it can also be applied *in vivo* for studying RNA structures and interactions.

Hydroxyl radical footprinting (HRF) technique employs sequence-nonspecific cleavage of the nucleic acid backbone in the presence of hydroxyl radicals that are generated by decomposition of hydrogen peroxide into a hydroxyl radical and a hydroxide ion in the presence of Fe(II)-EDTA [109,110]. Regions of the molecule that are buried within the structure or protected by interactions with other molecules are less sensitive to the hydroxyl cleavage than those exposed to the solvent. The effect of the cleavage is quantitated at each nucleotide position, and the resulting reactivity profile can be interpreted in the structural context. Fe(II)-EDTA can be dissolved in solution, providing a comprehensive overview of solvent-accessible compared with solvent-inaccessible residues in the molecule under study. Alternatively, it can be chemically tethered to a specific position to act as a local probe to cleave residues within the range of the tether, often approximately 25 A. The distance between a cleavage site and the probe is proportional to the intensity of the cleavage, which may provide pairwise distance restraints for modeling. Multiplexed hydroxyl radical (•OH) Cleavage Analysis (later renamed MOHCA-gel) utilizes the random incorporation of Fe(II)-EDTA-tagged nucleotide analogs followed by cleavage pattern readout via 2D gel electrophoresis [111].

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) [112,113] technique uses a sequence-unbiased reagent, e.g., NMIA [113], 1M7 [113,114], 1M6 [115], BzCN [116], NAI [117], NAI-N$_3$ [118], and FAI [117], to form an adduct with the sugar 2′-hydroxyl group of flexible or disordered nucleotides, and discriminates them from residues that are rigid, e.g., those base-paired in helices [109,110,112,113]. There is essentially no relationship between SHAPE reactivities and solvent accessibility, as many ribonucleotide residues with unconstrained 2′-hydroxyl groups that are fully buried, react readily with SHAPE reagents. SHAPE thus probes the local nucleotide flexibility of RNA nucleotides regardless of solvent accessibility [119].

High-throughput probing

High-throughput probing involves the application of structure-selective reagents to many RNA molecules at the same time, coupled with high-throughput sequencing. This method allows for the investigation of very large RNA molecules, different mixtures of RNAs (even whole transcriptomes), for analyzing different conformational states of RNAs (e.g., in cotranscriptional folding or in riboswitches), and for detecting correlated changes at different positions in the sequence, which may be indicative of long-range interactions. There are many varieties of high-throughput RNA probing techniques developed to probe RNA structure *in vitro* and/or *in vivo*. In particular, a number of methods were developed that use RT-stop as reactivity detection strategy, which differ in the probing reagents and in the priming strategy (defined or random primer) used to prepare a library for high-throughput sequencing. High-throughput methods for RNA enzymatic probing include PARS with RNase S1 vs RNase V1 [120], FragSeq with RNase P1 [121], PARTE with RNase V1 across different temperatures [122], and dsRNA-seq with RNase One [123]. HRF-seq is a parallel sequencing-based method for probing RNA accessibility with hydroxyl radicals [124]. MOHCA-seq is an adaptation of the MOHCA-gel technique, in which the detection of radical cleavage sites is detected by high-throughput sequencing [125]. Also, many high-throughput variations of the chemical probing technique were developed, which utilize RT-stop and/or RT-map approaches. These techniques have been recently reviewed in detail [126].

Mutational profiling (MaP) is a high-throughput chemical probing approach that employs detection of multiple chemical modification positions within a single RNA by the RT-map method. The advantage of the MaP method over other sequencing-based probing methods is that a single RNA molecule with the misincorporated nucleotide is analyzed, which eliminates the multiple ligation steps required during sequencing. The sequencing readout is then processed to obtain information about the reactivity of each nucleotide, which in turn can be used to infer structural

features. In SHAPE-MaP, the reactivities are used primarily for discriminating between paired and unpaired regions, and they can also be used for identifying competing and alternative RNA conformations or quantitate any process or function that modulates local RNA dynamics [127,128]. This technique can also be used for probing RNA *in vivo* with 1M7 as a probing reagent [129,130]. RNA interaction groups by MaP (RING-MaP) uses DMS as a modifying reagent and can be used to obtain information about through-space interaction networks that define the tertiary structure and govern sampling of multiple conformations [131].

Mutate-and-map (M^2) is a hybrid method, which combines mutational analysis with chemical mapping to detect interacting residues. It relies on a principle that sequence perturbations at specific sites may release interacting partners that are distant in sequence, leading to detectable changes in their modification by chemical reagents. Thus, the RNA molecule is systematically mutated at each position to generate a series of variants, whose structures are then analyzed by chemical probing. The correlations between mutation sites and the associated spatially distant perturbations in probing patterns are then used to infer pairwise interactions in particular canonical base pairs [132].

Comparison and combination of experimental techniques

Experimental techniques presented above provide different types of structural information, and their application poses distinct challenges. Many RNAs have proven recalcitrant to structure determination by individual methods, in particular, those aiming at capturing the complete structure at atomic resolution. However, a combination of various experimental and computational methods called a hybrid approach has proven to be a successful way of determining structures for even the most 'difficult' RNA targets, overcoming the weaknesses of methods applied separately.

MX and NMR are methods of choice to determine RNA 3D structures at the atomic resolution. Structures based on crystallographic data are often merely snapshots of the molecule that in the native conditions adopts multiple conformations, in contrast with those obtained through NMR provide an ensemble of conformers of the structure of the molecule under study. The shortages of MX and NMR, which include the size limitation of the molecule or complex under study, can be tuned by applying other techniques such as cryo-EM, SAS, and AFM that provide low-resolution information concerning the global shape of the studied targets in the native conditions.

SAS methods deliver low-resolution tertiary structure information of a molecule under physiological conditions, spanning also disordered or flexible regions, as well as facilitate the analysis of molecules' dynamics, i.e., conformational changes, complex formation, folding, movement of flexible domains, and structural alterations occurring in response to changes exerted by external conditions. SAXS and SANS data processed alone can be only used to build coarse-grained structural models that represent the averaged distribution of atoms in the molecule. SAS methods can also be used in combination with other techniques [133]. One of the recent examples is the use of SAXS together with the microfluidic technology as well as with ensemble modeling with MD to provide a landscape of RNA conformations [57].

Constraints regarding the medium or long-range interactions, or protein binding sites, can be measured by chemical probing and footprinting methods combined with MS, EPR, or FRET, which significantly improve the data interpretation. Moreover, scattering measurements are not limited by the size of the molecule of interest and they supplement the methods such as solution state NMR or cryo-EM, generally restricted to small and large targets, respectively. A combination of methods revealing the overall shape with techniques focussing on local structural information allows for the structure determination of molecules exceeding the limitations of a single approach, exemplified by solving the structure of the 30 kDa HIV-1 RNA Dimerization Signal by combining cryo-EM, NMR, and MD [64].

Table 1 summarizes the main features of experimental methods applicable to RNA structure determination. Figure 1 illustrates schematically the various stages of data acquisition and processing toward the determination of RNA structures by selected commonly used experimental approaches.

RNA 3D structure modeling methods informed by experimental data

Purely theoretical prediction of RNA structure, which disregards all experimental information, is currently not practical. While it is known that all biophysical processes, including RNA structure formation, ultimately depend on the quantum chemistry, the enormous mathematical complexity of computational quantum mechanics methods limits their application to systems up to a few hundreds of atoms [134,135]. However, RNA molecules alone are often composed of hundreds or thousands of atoms, and the simulation must also take into account solvent molecules. In general, all contemporary RNA structure prediction methods that are capable of making biologically useful predictions rely on experimental information.

Table 1 Experimental methods commonly used for RNA structure determination

Experimental technique	Data type		Main disadvantages and bottlenecks	Number of 3D structures in the PDB		Resolution [Å] of 3D structures in the PDB		RNA length (nt) in 3D structures in the PDB	
	Original	Interpreted		RNA	RNP	RNA	RNP	RNA	RNP
MX	Diffraction patterns	Electron density maps	● Obtaining well-diffracting crystals, especially for flexible or disordered molecules ● Phasing in the absence of a related structure	807	1680	0.61–11.5	1.14–11.5	4–2880	2–3396
SAXS/SANS	Electron/neutron scattering patterns	Low-resolution envelope	● Aggregation at high concentration ● Heterogeneity of the sample	NA	NA	NA	NA	NA	NA
EM (including cryo-EM)	2D projections of sample	3D density map	● Challenges of working with a liquid sample	29	478	4.5–18.8	2.5–34	12–2904	2–5070
Solution NMR	Free induction decay signal	NMR spectrum/atomic interactions	● Aggregation ● Size of the target biomolecule	503	120	NA	NA	5–155	4–101
ssNMR	Free induction decay signal	NMR spectrum/atomic interactions	● Broad line widths and resonance overlap	1	0	NA	NA	26	NA
SHAPE/ Probing methods	Stops/mutations in cDNA reversely transcribed from modified RNA	Reactivity profile	● Reproducibility ● Data interpretation in the structural context	NA	NA	NA	NA	NA	NA
FRET	Dampening of energy	Distances between interacting chromophores	● Reduced signal-to-noise ratio associated with acquiring the complete spectrum	1	0	NA	NA	NA	49
Fiber Diffraction	Diffraction patterns	Electron density maps	● Loss of structural information due to cylindrical averaging of diffraction data	0	5	NA	2.9–3.5	NA	3 - 3
Combination of experimental techniques			● Combining data obtained by different techniques	4	1	NA	NA	47–192	58
Total				1345	2284				

Thermodynamic experiments such as the UV optical melting and differential scanning calorimetric measurements were used to determine the folding stability of short RNA strands and to infer parameters for secondary structure formation, which are now commonly used in computational methods (review: [136]). For the development and testing of computational methods for RNA 3D structure modeling, datasets of experimentally determined macromolecular 3D structures are fundamental. In particular, datasets of high-resolution structures determined by MX and NMR, available in the PDB, were used to derive statistical parameters that describe the structural and dynamical features of RNA molecules. These parameters describe the frequency of different molecular features and can be encoded in the form of scoring functions used in computational folding simulations, e.g., to estimate the energy of different conformations, or to assess the models according to their similarity to structures observed previously. For instance, statistical information on torsion angles and sugar puckers can be used as prior knowledge in structure prediction algorithms that rely on conditional probabilities [137,138].

RNA 3D structure modeling approaches that use atomic coordinates of known structures as the primary source of experimental information

The experimentally determined macromolecular structures have been widely used as a source of 'spare parts' in the process of model building, from structures and conformations of chemical groups comprising several atoms, to fragments comprising several residues, to local 3D motifs, and to whole molecules. The most successful programs available for molecular modelers often combine multiple approaches, however for the purpose of this review we define the key modeling approaches separately from the actual implementations.

Comparative modeling (also called **homology modeling** or **template-based modeling**) technique is based on the observation that evolutionarily related (homologous) molecules adopt very similar structures, despite divergent mutations accumulated in their sequences [139]. Thus, a 'template' molecule can be used as a structural framework, in
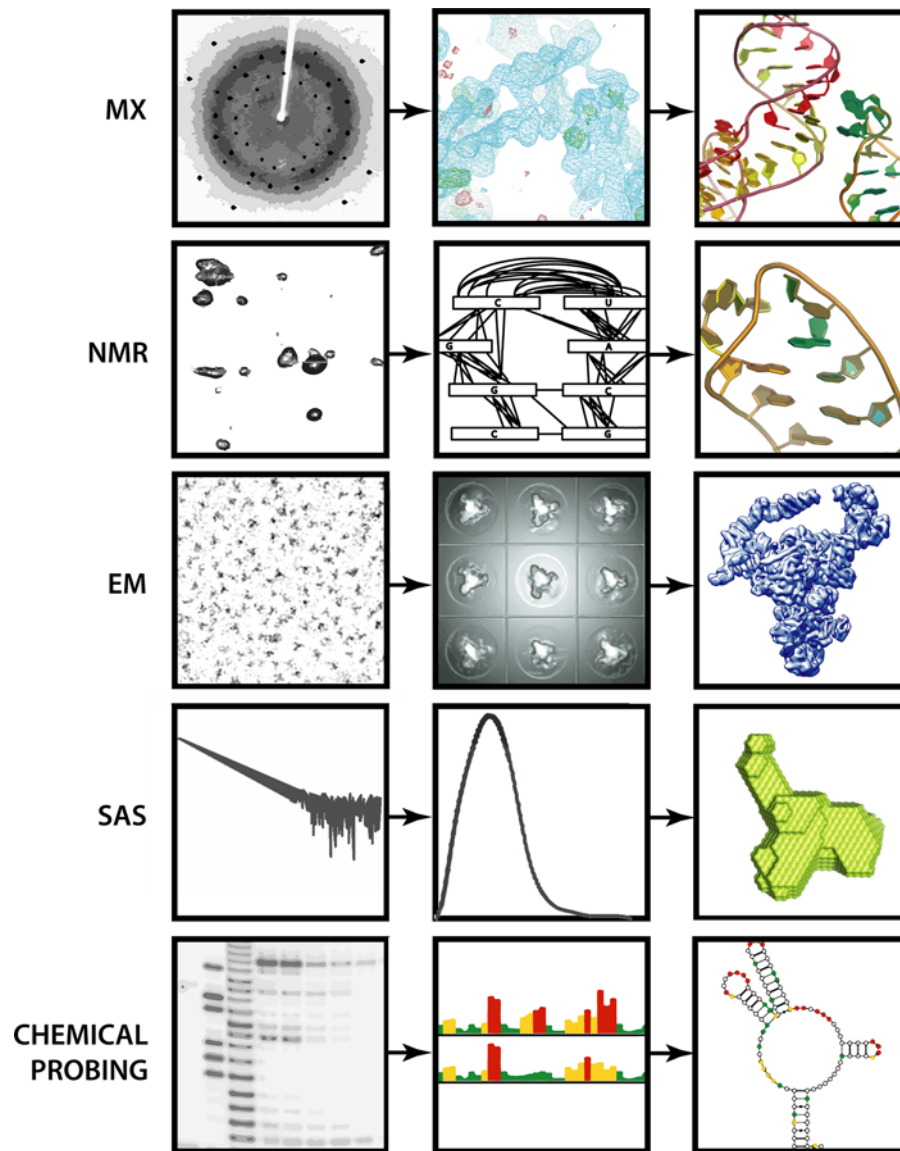
Figure 1. Schematic illustration of selected experimental methods applicable to RNA structure determination: from data acquisition, through data processing, to the generation of an atomic model, 3D shape, or secondary structure representation

which the original sequence is 'mutated' to become the sequence of the 'target' molecule while retaining most of the local and global structural features. The identification of a template structure (one or more) and aligning the target and template sequences (to determine residues that are spatially equivalent) is the first and the most crucial step of this kind of modeling. Thus, the success of comparative modeling is limited by the availability of experimentally solved RNA 3D structures that can serve as templates, and hence this approach is heavily dependent on macromolecular structure databases.

Typically, comparative modeling can produce highly accurate models for target molecules that exhibit high sequence similarity to the templates, indicative of a close evolutionary relationship. Evolutionarily related RNAs with divergent sequences often retain the ancestral tertiary folds, very much like proteins do [140]. However, with the increasing evolutionary distance (and sequence dissimilarity) between the target and the template, the likelihood of obtaining an accurate model decreases, due to the structural divergence and mounting difficulties in the generation of accurate sequence alignments. Technically, it is possible to carry out comparative modeling using templates that are evolutionarily unrelated to the target, but this requires the unrelated template to be structurally very similar to the

11

target. Comparative models based on wrong templates, or on correct templates with wrong alignments, are bound to be wrong.

In the absence of closely related molecules in the database of experimentally determined structures, the detection of templates becomes challenging. This issue is addressed by fold-recognition (FR) methods, which aim at predicting which of the known 3D structures is likely to exhibit a common architecture with the target sequence. FR has been developed for proteins and while it is applicable to RNA, to our knowledge it remains to be implemented in this context. FR can involve sequence similarity searches (typically to identify molecules with a known structure that are evolutionarily related to the target) or threading (finding structures that can accommodate the target sequence in such a way that favorable contacts are formed in 3D) or the combination of these approaches.

Comparative modeling methods often use atomic coordinates of experimentally determined structures in one of two ways. One group of methods, exemplified by ModeRNA [141,142], directly copies atomic coordinates of the structural core from the template, preserving the backbone in regions of continuous alignment, and 'patches up' backbone discontinuities caused by insertions and deletions by stitching the template with short fragments from other structures. Whenever the template structure or additional fragments disagree in sequence with the target, bases are replaced by geometry matching. These type of methods do not require any energy calculations, and scoring is used mostly to minimize gaps in the backbone and steric clashes between the inserted fragments and the core. Another group of comparative modeling methods, exemplified by Modeller [143] or MacroMolecular Builder (MMB, previously known as RNABuilder) [144], use target-template alignment to assign correspondences between residues, and then assign spatial restraints (e.g., bond lengths and torsion angles) to the target sequence, based on the values for corresponding residues in the template. The target sequence is then folded to satisfy these restraints. Also, energy minimization is performed with the use of a force-field that ensures the resulting structure is physically realistic. Additionally, geometry constraints can be used based on general statistics from RNA 3D structures to enforce proper stereochemistry. In both types of approaches, secondary structure information can also be used to constrain the selection of fragments or to restrain the helices to achieve proper interactions and geometry.

Fragment assembly is another approach for structure prediction, which can use atomic coordinates of experimentally determined structures. As opposed to comparative modeling, it does not require a match of the template sequence to any particular structure, so it is essentially a template-free approach. Instead, it builds the structure *de novo* by using segments extracted from known 3D structures as building blocks. Often, the modeling procedure involves the testing of various combinations of fragments, which are assessed by a scoring function that can be based on statistics, physics, or both. Typically, secondary structure information is used in the form of constraints in the process of fragment selection. Additional spatial restraints may also be used. Both in comparative modeling based on restraints and in fragment assembly, additional information may also be encoded: pairwise distance restraints are relatively straightforward to encode and are enabled by most methods, and some methods (e.g., MMB) also allow 3D shape restraints in the form of density maps. There exist methods for manual (semi-automated) modeling with the use of fragment assembly such as RNA2D3D [145] and Assemble [146,147]. There are a few fully automated methods based on a computational assembly of RNA fragments, such as RNAComposer [148] and Rosetta/FARNA/FARFAR [149,150].

RNA 3D structure modeling approaches that use local structural features and pairwise distance restraints (including contact information)

De novo **modeling** is an approach for structure prediction, which by definition does not use templates. '*Ab initio*' modeling methods that rely solely on first principles (i.e., on quantum physics calculations), even with various simplifications, are extremely costly and are incapable of predicting 3D structures for objects as large as typical RNA structures starting from sequence information alone [151]. On the other hand, most *de novo* folding methods enable simulating the RNA structure formation process based on sequence alone or supplemented only with secondary structure information. Therefore they introduce various simplifications and biases to make the prediction process plausible, by trading off accuracy for speed, given limited computing resources. Many of these methods use a coarse-grained representation and simplify or ignore various aspects of RNA structure and dynamics that are not essential for predicting the 3D structure that is as close to the real one, as possible and feasible [134,152]. All successful methods for *de novo* modeling of RNA structure, do utilize prior information on experimentally determined structures, very often in the form of statistics that guide the folding toward conformations that are similar to those observed earlier. It has been shown, e.g., in the series of community-wide RNA Puzzles experiments, that the use of additional information in the form of various additional restraints, contributes greatly to the accuracy of predictions [153–155]. In fact, for

large RNAs, it is still virtually impossible to predict 3D structures accurately (with RMSD to the native structure <10 Å) without the use of structural restraints.

Essentially all methods for RNA 3D structure prediction can use restraints or constraints on secondary structure. Many of the *de novo* folding, as well as fragment assembly methods, actually require a secondary structure to be included as one of the elements of the input, so they either take the secondary structure supplied by the user or they predict it using one or more of the external tools. Many programs can also use additional pairwise restraints derived from different types of experiments, beyond the information about the secondary structure. Multidimensional Chemical Mapping (MCM) is a computational workflow for RNA 3D structure modeling that utilizes Rosetta/FARFAR with restraints based on experimental probing with the M^2 and MOHCA-seq approaches, which provide secondary structure, and tertiary proximity information, respectively [156]. Examples of RNA 3D structure prediction methods that are capable of using secondary structure data as well as other data types, such as restraints, include NAST [157], iFoldRNA-v2 [158], and SimRNA [159].

A few modeling methods specialize in the use of restraints from NMR experiments. The Al-Hashimi group used ^{1}H chemical shift data to restrain MD simulation and suggested the prospects of using chemical shift data as restraints in RNA structural modeling as well as in the qualitative evaluation of RNA dynamic ensembles [160]. ^{1}H chemical shift data obtained from NMR experiments can also be used in an extension of the Rosetta method, Chemical-shift-ROSETTA for RNA (CS-ROSETTA-RNA), to model non-canonical RNA motifs using [161]. iFold-NMR can use inter-imino-proton distances and atomic inter-base distances associated with base-pairing configuration obtained through ^{1}H–^{1}H NMR NOESY and NMR $_{JN}$-COSY experiments, respectively [162].

The integration of XL-MS with computational techniques for molecular 3D structure determination has been initially developed for proteins and termed as MS3D [163]. The feasibility of MS3D for RNA structure determination has been demonstrated for the mouse mammary tumor virus (MMTV) ribosome frameshifting pseudoknot [164] and for the HIV-1 ψ-RNA packaging signal [165]. However, the analysis of RNA–RNA cross-links by XL-MS is difficult and thus far MS3D has not been established as a routine technique for structural studies of RNA.

RNA 3D structure modeling approaches that use 3D shape information

While the information about molecular shapes is the basis of several methods for macromolecular structure determination, it is rarely used as a restraint by modeling methods that are largely predictive. However, a number of computational methods were developed that blur the line between data-driven structure determination and data-driven predictive modeling. Such methods perform flexible fitting of molecular structures into density maps obtained from EM or MX experiments. They were developed to improve the model building in particular for maps that are of too low resolution for traditional structure determination methods to produce reliable atomic coordinates, e.g., at the resolution > 3.5 Å for MX. This type of modeling usually does not construct the whole structure from scratch and should be typically initiated with a 3D model that is relatively close to the expected final conformation, and only small to medium structural changes are introduced. For instance, the MD Flexible Fitting (MDFF) method incorporates the EM data as an external potential added to the MD force field, allowing the internal features present in the EM map to be used in the fitting process. In this approach, the expected conformational changes must be within the range of the MD method [166]. A similar task can be performed with different sampling algorithms, e.g., in the DEN approach, a Deformable Elastic Network is used to restrain the sampling to prior knowledge of an approximate structure, which reduces overfitting, especially at low resolution. The DEN restraints can also be used to enhance reciprocal space simulated annealing refinement [167]. The Das group has developed an extension of Rosetta called DRRAFTER (*De novo* RNP modeling in Real-space through Assembly of Fragments Together with Electron density in Rosetta), which can be used to build RNA coordinates into cryo-EM maps of RNAs and RNPs [168].

Hybrid modeling approaches that involve data-driven modeling

Given the scarcity of experimentally determined high-resolution structures of RNA molecules and their complexes and the wide range of heterogeneous biochemical data, computational techniques can be used to integrate existing data, guide structure elucidation, and subsequently determine the mechanisms of action and interactions between components. This approach is commonly called hybrid or integrative modeling. It has been developed and used mainly for protein complexes, but it has also been applied to systems involving RNA. Typically, the result of integrative modeling is not just a single structure, but rather a series of models that are maximally consistent with the input data. Hence, in the case of insufficient data, a user obtains multiple alternative solutions that can be utilized to plan additional experiments. Theoretical models of macromolecular structures can assist in understanding and guiding
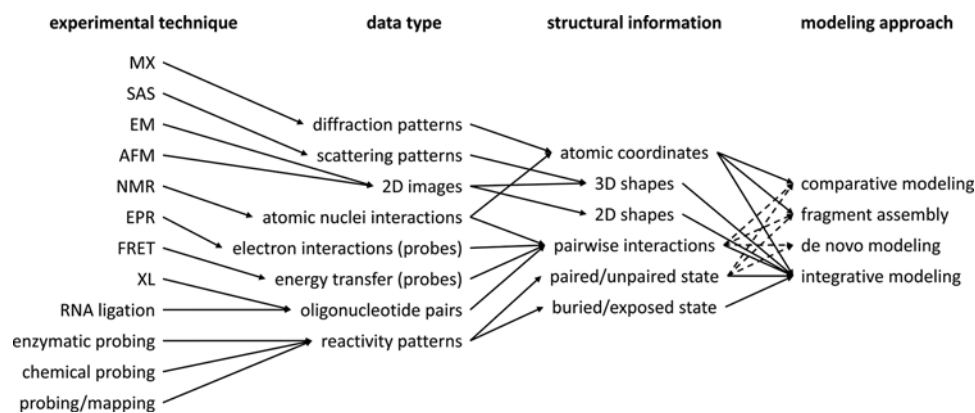
Figure 2. Information flow from experimental techniques to modeling approaches

the identification of important interaction surfaces and, more importantly, can provide a starting point for higher resolution descriptions.

Integrative Modeling Platform (IMP) is an example of a very comprehensive computational toolbox for hybrid modeling, which encompasses various applications for data-driven modeling. It supports the use of multiple types of restraints that are applicable for RNA modeling, including SAXS profiles, EM images and density maps, and sparse NMR data [37]. There are many different sampling protocols implemented. The preparation of the input data for IMP requires advanced knowledge of the system analyzed. IMP has been used to predict macromolecular structures involving RNA, such as ribosomes [169,170]. Another example of a hybrid modeling method is PyRy3D developed in our group (http://genesilico.pl/pyry3d/), which aimed at the relative simplicity of the input while retaining the ability to use various restraints. This method is particularly useful for data-driven modeling of complexes where the users have only limited knowledge about the system analyzed and wish to test alternative hypotheses, e.g., by using restraints that may be ambiguous or mutually incompatible. PyRy3D enables the construction of models for individual molecules as well as for large macromolecular complexes, and it combines the modeling of rigid bodies (e.g., for components with experimentally determined structures) with flexible shapes (e.g., for components with completely unknown structures or for regions of molecules predicted or determined to be disordered). PyRy3D has been used to model RNA and RNA–protein complexes such as the complex between the 2′,5′-oligoadenylate synthetase OAS1 and the 3′-terminal region of the WNV RNA genome [171]. M3 is yet another computational method that incorporates sparse and hybrid experimental data to develop molecular models, which is applicable to RNA [172]. The M3 protocol uses the HADDOCK framework [173], and accepts input in the form of interatomic distances and/or molecular shapes. Modeling with M3 can account for extensive structural rearrangements both at the domain and atomic levels. There also exist other programs and software packages that support integrative modeling, and their utility for RNA 3D structure modeling remains to be verified.

Table 2 summarizes the main features of modeling approaches, with emphasis on data requirements and limitations of applicability. Figure 2 illustrates the flow of information from experimental measurements to computational modeling of RNA 3D structure.

Assessment of RNA 3D structural models

The ability to discriminate correct models from incorrect models is of great importance for RNA 3D structure prediction. When the experimentally determined structure is available as a reference, a number of computational methods can be used to make *a posteriori* assessment [174] about the accuracy of models. However, in real life cases, the 'true' RNA structure is not available, which defines the purpose of RNA 3D structure prediction. Therefore, the challenge is not only to build 'a model' or 'models' of that RNA structure, but also to assess how similar these models may be to the unknown 'true' structure. Generally, *a priori* model assessment attempts to answer questions such as 'is a given structural model likely to be correct or is it likely to be wrong', 'which parts of the model are likely to be correct, and which parts are uncertain' and 'for a given set of models, which are likely to be the better than others'.

The most obvious way to assess structural models is to compare them with data that were not used in the process of structure prediction and can be considered independent. In crystallography, for instance the model is assessed with the R-work and R-free values. The R-work measures the agreement between the model and the actual experimental data, by calculating a simulated diffraction pattern from the model. The R-work value can be, however, biased

Table 2 Main computational approaches for RNA 3D structure modeling

Modeling technique	Obligatory input data for the target RNA (prior knowledge)	Non-obligatory additional data often used	Main advantages	Main disadvantages
Comparative modeling	• Target RNA sequence • 3D structure of another RNA (at least one) used as a template • Sequence alignment between the target and the template(s)	• Secondary structure	• Computationally inexpensive and hence very fast • Practically no size limitation	• Does not work without the template • Critically depends on the accuracy of the template and the alignment. Models based on a wrong template or on a correct template but with a wrong alignment are bound to be incorrect • Regions that depart from the template structure are modeled poorly
Fragment assembly	• Target RNA sequence • Database of fragments of RNA 3D structures • Assignment of structural fragments to sequence fragments (usually based on predefined secondary structure)	• Secondary structure • Pairwise distance restraints	• Computationally inexpensive and hence very fast • Practically no size limitation	• Usually depends on the accuracy of predicted secondary structure • Complex 3D topologies and pseudoknots are difficult to predict
De novo modeling	• Target RNA sequence	• Secondary structure • Pairwise distance restraints • Partial 3D structure	• Often capable of modeling small RNA molecules (<100 nt long) based on sequence data alone, without any additional information	• Computationally expensive and hence relatively slow • Limited to ~100 nt residues without additional data (longer sequences are modeled very poorly), the limit can be extended by the use of additional data (toward integrative/hybrid modeling)
Ab initio modeling	• Target RNA sequence • Detailed chemico-physical conditions of the system (e.g., ion concentration, pH, temperature etc.)		• Enables modeling based on first principles, only with the knowledge of the RNA sequence and environmental conditions, but without any restraints	• Computationally very expensive and hence very slow • Limited to very small systems, which are very well characterized
Integrative/hybrid modeling	• Target RNA sequence • Various types of experimental data (often partial 3D models, secondary structure, shape restraints, pairwise distance restraints)	• Most data types used are non-obligatory, but the modeling procedure is mostly data-driven	• Enables modeling based on the combination of various types of experimental data	• Applicable to large systems • Computationally expensive and hence relatively slow • Critically dependent on the quantity and quality of data

as the model can be refined to better agree with the experimental data. A less biased measure is the R-free, which measures how well the model can predict a test set of reflections (5–10% of the dataset) that were not used during refinement [175]. For other types of experimental data, there exist specialized methods dedicated to scoring features of 3D models. Many RNA 3D structure modeling methods, such as MC-Sym [176], NAST [157], RACER [39,177], and SimRNA [159] can use SAXS raw data as a part of the model scoring procedure, to 'filter out' conformations that disagree with the experimental shape measurements. In this approach, the theoretical scattering curve of the generated models is calculated and then compared with the experimental SAXS data. MC-Sym generated models obtained with any method can also be scored/filtered with standalone methods, such as Fast-SAXS-RNA [178]. RS3D is another program that uses SAXS data, together with the secondary structure and tertiary contact information, to score RNA structures obtained in a Monte-Carlo simulation [179]. A general purpose method FILTREST3D can score structural models (including RNA 3D structures and RNP complexes) based on a combination of distance restraints with other factors such as local or global structure or molecule shape, and it implements logical operators to enable sets of alternative restraints [180].

What can be done if no additional experimental data exist? A 3D structural model can be assessed statistically: first, if the modeled structure exhibits physico-chemical and geometrical features that are generally expected from a well-folded molecule and second, if a particular model is similar to other models, predicted independently by other computational methods. Many methods for the theoretical prediction of 3D structural model accuracy based on these principles have been developed for proteins, and they have been evaluated regularly in the CASP experiment [181]. Thus far, quality assessment of RNA 3D structures have lagged behind, and only a few such methods are available,

including standalone potentials for scoring 3D structure models [182–184] and internal scoring functions of RNA 3D structure prediction methods (e.g., ROSETTA, SimRNA, and others). We have recently developed a web interface for assessing RNA 3D structure models with a variety of scoring functions (http://genesilico.pl/mqapRNA/, Marcin Magnus and Janusz M. Bujnicki, unpublished) to facilitate the comparison of different approaches.

Results of RNA Puzzles experiments conducted so far indicate that the assessment of RNA 3D structure model quality is a significant bottleneck in RNA structure prediction, as even the best modelers are unable to predict with confidence which of their own models are likely to be better than others [153–155].

Figure 3 illustrates a general workflow of macromolecular modeling applicable to RNA, highlighting the use of experimental data. Major methodological bottlenecks are indicated to suggest areas for future improvement.

Which types of experimental data are available, but are not yet commonly used by modeling methods? How can they be implemented?

Large-scale structural studies of RNA gained momentum much later than efforts to systematically determine protein structures such as 'structural genomics'. In the last decade, many of the strategies used to model 3D structures of proteins have been adapted for RNA structural modeling [185]. Nonetheless, there are some types of experimental data used in protein modeling that have not yet been implemented for RNA 3D structure modeling.

In MX, experimental phasing methods are required for *de novo* structure determination in order to solve the so-called phase problem. Both anomalous dispersion and isomorphous replacement approaches require the incorporation of heavy atoms within the crystals, and the structure determination process relies on the determination of their location (substructure determination) through Patterson or direct methods. The diffraction data collected in the presence of heavy atoms is usually of lower resolution than the data collected from unmodified crystals, and thus the electron density maps obtained after phasing suffer from a lack of features that can make the manual model building very challenging or even impossible. The situation can drastically improve if native data at better resolution and a partial model are available for phase extension. Obtaining the latter could be used by modeling programs that, taking advantage of low-resolution electron density maps, would be able to place helices, known 3D structural motifs or other distinguishable structural features to yield a partial model for phase extension and subsequent all-atom model building.

As mentioned earlier, electron density maps from MX and EM have been used as restraints for modeling of RNA and RNP 3D structures. However, this has been largely limited to data at medium resolution, where individual residues can be distinguished from each other. On the other hand, the EM technique can generate low (e.g., >10 Å) and very low (>>10 Å) resolution maps for macromolecules of unknown structure, which could be used as potentially useful restraints by the modeling methods. In principle, the restriction of the conformational space by any shape restraints is very helpful for modeling, as it saves the time that could be otherwise wasted on analyzing completely wrong molecular architectures. Instead, conformational sampling can be focussed on testing models that are more likely to be similar to the true structure.

2D class averages calculated from cryo-EM micrographs are used in predicting 3D structures of protein assemblies [186,187]. This approach bypasses the necessity of single-particle reconstruction step to obtain a complete 3D density map. Instead, 2D projections of computationally generated 3D models are generated and compared with experimentally obtained 2D projections. This approach has not yet been commonly used for RNA molecules and could be incorporated in RNA 3D structure modeling programs.

Another type of data that could be very useful for RNA modeling is offered by AFM. With the recent advances in this technique, RNA molecules can be visualized to reveal features such as double- and single-stranded regions, and their interactions [66,188]. AFM can be particularly useful for structural studies of relatively big RNA molecules with multiple secondary structure elements that can be arranged on a surface while preserving the geometry of local motifs [67]. Importantly, AFM, as a single-molecule technique can reveal alternative conformations. In principle, AFM images could be used to encode shape restraints in a similar way to 2D class averages from cryo-EM experiments.

Restraints from XL-MS and MS-MS experiments are used in modeling of proteins and their complexes with other macromolecules including RNA–protein complexes [100,189,190]. There are multiple approaches available to access the quality of the theoretically predicted 3D model of proteins using XL-MS data [191–193]. However, despite its success in protein structure analysis and determination, the use of XL-MS has been limited in RNA structural studies, mostly due to the difficulties associated with experimental data collection and analysis. RNA sequences are composed of only four basic building blocks, therefore the unambiguous determination of short oligonucleotides in cross-linking
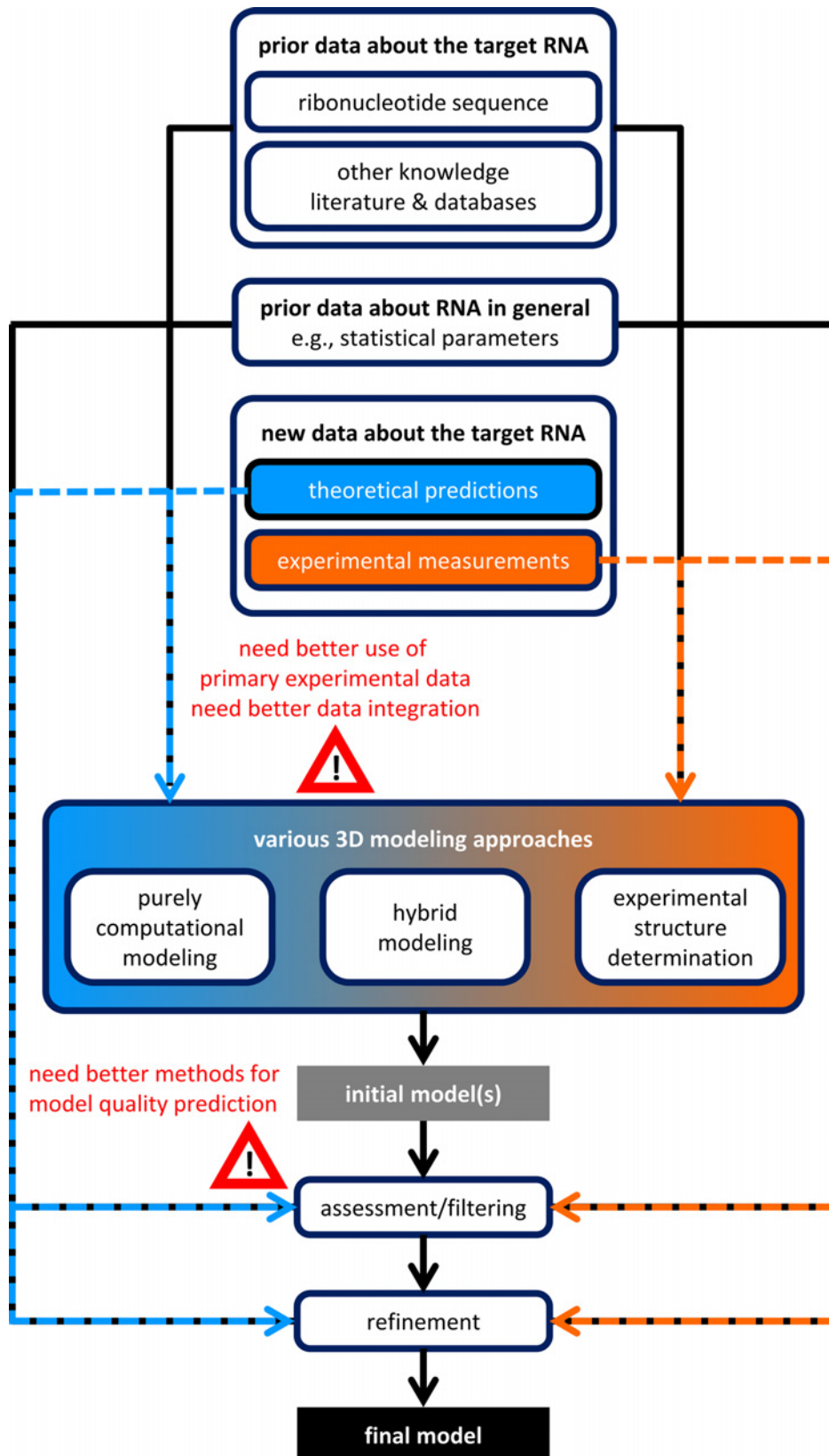
Figure 3. Data workflow in RNA 3D structure determination, with various experimental and theoretical approaches, and combinations thereof

products is much more difficult than for peptides. However, some methods for computational modeling of RNA 3D structure are capable of handling ambiguous restraints or could be modified to enable such functionality. Thus, even if the XL-MS experiment yields a highly unambiguous dataset in which every cross-link can be assigned to a whole list of pairs of oligonucleotide fragments from a given sequence, this could be used as an element of the scoring function, e.g., to eliminate models that are largely incompatible with any of the XL-MS assignments.

Last, but not the least, computational methods for RNA 3D structure prediction could be improved in the way how the footprinting data are used. Reactivity profiles, e.g., from SHAPE and related experiments, are usually first used as an input for specialized secondary structure prediction methods, and only then the predicted secondary structures are used as constraints or restraints in the RNA 3D structure modeling process. However, for some RNA modeling methods, it should be possible, and potentially advantageous, to use the probing data directly at the stage of model building. For example, probing data that indicate the likelihood of individual residues to be paired compared with unpaired could be used as an element of the scoring function without the necessity to specify the binding partners. This could be particularly advantageous for predicting complex RNA structures with pseudoknots that are notoriously difficult to infer for secondary structure prediction tools. On the other hand, modeling tools such as our own program SimRNA, which make and break base pairs and entire secondary structures as a part of the conformational sampling process, do not require any special treatment for pseudoknots, and could be guided to maximize the agreement of the 3D model directly with the probing pattern [159].

Ultimately, we propose that methods for RNA 3D structure modeling that enable the use of experimental data as restraints could be significantly improved. First, there exists a number of experimental techniques that are incapable of unambiguous structure determination, but they generate data that can be potentially used as restraints in molecular modeling. Many of these data are low-resolution and ambiguous, hence modeling methods would have to be adapted to treat them accordingly, and use appropriate statistical frameworks to allow selection amongst multiple alternatives and to tolerate a relatively high level of restraints that could be completely wrong. Second, modeling methods often use highly processed (interpreted) experimental data, which may be misleading. For example, MX and EM data are usually used at the level of 3D density maps and not at the level of diffraction patterns or 2D images. Likewise, footprinting data are not used directly, but in the form of secondary structure predictions. Therefore, we suggest that significant advancement in 3D structure prediction could be made by adapting the existing methods to use experimental data at the possibly early stage of processing, to avoid unnecessary biases. This may require the development of novel ways of treating the experimental data in the prediction process, and as a result, would bring the theoretical approaches closer to what kind of information is actually recorded in the experiments. Such a direction of development may also prompt the theoreticians to collaborate with experimentalists more closely. Computational approaches that deal with the original data may better capture details of biological processes studied experimentally, and could potentially lead to discoveries of new biological phenomena.

Competing interests
The authors declare that there are no competing interests associated with the manuscript.

Abbreviations
AFM, atomic force microscopy; DEN, deformable elastic network; DMS, dimethyl sulphate; EM, electron microscopy; EPR, electron paramagnetic resonance; FR, fold-recognition; FRET, Förster resonance energy transfer; HRF, hydroxyl radical footprinting; IMP, Integrative Modeling Platform; MaP, mutational profiling; MD, molecular dynamics; MMB, MacroMolecular Builder; MR, molecular replacement; MS, mass spectrometry; MX, macromolecular X-ray crystallography; M2, mutate-and-map; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; SAS, small angle scattering; SANS, small angle neutron scattering; SAXS,

PORTLAND PRESS

small angle X-ray scattering; SHAPE, selective 2′-hydroxyl acylation analyzed by primer extension; smFRET, single-molecule FRET; ssNMR, solid-state NMR; XL,   cross-linking; XL-MS,   cross-linking MS; WC, Watson–Crick; 4SU, 4-thiouridine.

References

1   He, L. and Hannon, G.J. (2004) Correction: MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 631–631, https://doi.org/10.1038/nrg1415

2   Fire, A. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811, https://doi.org/10.1038/35888

3   Cech, T.R. and Steitz, J.A. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94, https://doi.org/10.1016/j.cell.2014.03.008

4   Rinn, J.L. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323, https://doi.org/10.1016/j.cell.2007.05.022

5   Berget, S.M., Moore, C. and Sharp, P.A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171–3175, https://doi.org/10.1073/pnas.74.8.3171

6   Chow, L.T. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8, https://doi.org/10.1016/0092-8674(77)90180-5

7   Knapp, G. (1978) Transcription and processing of intervening sequences in yeast tRNA genes. *Cell* **14**, 221–236, https://doi.org/10.1016/0092-8674(78)90109-5

8   Alberts, B., Wilson, J. and Hunt, T. (2008) *Molecular Biology of the Cell*, 5th, Garland Science 1725

9   Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356, https://doi.org/10.1016/S0022-2836(61)80072-7

10   Moore, P.B. and Steitz, T.A. (2011) The roles of RNA in the synthesis of protein. *Cold Spring Harb. Perspect. Biol.* **3**, a003780, https://doi.org/10.1101/cshperspect.a003780

11   Will, C.L. and Luhrmann, R. (2010) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707, https://doi.org/10.1101/cshperspect.a003707

12   Zhang, X. (2018) Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28**, 307–322, https://doi.org/10.1038/cr.2018.14

13   Wilkinson, M.E. (2017) Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection. *Science* **358**, 1283–1288, https://doi.org/10.1126/science.aar3729

14   Wimberly, B.T. (2000) Structure of the 30S ribosomal subunit. *Nature* **407**, 327, https://doi.org/10.1038/35030006

15   Schluenzen, F. (2000) Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623, https://doi.org/10.1016/S0092-8674(00)00084-2

16   Ban, N. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905–920, https://doi.org/10.1126/science.289.5481.905

17   Liu, Z. (2018) Cryo-EM structure of human dicer and its complexes with a pre-miRNA substrate. *Cell* **173**, 1191–1203.e12, https://doi.org/10.1016/j.cell.2018.05.031

18   Berman, H.M. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242, https://doi.org/10.1093/nar/28.1.235

19   Brenner, S.E. (2001) A tour of structural genomics. *Nat. Rev. Genet.* **2**, 801–809, https://doi.org/10.1038/35093574

20   Chandonia, J.M. and Chandonia, J. M. (2006) The impact of structural genomics: expectations and outcomes. *Science* **311**, 347–351, https://doi.org/10.1126/science.1121018

21   Friedmann, D., Messick, T. and Marmorstein, R. (2011) Crystallization of macromolecules. *Curr. Protoc. Protein Sci.* **Chapter 17**, Unit 17.4, https://doi.org/10.1002/0471140864.ps1704s66

22   Zhang, J. and Ferré-D'Amaré, A.R. (2015) Post-crystallization improvement of RNA crystal diffraction quality. *Methods Mol. Biol.* **1316**, 13–24, https://doi.org/10.1007/978-1-4939-2730-2˙2

23   Thompson, R.F. (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3–15, https://doi.org/10.1016/j.ymeth.2016.02.017

24   Camacho-Zarco, A.R. (2015) Multiple paramagnetic effects through a tagged reporter protein. *Angew. Chem. Int. Ed. Engl.* **54**, 336–339, https://doi.org/10.1002/anie.201408615

25   Ponce-Salvatierra, A. (2016) Crystal structure of a DNA catalyst. *Nature* **529**, 231–234, https://doi.org/10.1038/nature16471

26   Shechner, D.M. (2009) Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **326**, 1271–1275, https://doi.org/10.1126/science.1174676

27   McPherson, A. (1999) *Crystallization of biological macromolecules*, vol. 586, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

28   Hoogstraten, C.G., Wank, J.R. and Pardi, A. (2000) Active site dynamics in the lead-dependent ribozyme. *Biochemistry* **39**, 9951–9958, https://doi.org/10.1021/bi0007627

29   Blad, H. (2005) Dynamics and metal ion binding in the U6 RNA intramolecular stem-loop as analyzed by NMR. *J. Mol. Biol.* **353**, 540–555, https://doi.org/10.1016/j.jmb.2005.08.030

30   Dethoff, E.A. (2012) Visualizing transient low-populated structures of RNA. *Nature* **491**, 724–728, https://doi.org/10.1038/nature11498

31   Marcia, M. (2013) Solving nucleic acid structures by molecular replacement: examples from group II intron studies. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 2174–2185, https://doi.org/10.1107/S0907444913013218

32 Deigan, K.E. et al. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 97–102,
    https://doi.org/10.1073/pnas.0806929106

33 Cordero, P. (2012) Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry* **51**, 7037–7039,
    https://doi.org/10.1021/bi3008802

34 Petoukhov, M.V. and Svergun, D.I. (2013) Applications of small-angle X-ray scattering to biomacromolecular solutions. *Int. J. Biochem. Cell Biol.* **45**,
    429–437, https://doi.org/10.1016/j.biocel.2012.10.017

35 Xu, X. (2015) Integrative modeling of macromolecular assemblies from low to near-atomic resolution. *Comput. Struct. Biotechnol. J.* **13**, 492–503,
    https://doi.org/10.1016/j.csbj.2015.08.005

36 Magnus, M. (2014) Computational modeling of RNA 3D structures, with the aid of experimental restraints. *RNA Biol* **11**, 522–536,
    https://doi.org/10.4161/rna.28826

37 Russel, D. (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS*
    Biol. **10**, e1001244, https://doi.org/10.1371/journal.pbio.1001244

38 Smith, L.G. (2017) Physics-based all-atom modeling of RNA energetics and structure. *Wiley Interdiscip. Rev. RNA* **8**, e1422,
    https://doi.org/10.1002/wrna.1422

39 Xia, Z. (2013) RNA 3D structure prediction by using a coarse-grained model and experimental data. *J. Phys. Chem. B* **117**, 3135–3144,
    https://doi.org/10.1021/jp400751w

40 Šulc, P. (2014) A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.* **140**, 235102, https://doi.org/10.1063/1.4881424

41 Shi, Y. (2014) A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995–1014, https://doi.org/10.1016/j.cell.2014.10.051

42 Hammond, C. (2015) *The basics of crystallography and diffraction*, Oxford University Press, Oxford,
    https://doi.org/10.1093/acprof:oso/9780198738671.001.0001

43 Evans, P. and McCoy, A. (2008) An introduction to molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 1–10,
    https://doi.org/10.1107/S0907444907051554

44 McCoy, A.J. and Read, R.J. (2010) Experimental phasing: best practice and pitfalls. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 458–469,
    https://doi.org/10.1107/S0907444910006335

45 Ferré-D'Amaré, A.R., Zhou, K. and Doudna, J.A. (1998) A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621–631,
    https://doi.org/10.1006/jmbi.1998.1789

46 Ke, A. (2004) Crystallization of RNA and RNA/protein complexes. *Methods* **34**, 408–414, https://doi.org/10.1016/j.ymeth.2004.03.027

47 Shoffner, G.M. et al. (2018) In crystallo selection to establish new RNA crystal contacts. *Structure* **26**, 1275–1283.e3,
    https://doi.org/10.1016/j.str.2018.05.005

48 Blakeley, M.P. et al. (2008) Neutron crystallography: opportunities, challenges, and limitations. *Curr. Opin. Struct. Biol.* **18**, 593–600,
    https://doi.org/10.1016/j.sbi.2008.06.009

49 Singh, V., Fedeles, B.I. and Essigmann, J.M. (2015) Role of tautomerism in RNA biochemistry. *RNA* **21**, 1–13,
    https://doi.org/10.1261/rna.048371.114

50 Svergun, D.I. and Koch, M.H.J. (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735–1782,
    https://doi.org/10.1088/0034-4885/66/10/R05

51 Glatter, O. and Kratky, O. (1982) *Small Angle X-ray Scattering*, p. 515, Academic Press

52 Byron, O. and Gilbert, R.J. (2000) Neutron scattering: good news for biotechnology. *Curr. Opin. Biotechnol.* **11**, 72–80,
    https://doi.org/10.1016/S0958-1669(99)00057-9

53 Koch, M.H., Vachette, P. and Svergun, D.I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological
    macromolecules in solution. *Q. Rev. Biophys.* **36**, 147–227, https://doi.org/10.1017/S0033583503003871

54 Hyland, L.L., Taraban, M.B. and Yu, Y.B. (2013) Using small-angle scattering techniques to understand mechanical properties of biopolymer-based
    biomaterials. *Soft Matter* **9**, 10218–10228, https://doi.org/ 10.1039/c3sm51209f

55 Schnablegger, H. and Singh, Y. (2011) The SAXS guide: getting acquainted with the principles. *Austria: Anton Paar GmbH*

56 Skou, S., Gillilan, R.E. and Ando, N. (2014) Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protoc.* **9**, 1727–1739,
    https://doi.org/10.1038/nprot.2014.116

57 Plumridge, A. (2018) Revealing the distinct folding phases of an RNA three-helix junction. *Nucleic Acids Res.* **46**, 7354–7365,
    https://doi.org/10.1093/nar/gky363

58 Jiang, J. (2013) The architecture of Tetrahymena telomerase holoenzyme. *Nature* **496**, 187–192, https://doi.org/10.1038/nature12062

59 Khoshouei, M. (2017) Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* **8**, 16099,
    https://doi.org/10.1038/ncomms16099

60 Merk, A. et al. (2016) Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell* **165**, 1698–1707,
    https://doi.org/10.1016/j.cell.2016.05.040

61 Amunts, A. (2014) Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489, https://doi.org/10.1126/science.1249410

62 Galej, W.P. (2016) Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197, https://doi.org/10.1038/nature19316

63 Qu, G. (2016) Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* **23**, 549,
    https://doi.org/10.1038/nsmb.3220

64 Zhang, K. (2018) Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid cryo-EM, NMR, and molecular dynamics approach. *Structure* **26**,
    490.e3–498.e3, https://doi.org/10.1016/j.str.2018.01.001

65 Shahin, V. and Barrera, N.P. (2008) Providing unique insight into cell biology via atomic force microscopy. *Int. Rev. Cytol.* **265**, 227–252,
    https://doi.org/10.1016/S0074-7696(07)65006-2

PORTLAND
PRESS

66 Schön, P. (2018) Atomic force microscopy of RNA: State of the art and recent advancements. *Semin. Cell Dev. Biol.* **73**, 209–219, https://doi.org/10.1016/j.semcdb.2017.08.040

67 Gilmore, J.L. (2017) Visualization of conformational variability in the domains of long single-stranded RNA molecules. *Nucleic Acids Res.* **45**, 8493–8507, https://doi.org/10.1093/nar/gkx502

68 Gilmore, J.L. (2014) Nanoimaging of ssRNA: genome architecture of the hepatitis C virus revealed by atomic force microscopy. *J. Nanomed. Nanotechnol.* **S5**, 1–7 , https://doi.org/10.4172/2157-7439.S5-010

69 Pallesen, J. (2009) Structure of the HIV-1 Rev response element alone and in complex with regulator of virion (Rev) studied by atomic force microscopy. *FEBS J.* **276**, 4223–4232, https://doi.org/10.1111/j.1742-4658.2009.07130.x

70 Mlynárik, V. (2017) Introduction to nuclear magnetic resonance. *Anal. Biochem.* **529**, 4–9, https://doi.org/10.1016/j.ab.2016.05.006

71 Anglister, J., Srivastava, G. and Naider, F. (2016) Detection of intermolecular NOE interactions in large protein complexes. *Prog. Nucl. Magn. Reson. Spectrosc.* **97**, 40–56, https://doi.org/10.1016/j.pnmrs.2016.08.002

72 Marion, D. (2013) An introduction to biological NMR spectroscopy. *Mol. Cell. Proteomics* **12**, 3006–3025, https://doi.org/10.1074/mcp.0113.030239

73 Mittermaier, A. and Kay, L.E. (2006) New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228, https://doi.org/10.1126/science.1124964

74 Fürtig, B. (2003) NMR spectroscopy of RNA. *Chembiochem* **4**, 936–962, https://doi.org/10.1002/cbic.200300700

75 Davis, J.H. (2005) RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex. *J. Mol. Biol.* **351**, 371–382, https://doi.org/10.1016/j.jmb.2005.05.069

76 Wu, M. and Tinoco, Jr, I. (1998) RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11555–11560, https://doi.org/10.1073/pnas.95.20.11555

77 Barnwal, R.P., Yang, F. and Varani, G. (2017) Applications of NMR to structure determination of RNAs large and small. *Arch. Biochem. Biophys.* **628**, 42–56, https://doi.org/10.1016/j.abb.2017.06.003

78 Davis, B. (2004) Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *J. Mol. Biol.* **336**, 343–356, https://doi.org/10.1016/j.jmb.2003.12.046

79 Burke, J.E. (2012) Structure of the yeast U2/U6 snRNA complex. *RNA* **18**, 673–683, https://doi.org/10.1261/rna.031138.111

80 Keane, S.C. et al. (2015) RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* **348**, 917–921, https://doi.org/10.1126/science.aaa9266

81 Marchanka, A. (2015) RNA structure determination by solid-state NMR spectroscopy. *Nat. Commun.* **6**, 7024, https://doi.org/10.1038/ncomms8024

82 Marchanka, A., Ahmed, M. and Carlomagno, T. (2018) A solid view on RNA: solid-state NMR of RNA and RNP complexes. *Biophys. J.* **114**, 366a, https://doi.org/10.1016/j.bpj.2017.11.2029

83 Yang, Y. and Wang, S. (2018) RNA characterization by solid-state NMR spectroscopy. *Chemistry* **24**, 8698–8707, https://doi.org/10.1002/chem.201705583

84 Huang, W. (2017) Ultraslow domain motions in HIV-1 TAR RNA revealed by solid-state deuterium NMR. *J. Phys. Chem. B* **121**, 110–117, https://doi.org/10.1021/acs.jpcb.6b11041

85 Qin, P.Z. and Dieckmann, T. (2004) Application of NMR and EPR methods to the study of RNA. *Curr. Opin. Struct. Biol.* **14**, 350–359, https://doi.org/10.1016/j.sbi.2004.04.002

86 Frolow, O. (2008) Nitroxide spin labeled RNA for long range distance measurements by EPR-PELDOR. *Nucleic Acids Symp. Ser.* **52**, 153–154, https://doi.org/10.1093/nass/nrn078

87 Halbmair, K. (2016) High-resolution measurement of long-range distances in RNA: pulse EPR spectroscopy with TEMPO-labeled nucleotides. *Chem. Sci.* **7**, 3172–3180, https://doi.org/10.1039/C5SC04631A

88 Sigurdsson, S. (2011) EPR spectroscopy for the study of RNA–ligand interactions. In *Methods for Studying Nucleic Acid/Drug Interactions* (Wanunu, Y., ed.), CRC Press

89 Edwards, T.E. and Sigurdsson, S.T. (2005) EPR spectroscopic analysis of U7 hammerhead ribozyme dynamics during metal ion induced folding. *Biochemistry* **44**, 12870–12878, https://doi.org/10.1021/bi050549g

90 Duss, O. (2015) Combining NMR and EPR to determine structures of large RNAs and protein-RNA complexes in solution. *Methods Enzymol* **558**, 279–331, https://doi.org/10.1016/bs.mie.2015.02.005

91 Stephenson, J.D. (2016) Characterizing 3D RNA structure by single molecule FRET. *Methods* **103**, 57–67, https://doi.org/10.1016/j.ymeth.2016.02.004

92 Sekar, R.B. and Periasamy, A. (2003) Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160**, 629–633, https://doi.org/10.1083/jcb.200210140

93 Millar, D.P. (1996) Fluorescence studies of DNA and RNA structure and dynamics. *Curr. Opin. Struct. Biol.* **6**, 322–326, https://doi.org/10.1016/S0959-440X(96)80050-9

94 Manz, C. (2017) Single-molecule FRET reveals the energy landscape of the full-length SAM-I riboswitch. *Nat. Chem. Biol.* **13**, 1172–1178, https://doi.org/10.1038/nchembio.2476

95 Zhang, Q. (2006) Toward building a database of bifunctional probes for the MS3D investigation of nucleic acids structures. *J. Am. Soc. Mass Spectrom.* **17**, 1570–1581, https://doi.org/10.1016/j.jasms.2006.06.002

96 Harris, M.E. and Christian, E.L. (2009) RNA crosslinking methods. *Methods Enzymol.* **468**, 127, https://doi.org/10.1016/S0076-6879(09)68007-1

97 Hafner, M. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141, https://doi.org/10.1016/j.cell.2010.03.009

98 Kellersberger, K.A. (2004) Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal. Chem.* **76**, 2438–2445, https://doi.org/10.1021/ac0355045

99   Yu, E.T. (2008) The collaboratory for MS3D: a new cyberinfrastructure for the structural elucidation of biological macromolecules and their assemblies using mass spectrometry-based approaches. *J. Proteome Res.* **7**, 4848–4857, https://doi.org/10.1021/pr800443f

100 Dorn, G. (2017) Mapping protein-RNA interactions with single residue resolution by CLIR-MS/MS. *Protocol Exchange*, https://doi.org/10.1038/protex.2017.031

101 Kramer, K. (2014) Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064–1070, https://doi.org/10.1038/nmeth.3092

102 Lu, Z. (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279, https://doi.org/10.1016/j.cell.2016.04.028

103 Aw, J.G.A (2016) *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* **62**, 603–617, https://doi.org/10.1016/j.molcel.2016.04.028

104 Turner, K.B. et al. (2009) SHAMS: combining chemical modification of RNA with mass spectrometry to examine polypurine tract-containing RNA/DNA hybrids. *RNA* **15**, 1605–1613, https://doi.org/10.1261/rna.1615409

105 Fourmy, D. and Yoshizawa, S. (2012) Protein-RNA footprinting: an evolving tool. *Wiley Interdiscip. Rev. RNA* **3**, 557–566, https://doi.org/10.1002/wrna.1119

106 Nichols, N.M. and Yue, D. (2008) Ribonucleases. *Curr. Protoc. Mol. Biol.* **Chapter 3**, Unit 3.13, https://doi.org/10.1002/0471142727.mb0313s84

107 Stern, S., Moazed, D. and Noller, H.F. (1988) Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**, 481–489, https://doi.org/10.1016/S0076-6879(88)64064-X

108 Feng, C. (2018) Light-activated chemical probing of nucleobase solvent accessibility inside cells. *Nat. Chem. Biol.* **14**, 276–283, https://doi.org/10.1038/nchembio.2548

109 Latham, J. and Cech, T. (1989) Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276–282, https://doi.org/10.1126/science.2501870

110 McGinnis, J.L., Duncan, C.D.S. and Weeks, K.M. (2009) High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol.* **468**, 67–89, https://doi.org/10.1016/S0076-6879(09)68004-6

111 Das, R. (2008) Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4144–4149, https://doi.org/10.1073/pnas.0709032105

112 Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616, https://doi.org/10.1038/nprot.2006.249

113 Merino, E.J. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231, https://doi.org/10.1021/ja043822v

114 Mortimer, S.A. and Weeks, K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145, https://doi.org/10.1021/ja0704028

115 Steen, K.-A., Rice, G.M. and Weeks, K.M. (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. *J. Am. Chem. Soc.* **134**, 13160–13163, https://doi.org/10.1021/ja304027m

116 Mortimer, S.A. and Weeks, K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* **4**, 1413–1421, https://doi.org/10.1038/nprot.2009.126

117 Spitale, R.C. (2012) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20, https://doi.org/10.1038/nchembio.1131

118 Spitale, R.C. (2015) Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* **519**, 486–490, https://doi.org/10.1038/nature14263

119 Weeks, K.M. and Mauger, D.M. (2011) Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* **44**, 1280–1291, https://doi.org/10.1021/ar200051h

120 Kertesz, M. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107, https://doi.org/10.1038/nature09322

121 Underwood, J.G. et al. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001, https://doi.org/10.1038/nmeth.1529

122 Wan, Y. (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706–709, https://doi.org/10.1038/nature12946

123 Zheng, Q. (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in *Arabidopsis*. *PLoS Genet.* **6**, e1001141, https://doi.org/10.1371/journal.pgen.1001141

124 Kielpinski, L.J. and Vinther, J. (2014) Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res.* **42**, e70, https://doi.org/10.1093/nar/gku167

125 Cheng, C. et al. (2014) MOHCA-seq: RNA 3D models from single multiplexed proximity-mapping experiments. *bioRxiv.* 004556, https://doi.org/10.1101/004556

126 Strobel, E.J., Yu, A.M. and Lucks, J.B. (2018) High-throughput determination of RNA structures. *Nat. Rev. Genet.* **19**, 615–634, https://doi.org/10.1038/s41576-018-0034-x

127 Siegfried, N.A. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965, https://doi.org/10.1038/nmeth.3029

128 Smola, M.J. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **10**, 1643–1669, https://doi.org/10.1038/nprot.2015.103

129 Smola, M.J. (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10322–10327, https://doi.org/10.1073/pnas.1600008113

130 Lavender, C.A. (2015) Model-free RNA sequence and structure alignment informed by SHAPE probing reveals a conserved alternate secondary structure for 16S rRNA. *PLoS Comput. Biol.* **11**, e1004126, https://doi.org/10.1371/journal.pcbi.1004126

131 Homan, P.J. (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13858–13863, https://doi.org/10.1073/pnas.1407306111

132 Kladwang, W. (2011) A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* **3**, 954–962, https://doi.org/10.1038/nchem.1176

133 Patel, T.R. (2017) Structural studies of RNA-protein complexes: a hybrid approach involving hydrodynamics, scattering, and computational methods. *Methods* **118-119**, 146–162, https://doi.org/10.1016/j.ymeth.2016.12.002

134 Dawson, W.K. and Bujnicki, J.M. (2016) Computational modeling of RNA 3D structures and interactions. *Curr. Opin. Struct. Biol.* **37**, 22–28, https://doi.org/10.1016/j.sbi.2015.11.007

135 Šponer, J. (2018) RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **118**, 4177–4338, https://doi.org/10.1021/acs.chemrev.7b00427

136 Andronescu, M. (2014) The determination of RNA folding nearest neighbor parameters. *Methods Mol. Biol.* **1097**, 45–70, https://doi.org/10.1007/978-1-62703-709-9˙3

137 Kerpedjiev, P., Höner Zu Siederdissen, C. and Hofacker, I.L. (2015) Predicting RNA 3D structure using a coarse-grain helix-centered model. *RNA* **21**, 1110–1121, https://doi.org/10.1261/rna.047522.114

138 Li, J. (2016) Structure prediction of RNA loops with a probabilistic approach. *PLoS Comput. Biol.* **12**, e1005032, https://doi.org/10.1371/journal.pcbi.1005032

139 Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826, https://doi.org/10.1002/j.1460-2075.1986.tb04288.x

140 Capriotti, E. and Marti-Renom, M.A. (2010) Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC Bioinformatics* **11**, 322, https://doi.org/10.1186/1471-2105-11-322

141 Rother, M. et al. (2011) ModeRNA server: an online tool for modeling RNA 3D structures. *Bioinformatics* **27**, 2441–2442, https://doi.org/10.1093/bioinformatics/btr400

142 Rother, M. et al. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* **39**, 4007–4022, https://doi.org/10.1093/nar/gkq1320

143 Šali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815, https://doi.org/10.1006/jmbi.1993.1626

144 Flores, S.C. (2010) Predicting RNA structure by multiple template homology modeling. *Pac. Symp. Biocomput.* 216–227, https://doi.org/10.1142/9789814295291˙0024

145 Martinez, H.M., Maizel, Jr, J.V. and Shapiro, B.A. (2008) RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J. Biomol. Struct. Dyn.* **25**, 669–683, https://doi.org/10.1080/07391102.2008.10531240

146 Jossinet, F., Ludwig, T.E. and Westhof, E. (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057–2059, https://doi.org/10.1093/bioinformatics/btq321

147 Jossinet, F. (2015) Assemble2: an interactive graphical environment dedicated to the study and construction of RNA architectures, in 2015 IEEE 1st International Workshop on Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR). 37–38, https://doi.org/10.1109/VARMS.2015.7151721

148 Popenda, M. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40**, e112, https://doi.org/10.1093/nar/gks339

149 Das, R. and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14664–14669, https://doi.org/10.1073/pnas.0703836104

150 Das, R., Karanicolas, J. and Baker, D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* **7**, 291–294, https://doi.org/10.1038/nmeth.1433

151 Sponer, J.í. et al. (2018) RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **118**, 4177–4338, https://doi.org/10.1021/acs.chemrev.7b00427

152 Dawson, W.K. (2016) Coarse-grained modeling of RNA 3D structure. *Methods* **103**, 138–156, https://doi.org/10.1016/j.ymeth.2016.04.026

153 Cruz, J.A. et al. (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* **18**, 610–625, https://doi.org/10.1261/rna.031054.111

154 Miao, Z. et al. (2015) RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* **21**, 1066–1084, https://doi.org/10.1261/rna.049502.114

155 Miao, Z. et al. (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23**, 655–672, https://doi.org/10.1261/rna.060368.116

156 Cheng, C.Y. (2015) Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife* **4**, e07600, https://doi.org/10.7554/eLife.07600

157 Jonikas, M.A. (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* **15**, 189–199, https://doi.org/10.1261/rna.1270809

158 Krokhotin, A., Houlihan, K. and Dokholyan, N.V. (2015) iFoldRNA v2: folding RNA with constraints. *Bioinformatics* **31**, 2891–2893, https://doi.org/10.1093/bioinformatics/btv221

159 Boniecki, M.J. et al. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **44**, e63, https://doi.org/10.1093/nar/gkv1479

160 Frank, A.T. (2013) Utility of 1H NMR chemical shifts in determining RNA structure and dynamics. *J. Phys. Chem. B* **117**, 2045–2052, https://doi.org/10.1021/jp310863c

161 Sripakdeevong, P. (2014) Structure determination of noncanonical RNA motifs guided by ^{1}H NMR chemical shifts. *Nat. Methods* **11**, 413–416, https://doi.org/10.1038/nmeth.2876

162 Williams, II, B. et al. (2017) Structure modeling of RNA using sparse NMR constraints. *Nucleic Acids Res.* **45**, 12638–12647, https://doi.org/10.1093/nar/gkx1058

163 Young, M.M. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5802–5806, https://doi.org/10.1073/pnas.090099097

164 Yu, E.T., Zhang, Q. and Fabris, D. (2005) Untying the FIV frameshifting pseudoknot structure by MS3D. *J. Mol. Biol.* **345**, 69–80, https://doi.org/10.1016/j.jmb.2004.10.014

165 Yu, E.T. (2008) MS3D structural elucidation of the HIV-1 packaging signal. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12248–12253, https://doi.org/10.1073/pnas.0800509105

166 Trabuco, L.G. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683, https://doi.org/10.1016/j.str.2008.03.005

167 Schröder, G.F., Brunger, A.T. and Levitt, M. (2007) Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **15**, 1630–1641, https://doi.org/10.1016/j.str.2007.09.021

168 Kappel, K. (2018) De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nature Methods* **15**, 947–954, https://doi.org/10.1038/s41592-018-0172-2

169 Taylor, D.J. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure* **17**, 1591–1604, https://doi.org/10.1016/j.str.2009.09.015

170 Chandramouli, P. (2008) Structure of the mammalian 80S ribosome at 8.7 A resolution. *Structure* **16**, 535–548, https://doi.org/10.1016/j.str.2008.01.007

171 Deo, S. (2015) Characterization of the termini of the West Nile virus genome and their interactions with the small isoform of the 2′,5′-oligoadenylate synthetase family. *J. Struct. Biol.* **190**, 236–249, https://doi.org/10.1016/j.jsb.2015.04.005

172 Karaca, E. et al. (2017) M3: an integrative framework for structure determination of molecular machines. *Nat. Methods* **14**, 897–902, https://doi.org/10.1038/nmeth.4392

173 Dominguez, C., Boelens, R. and Bonvin, A.M.J.J. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737, https://doi.org/10.1021/ja026939x

174 Parisien, M. (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* **15**, 1875–1885, https://doi.org/10.1261/rna.1700409

175 Kleywegt, G.J. and Alwyn Jones, T. (1997) Model building and refinement practice. *Methods Enzymol.* **277**, 208–230, https://doi.org/10.1016/S0076-6879(97)77013-7

176 Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55, https://doi.org/10.1038/nature06684

177 Bell, D.R. (2017) Capturing RNA folding free energy with coarse-grained molecular dynamics simulations. *Sci. Rep.* **7**, 45812, https://doi.org/10.1038/srep45812

178 Yang, S. et al. (2010) RNA structure determination using SAXS data. *J. Phys. Chem. B* **114**, 10039–10048, https://doi.org/10.1021/jp1057308

179 Bhandari, Y.R. (2017) Topological structure determination of RNA using small-angle X-ray scattering. *J. Mol. Biol.* **429**, 3635–3649, https://doi.org/10.1016/j.jmb.2017.09.006

180 Gajda, M.J. et al. (2010) FILTREST3D: discrimination of structural models using restraints from experimental data. *Bioinformatics* **26**, 2986–2987, https://doi.org/10.1093/bioinformatics/btq582

181 Elofsson, A. (2018) Methods for estimation of model accuracy in CASP12. *Proteins* **86**, 361–373, https://doi.org/10.1002/prot.25395

182 Capriotti, E. (2011) All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* **27**, 1086–1093, https://doi.org/10.1093/bioinformatics/btr093

183 Norambuena, T. et al. (2013) WebRASP: a server for computing energy scores to assess the accuracy and stability of RNA 3D structures. *Bioinformatics* **29**, 2649–2650, https://doi.org/10.1093/bioinformatics/btt441

184 Bernauer, J. (2011) Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA* **17**, 1066–1075, https://doi.org/10.1261/rna.2543711

185 Rother, K. et al. (2011) RNA and protein 3D structure modeling: similarities and differences. *J. Mol. Model.* **17**, 2325–2336, https://doi.org/10.1007/s00894-010-0951-x

186 Velázquez-Muriel, J. (2012) Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18821–18826, https://doi.org/10.1073/pnas.1216549109

187 Chen, Q. (2017) Reconstruction of 3D structures of MET antibodies from electron microscopy 2D class averages. *PLoS ONE* **12**, e0175758, https://doi.org/10.1371/journal.pone.0175758

188 Schön, P. (2016) Imaging and force probing RNA by atomic force microscopy. *Methods* **103**, 25–33, https://doi.org/10.1016/j.ymeth.2016.05.016

189 Bullock, J.M.A. (2018) Modeling protein complexes using restraints from crosslinking mass spectrometry. *Structure* **26**, 1015–1024.e2, https://doi.org/10.1016/j.str.2018.04.016

190 Rappsilber, J. (2011) The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540, https://doi.org/10.1016/j.jsb.2010.10.014

191 Heymann, M. et al. (2008) MSX-3D: a tool to validate 3D protein models using mass spectrometry. *Bioinformatics* **24**, 2782–2783, https://doi.org/10.1093/bioinformatics/btn510

24

192 Bullock, J.M.A., Thalassinos, K. and Topf, M. (2018) Jwalk and MNXL Web Server: model validation using restraints from crosslinking mass spectrometry. *Bioinformatics* **34**, 3584–3585, https://doi.org/10.1093/bioinformatics/bty366

193 Kahraman, A. (2013) Cross-link guided molecular modeling with ROSETTA. *PLoS ONE* **8**, e73411, https://doi.org/10.1371/journal.pone.0073411