Published in final edited form as:

Curr Opin Struct Biol. 2009 June ; 19(3): 235–238. doi:10.1016/j.sbi.2009.04.004.

Nucleic acids: how high resolution structural biology help us to understand Darwinian evolution

Eric Westhof and

Architecture et réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg, France

Dinshaw J Patel

Structural Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA

Eric Westhof: e.westhof@ibmc.u-strasbg.fr; Dinshaw J Patel: pateld@mskcc.org

The reviews in this issue of *Current Opinion in Structural Biology* are organized from the analysis of DNA and RNA structures to that of nucleic acid complexes and assemblies. We selected elements with intricate structures, from the assemblies of G-quadruplexes to the mysterious and awesome complexity of the group II intron fold. The structures range from the exquisitely precise molecular recognition modes displayed by riboswitches bound to their ligands, to the apparently loose folds of the internal ribosome entry site RNAs. However, all these structures present a striking diversity in architectures; for example, the various internal linkages of the G-quadruplexes, the variety of folds of riboswitches for similar size ligands, and sometimes the same ligand like S-adenosylmethionine (SAM). They all illustrate the modularity of the assembly of nucleic acids, as well as the basis of the amazing functional diversity that Darwinian evolution exploits. In protein complexes with nucleic acids, the recurrence of folds, sometimes surprising and rarely suspected, is also the rule. Two reviews contribute to this point. The chapter on the photolyases and the cryptochromes illustrates how structural similarity can be coupled to functional difference. The chapter on the RNA interference (RNAi) machinery adds further evolutionary comparisons on how variations on similar structures can lead to changes in biological function.

Stephen Neidle has written a comprehensive review on the current status of four-stranded G-quadruplex structures formed by guanine-rich DNA and RNA tracts in the crystalline and solution states. He has outlined the range of four-stranded G-quadruplex folding topologies adopted by telomeric and oncogenic promotor sequences, with the contribution of connecting loop sequences to strand directionality around the G-quadruplex scaffold. Indeed, there is considerable polymorphism in the G-quadruplex fold adopted by four-repeat human telomere sequences, with coexistence of distinct topologies in equilibrium, controlled by flanking sequence and monovalent cation type. Thus, an all parallel-stranded G-quadruplex is favored under crystallization and viscous conditions, while G-quadruplexes adopting so-called 3+1 topologies facilitated by a chain reversal loop are favored in K⁺-containing solution. Further, in contrast to previous assumptions, even isolated guanines can participate in G-tetrad formation, as strikingly demonstrated for the *c-kit* oncogenic promotor G-quadruplex. A future challenge includes structure determination of higher order

alignments of two adjacent G-quadruplexes, as could occur in the biological context of telomeric ends. Attention is also turning to the folding topology of RNA G-quadruplexes, given the presence of guanine-rich tracts in telomeric RNA and in the 5'-untranslated regions of oncogenic RNA sequences, upstream of translation start sites. The unique diversity of G-quadruplex folds, their planar G-tetrad platforms, together with the availability of four grooves, makes this higher order nucleic acid scaffold an attractive therapeutic target. Thus, ligands have been designed that target and stabilize telomeric G-quadruplex folds, thereby impacting on the catalytic activity of the enzyme telomerase, which requires access to single-stranded TTAGGG repeats for telomere elongation. Such ligands, designed so as to exhibit greater selectivity for G-quadruplex over duplex DNA, attempt to maximize stacking of their planar chromophores with the terminal G-tetrads, as well as position their basic side chains for alignment with the backbone phosphates lining the G-quadruplex grooves. A future challenge will involve the development of combinatorial and click chemistry approaches to identify promising ligands that target specific G-quadruplex scaffolds, which could subsequently be fine-tuned using structure-based design approaches, to maximize shape complementarity and hydrogen bond recognition between the components of the complex.

Recent structural research on the sensing domains of riboswitches in their ligand bound states have provided unique insights from the structural perspective on RNA architecture and small molecule recognition, and from the functional perspective on RNA-mediated gene regulation through transcription termination, translational inhibition, and RNA splicing. Alexander Serganov has nicely captured the ongoing excitement in this field where new structures have been reported in the past year on the sensing domains of riboswitches bound to the amino acid lysine, the coenzymes S-adenosylmethionine (SAM) and flavin mononucleotide (FMN), and a Mg^{2+} sensor, as well as efforts to monitor riboswitch conformational transitions on complex formation by biochemical and biophysical methods. Several groups have met the challenge to structurally determine riboswitch structures in the 150–250-nt range and define the conformations of these multi-junctional scaffolds. Striking examples include encapsulation of a bound lysine within a five-helical junctional pocket, where some of the intermolecular contacts are mediated by a bound monovalent K^+ cation, and the anchoring of bound FMN within the six-helical junctional site associated with an unexpected butterfly-like scaffold. A bound Mg^{2+} mediates interaction between the FMN phosphate and the base edges of the RNA in the FMN complex, while a cluster of Mg^{2+} divalent cations are integral to the fold and stability of the binding pocket of the Mg^{2+} sensor, through their ability to mediate long-range interactions. There are now three published structures of distinct SAM riboswitches, with the ligand bound in a different conformation in each case, emphasizing the distinct solutions nature has devised to recognize the same ligand in a diverse range of binding pocket conformations of varying size and complexity. Issues related to riboswitch folding and function have been probed from chemical footprinting, calorimetric, and single-molecule methods. To this end, photolabile riboswitch ligands and force field microscopy have been used in attempts to decipher individual steps and their energetics during the folding trajectory. It is also becoming clear that there is a level of tertiary structure preorganization in the ligand-free state and that gene regulation most probably involves an interplay of thermodynamic and

kinetic components to the switching process. Several structural challenges remain, such as efforts directed toward understanding at the molecular level how the glycine riboswitch utilizes its two tandem sensing domains, to bind this smallest of amino acids in a cooperative manner.

A highlight of the year 2008 has been the resolution of the first crystal structure of a group II intron. It belongs to the IIC subgroup and stems from a halophilic and alkaliphilic bacterium. Introns belonging to subgroup IIC possess distinct characteristics compared to the other two subgroups. First, although the secondary structure displays the six main domains, it contains less structural elements. Second, *in vitro*, the intron is released in the linear form, which implies that attack at the 5'-exon junction is done by a water molecule and not by the hydroxyl group of an internal A. Third, such introns insert 3' of rho-independent transcription terminator helices and are thus restricted to those bacterial lineages. The structure is striking. It shows how domain I encapsulates domain V, carrying the active site, which stands on the other domains II to IV (domain VI is missing). The authors, Navtej Toor, Kevin Keating, and Anna Marie Pyle, describe the achievement and make comparisons to other ribozymes, especially the group I introns for which several crystal structures are also available. Further, the authors compare domain V to the intramolecular helix of the U6 snRNA, central to catalysis in spliceosomal intron removal and discuss recent work on the prp8 protein, which should be located very close to the U2-U6 complex in the active site. The question on how proteins in the spliceosome replace the RNA domains of group II introns, whether they do or do not promote catalysis, will require a long series of crystal structures.

Filbin and Kieft in the following review describe our current structural knowledge of RNA sequences at the 5'-end of viral and cellular mRNAs, which act as internal ribosome entry site (IRES), bypassing the need for one or several of the initiation protein cofactors necessary for translation initiation in eukaryotes (see below the review by Myasnikov and collaborators). Thus, in contrast to the preceding discussion (how RNA is replaced by proteins), here we confront how RNA structures can replace protein cofactors and manipulate ribosome function. The authors, with due caution, make the unexpected observation that in viral RNAs, the more there is of RNA structure, the less is the requirement for the number of protein cofactors. However, in cellular IRESs, it appears that RNA structures are inhibitory. Thus, IRES trans-acting factors (ITAFs) are required to promote translation initiation. Biologically, such differences may reflect the need for rapid translation in viral infection bypassing control by cellular factors, while cellular IRESs should be highly controlled in normal conditions and be used solely under specific conditions like stress.

Markus Muller and Thomas Carell report on recent advances providing structural perspectives on photolyases that use blue light to reverse cyclobutane pyrimidine dimers (CPDs) and 6-4 pyrimidine-pyrimidine lesions, thereby, clearing bacterial genomes of cytotoxic and mutagenic UV-induced photolesions. Available structures of the apo-proteins and their complexes with photolesion-containing DNA duplexes establish that the dinucleotide lesion is flipped out into a hydrophobic binding pocket on the protein surface and precisely oriented within 3 Å of a reduced and deprotonated flavin adenine dinucleotide

(FADH⁻). A second cofactor, involving an antenna chromophore, is positioned about 9 Å from the electron-donating FADH⁻, and together with a nearby array of stacked Trp ring systems, provide insights into the catalytic pathway for electron and energy transfer. Essentially, absorption of blue light by the antenna pigments results in energy transfer to FADH⁻, followed by electron transfer to the photolesion, resulting in the CPD and 6-4 lesions splitting back to dipyrimidine dinucleotides. Mechanisms have also been proposed for repair of the 6-4 lesions, which requires both cleavage of a C–C bond and hydroxyl group transfer. The authors also discuss recent progress on cryptochromes, a related class of blue light photoreceptors, containing additional signal-transmitting domains, which impact on processes such as circadian rhythms. It appears that semi-reduced FADH[•] is the activated cofactor in cryptochromes, suggestive of different operational FAD redox potentials between photolyases and cryptochromes. Structural insights have emerged from the structure of a nucleic acid complex involving a DASH-type cryptochrome, which are now categorized as single-strand specific CPD-photolyases. It is anticipated that future structure–function studies will focus on cryptochromes that are currently less well understood than their photolyase counterparts.

There has been a remarkable outpouring of structures of protein–RNA complexes that participate in the RNA silencing pathway. The review by Nowotny and Wang provide a timely overview of the impact of these structures in addressing mechanistic issues related to various steps associated with recognition, positioning, and cleavage of both long double-stranded RNA (dsRNA) and target RNA message. The review starts by describing the principles involved in recognition of the 2-nt 3'-ends of short interfering RNAs (siRNAs) by the PAZ domain, a small RNA-binding module common to both Dicer and Argonaute (Ago) proteins, followed by rules governing 5'-phosphate recognition by the Mid-PIWI module, corresponding to the C-terminal half of Ago. In each case, the protein targets the nucleic acid sugar-phosphate backbone, consistent with sequence-independent recognition. By contrast, the protein Dicer uses a pair of RNase III endonuclease (endoND) domains to generate a composite binding pocket that asymmetrically cleaves both strands of long dsRNA to generate siRNAs. Insights into this process have emerged from the structure of *Giardia intestinalis* Dicer (composed of PAZ and a pair of endoND domains) in the free state and *Aquifex aeolicus* RNase III (composed of a dimer of endoND and double-stranded RNA-binding domains) bound to a target RNA duplex. These structures support the role of a pair of divalent cations in RNase III-mediated cleavage chemistry and propose that the length of the siRNA cleavage products reflects the spatial separation between the PAZ and the composite endoND domains within the Dicer scaffold. The last section focuses on structures of archaeal and eubacterial Ago proteins in the free state and in complexes with bound guide and RNA target strands. The review summarizes mechanistic insights learnt from recently published structures of complexes of *Thermus thermophilus* Ago bound to a guide DNA strand and with added target RNA. These structural studies define the nucleic acid-binding channel, identify the segment of the guide strand in the binary complex available for recognition by target RNA, and alignment of the guide-target duplex spanning the seed segment in the ternary complex. There remains much to learn about structures of complexes in the microRNA biogenesis pathway and the recently identified piRNA pathway, and will most probably require a combination of crystallographic and cryo-

electron microscopy approaches to overcome challenges associated with the increasing complexity of these large multi-component complexes.

Bruekner, Ortiz, and Cramer produce a review very original in its presentation, since the main core of the subject matter is encapsulated into a movie of RNA polymerase. Six X-ray structures (five of yeast pol II and one from *T. thermophilus* holoenzyme) were used to produce the movie, which incorporates eight sets of atomic coordinates. Clearly, the movie does not include the thermal fluctuations and the proper molecular dynamics movements. The movie is centered on the Nucleotide Addition Cycle and the accompanying movements of the central bridge helix and the trigger loop occurring during NTP binding and translocation. Such visualization tools based on several isolated and static high resolution crystal structures, locked in a different state by an antibiotic or the absence of a substrate, contribute in non-descriptable ways to our deep understanding of such complex macromolecular assemblies.

In bacteria and eukaryotes, initiation of translation is the rate limiting step and the most tightly controlled process. In bacteria, initiation of translation involves only three protein cofactors, while in eukaryotes, numerous large protein factors have been implicated. Myasnikov, Simonetti, Marzi, and Klaholz review the recent progress made in our understanding of these processes, especially in light of data gathered from cryo-electron microscopy. They further compare the processes in the two kingdoms. Although striking progress has been made, still many questions remain. The authors raise, for example, the question as to whether the factors IF1 and IF2 interact directly, as shown for the eukaryotic homologs, eIF1A and eIF5B. They also analyze the question of the localization of the proof-reading activity of the initiator tRNA by IF3.

The review emphasizes the power of coupling cryo-electron microscopy with high resolution crystal or NMR data and homology modeling.

Structural biology remains as fascinating as ever. We now have powerful tools, including very high field NMR, synchrotron radiation at various wavelengths, efficient computer visualization, and calculation programs. Such tools, coupled with organic chemistry know-how and biochemical expertise, should continue to yield astonishing returns in our understanding of molecular life and evolution.