Title
Protein–nucleic acid interactions

Journal
Current Opinion in Structural Biology, 13(1)

ISSN
0959-440X

Authors
Aggarwal, Aneel K
Doudna, Jennifer A

Publication Date
2003-02-01

DOI
10.1016/s0959-440x(03)00004-6

Published in final edited form as:

Curr Opin Struct Biol. 2003 February ; 13(1): 3–5. doi:10.1016/s0959-440x(03)00004-6.

Protein–nucleic acid interactions

Aneel K Aggarwal,

Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, NY 10029, USA

Jennifer A Doudna

Department of Molecular and Cell Biology, Department of Chemistry, Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA

An extraordinary range of protein–nucleic acid interactions negotiates the flow of genomic information between DNA, RNA and protein. The reviews in this section of *Current Opinion in Structural Biology* cover the spectrum of gene expression from DNA packaging, unwinding and repair, mechanisms and regulation of RNA synthesis and post-transcriptional modification, the initiation of protein synthesis to the targeting of nascent polypeptides for secretion or membrane insertion. A recurrent theme in these reviews is the interplay between structure and the dynamics of macromolecular assembly, catalysis and regulation.

Akey and Luger, in the first review, highlight advances in our structural understanding of the packaging of vast linear sequences of DNA into compact units: the nucleosomes. The nucleosome core particle (NCP), consisting of 146 base pairs of DNA wound twice around an octameric core of four histone proteins, is the fundamental building block of packaged DNA (or chromatin) in eukaryotic cells. The 1997 atomic structure of NCP, a milestone in the study of chromatin structure and function, provided the starting point for more recent high-resolution structures, including the NCP containing histone variants that differ subtly in amino acid sequence. Akey and Luger raise the intriguing possibility that these histone variants may modulate the accessibility of the genome to transcription factors. These NCP structures have also set the stage for investigating the roles of chaperones and chromatin remodeling complexes in organizing chromatin. In this context, the authors highlight the recent crystal structure(s) of oligomeric nucleoplasmin, a histone chaperone. The structure suggests a model for histone storage during oogenesis and foreshadows future work on other chaperones that regulate nucleosome assembly.

In the second review, we are aptly reminded that “DNA is not a static and unchanging molecule”, but one that is continually modified by specialized enzymes such as recombinases, helicases and topoisomerases, among many others. Mondragón and colleagues offer a tantalizing look at the mechanisms of some of these enzymes, with an emphasis on intermediates that catch these DNA ‘manipulators’ at different stages of their catalytic pathways. Multiple structures of Cre–*LoxP*, for instance, now provide a fairly complete picture of the sequence of events underlying site-specific recombination, including

aggarwal@inka.mssm.edu.

the order in which the DNA strands are cleaved, exchanged and reconnected. Structures of transposase Tn5 in both presynaptic and postsynaptic DNA complexes are now available, and support a two-metal-ion mechanism of catalysis. The recent structure of *Escherichia coli* DNA topoisomerase III bound to single-stranded DNA frames an important intermediate during the topological rearrangement of DNA. The authors point out similarities between the different enzymes, best exemplified by the recent structure of *Archaeoglobus fulgidus* reverse gyrase, which contains both helicase and topoisomerase domains. These DNA manipulations are the stuff of life, preparing DNA for all manner of cellular processes, including transcription and replication.

The discovery of *E. coli* DNA polymerase I (PolI) in 1956 was followed over the years by the discovery of PolII and PolIII in prokaryotes, and Polα–ε in eukaryotes. Remarkably, in just the past four years, the number of known DNA polymerases has more than doubled with the discovery of the Y-family. These newly discovered polymerases allow cells to cope with unrepaired DNA damage by promoting replication through lesions that would otherwise stall the replication fork. In her review, Yang outlines the main structural features of Y-family DNA polymerases, as derived from recent crystal structures of archaeal Dbh and Dpo4, and eukaryotic Polη. The structures reveal an architecture consisting of the palm, fingers and thumb domains common to all DNA polymerases, and a unique C-terminal domain described variously as PAD (polymerase associated domain) in Polη, a little finger domain in Dpo4 and a wrist domain in Dbh. The thumb and fingers domains are unusually small and stubby when compared to those of replicative DNA polymerases, resulting in a more ‘open’ active site that can accommodate various DNA-distorting lesions. However, despite this coterie of new structures — published within four months of each other — the author notes that we still have much to learn about how each member of the Y-family bypasses a specific lesion, and how replication is coordinated between replicative and repair polymerases in cellular organisms.

A multisubunit RNA polymerase (RNAP) synthesizes mRNA in all cellular organisms. In the past few years, an extraordinary set of multisubunit RNAP structures has begun to uncover the details of how mRNA is transcribed in both prokaryotes and eukaryotes. The past year yielded another crop of exciting RNAP structures, which includes, for the first time, bacterial holoenzymes with the promoter specificity factor, the σ subunit. In their review, Murakami and Darst discuss the rapid developments in our structural understanding of how transcription is initiated, derived from both ‘free’ (from *Thermus aquaticus* [*Taq*] and *Thermus thermophilus* [*Tth*]) and promoter-DNA-bound (*Taq*) holoenzymes. The authors highlight the essential features of the σ subunit, including those that permit the recognition of the conserved −10 and −35 elements of the promoter. Together, the structures help to map the progression of bacterial transcription, from initial promoter recognition, melting of the DNA, the beginning and end of abortive initiation, promoter escape, to the final release of the σ subunit. In addition, despite the structural complexity of the multisubunit RNAPs, the authors note intriguing similarities to the simpler single-subunit phage RNAPs, including channels for nucleotide entry and protein elements that block the elongating RNA product.

The recruitment of RNAP to a promoter is regulated by a mix of transcription factors commonly bound to upstream DNA elements. The stereospecific assembly of these factors

in response to developmental/extracellular cues is central to survival. Ogata, Sato and Tahirov review recent structures of Runx1–CBFβ–DNA and c-Myb–C/EBPβ–DNA complexes that offer insights into cases in which one transcription factor is DNA bound (Runx1) and the other is not (CBFβ), and in which two transcription factors are bound to widely separated DNA sites. We learn that the binding of CBFβ does not alter the structure of Runx1 *per se*, but stabilizes its conformation for optimal DNA recognition. c-Myb and C/EBPβ interact via their DNA-binding domains attached to distant sites, with the intervening DNA looped out, as visualized by atomic force microscopy (AFM). A striking AFM image is one in which oncogenic v-Myb—known not to cooperate with C/EBPβ — does not loop out the intervening DNA. The authors discuss some of the factors influencing DNA looping in both eukaryotic and prokaryotic transcriptional systems, the connotations of which extend to genetic events ranging from nucleosome assembly to DNA transposition.

A critical post-transcriptional event is the chemical modification of RNA in organisms from bacteria to man. The extensively modified bases in tRNAs, for instance, are a reminder of the importance of these modifications to cellular life. In his review, Ferré-D'Amaré presents advances in our structural understanding of two types of RNA-modifying enzymes: pseudouridine synthases, which catalyze the isomerization of a specific uridine to a pseudouridine, and tRNA guanine transglycosylases (TGTs), which replace a guanine base with a 7-deaza-guanine derivative. Pseudouridine (ψ) is the most frequent base modification in RNA and is familiar as the conserved middle base in the TψC loop of tRNAs. Structures of several ψ synthases from *E. coli* have been determined recently, including 'free' TruA, RsuA in complex with uracil/UMP and TruB bound to a *22*-nucleotide portion of a tRNA. In spite of a lack of sequence similarity, the structures reveal a common catalytic domain with an invariant catalytic aspartic acid residue. The TruB–RNA complex shows that the enzyme gains access to the substrate by using a base-flipping mechanism that is likely to be shared by other ψ synthases, as well as by ribonucleoproteins responsible for producing ψ residues in rRNAs. However, as the author points out, the mechanism by which the substrate base is actually detached from the sugar, rotated and then reattached to the sugar is a puzzle at present. Ferré-D'Amaré also draws parallels between ψ synthases and the recently determined structure of an archaeal TGT. Like ψ synthases, the TGT contains a catalytic aspartic acid residue and a structurally homologous domain for interactions with RNA. Whether this kind of sharing of domains among unrelated RNA-modifying enzymes is widespread will become clearer with further structural work.

The initiation of protein synthesis is a complex process involving the assembly of a translation-competent ribosome at the start codon of an mRNA. In several coordinated steps, initiator tRNA binds to the small ribosomal subunit, this complex associates with the mRNA and locates the initiation codon, and the large ribosomal subunit joins to produce a translationally active ribosome. As reviewed by Sonenberg and Dever, structural studies of the initiation factors that mediate this assembly process have revealed several folds common to both prokaryotes and eukaryotes, suggesting evolutionary conservation of a core set of proteins essential for translation. Factors that recruit ribosomes to mRNA, unique to eukaryotes, form large multisubunit complexes that bind to the ribosome and both ends of the mRNA. The authors discuss recent progress in understanding the molecular basis of these interactions from both NMR and X-ray crystallographic data for proteins that bind the

5′ and 3′ ends, respectively, of mRNA—eIF4E and polyA-binding protein (PABP)—as well as for parts of eIF4G involved in multiple recognition events. Whereas current work has focused on single polypeptide factors, future structural studies will focus increasingly on multisubunit factors and complexes, as well as the interactions of these proteins with the ribosome and the translating mRNA.

A secretory or membrane protein is escorted to the membrane as soon as its nascent polypeptide emerges from the ribosome. The escort is the signal recognition particle (SRP), a ribonucleoprotein that binds the emerging polypeptide and delivers it to the endoplasmic reticulum in eukaryotes or to the plasma membrane in prokaryotes. The mammalian SRP consists of a ~300 nucleotide RNA that can be divided into S and Alu domains, and six proteins: SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72. In their review, Sauer-Eriksson and Hainzl stress recent crystallographic studies that elucidate the roles of SRP19 and SRP54 in S-domain assembly. These new structures, together with earlier work on SRPs (including bacterial SRP), offer compelling comparisons between species, between binary and ternary complexes, and between cognate and noncognate complexes. From these comparisons, the authors point out intriguing commonalities and differences in RNA–RNA and RNA–protein interactions, as well as in the sequence of conformational transitions in S-domain assembly in all three kingdoms of life.

Together, these reviews truly capture the pace and excitement of structural biology in elucidating genomic interactions. The remarkable snapshots of the assemblages and ‘molecular machines’ provide unprecedented detail on the processing of genomic information. These, combined with the challenge of looking at even larger assemblies and their underlying dynamics, presage an era of increasing synergism between X-ray crystallography, NMR, cryo-electron microscopy and single-molecule techniques.

Biography

The Aggarwal group studies the structures of enzymes and transcriptional/translational regulators, and their interactions with nucleic acids.

The Doudna laboratory studies the molecular structure and function of RNA, focusing on ribozymes and the large RNAs involved in protein synthesis initiation and protein trafficking.

Abbreviations

Ψ	pseudouridine
AFM	atomic force microscopy
NCP	nucleosome core particle
RNAP	RNA polymerase
SRP	signal recognition particle
TGT	tRNA guanine transglycosylase