Published in final edited form as:

Biopolymers. 2013 December ; 99(12): 910–915. doi:10.1002/bip.22311.

Physical Chemistry of Nucleic Acids and their Complexes

Rodolfo Ghirlando and Gary Felsenfeld

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892

Abstract

Studies of the physical properties of nucleic acids began almost immediately following the discovery of the DNA structure. Early investigations focused on the stability and specificity of multi-strand polynucleotide complexes, then gradually on their interaction with other molecules, particularly proteins. As molecular and structural biology expanded to provide detailed information about biochemical mechanisms, physical studies eventually acquired the additional constraint that they should be relevant to functioning biological systems. We describe work in our laboratory that began with investigations of relatively simple questions about the role of electrostatic interactions in the stabilization of multi-strand nucleic acid structures, and evolved to studies of chromatin structure in vitro and within the nucleus.

Since the discovery of the DNA double helix, physical chemists have continued to ask the same basic questions, using increasingly sophisticated tools and obtaining increasingly detailed answers. These questions include: 1. What governs the stability and base pairing specificity of structures that bring together polynucleotide chains with high negative charge density? 2. What structures other than the A and B forms can be formed by DNA, and what structures can be formed by single strand RNA? 3. What proteins and other molecules of biological importance can interact with nucleic acids? Are these interactions specific for structure or nucleotide sequence and, if they are, what physical and chemical mechanisms are involved? Work in our laboratory over many years has addressed aspects of most of these questions.

Almost immediately following the publication of the DNA structure, efforts were made to discover whether the strands of the duplex could be dissociated, and more importantly whether they could then be reassociated properly with the correct base pairing pattern re-established. The success of those experiments led ultimately to the development of the DNA hybridization techniques that became a powerful analytic tool for molecular biologists. More controlled studies of inter-strand interactions were made possible by the development of methods for synthesizing both ribo- and deoxyribonucleotide homopolymers. In this way the kinetics and thermodynamics of helix formation by well defined complementary polynucleotide strands could be studied. These early studies also led to the discovery of a triple stranded polyribonucleotide structure in which a poly (rA)-poly(rU) Watson-Crick duplex accommodates a second poly(rU) running anti-parallel to the first ^{1,2}. Other three

stranded and four stranded structures were discovered in the following period, as well as other forms of double stranded DNA – notably the left-handed Z form ³.

Early studies in our laboratory were concerned with determining the stability of polynucleotide structures as a function of temperature and ionic conditions. It was shown that the stability of the poly (rA)·poly(rU) duplex depended on ionic strength, and that in the absence of other supporting electrolyte roughly an equivalent of Mg⁺⁺ per mole of nucleotide was needed to allow complete duplex formation. ⁴ This work was gradually extended to the study of interactions with other small cations, such as spermine and spermidine, and then with polylysine and polyarginine as models for positively charged proteins. ⁵ We also studied, in collaboration with Martin Gellert, the binding of actinomycin to DNA and showed that the principal driving force for the reaction was entropic, attributable to the loss upon binding of water molecules associated with actinomycin's cyclic peptides. ⁶

The polymer properties of single strand polynucleotides were, at that time, another important focus of nucleic acid physical chemistry. We had shown ⁷ through thermal denaturation studies that poly (rA) in solution exhibited non-cooperative base stacking interactions similar to those seen in the dimer ApA; hydrogen bonding between adenines did not seem to contribute to this structure. This led us to a more extensive study in which we applied the methods previously developed for study of synthetic polymers to this polynucleotide. ⁸ Work by Flory ⁹ and others had characterized the ideal properties of polymers under θ solvent conditions, but no measurements had been carried out on any polyelectrolyte. We were able to find θ solvent conditions for poly (rA), and showed that, under these conditions, the length dependence of radius of gyration measured by light scattering, as well as the length dependence of sedimentation coefficient and viscosity, reflected ideal random coil behavior. Further analysis led to the conclusion that at low temperature poly (rA) could form an extended base-stacked structure that would be disrupted non-cooperatively as the temperature was raised. ¹⁰ We later extended these studies to poly (rU) ¹¹ and apurinic acid ¹². In the case of poly (rU) it was again possible to make measurements under θ solvent conditions. Unlike poly (rA), no temperature-dependent base stacking was observed in poly (rU), and the overall polymer dimensions were quite similar to those of poly (rA) at high temperature, where base stacking does not occur. A principal point of these studies was to determine the relative contributions of base stacking and intrinsic backbone rigidity to overall polynucleotide dimensions. From the available data on rotational restrictions, bond angles and distances, and with the use of Flory's methods ¹³ of introducing local constraints in the prediction of overall polymer dimensions, it was possible to account reasonably well for the observed properties of poly (rU). Although these studies provided a clear picture of the solution properties of these synthetic homopolymers, they could of course say nothing about the more interesting question of the structure of naturally occurring RNAs (see below).

Our interest in interactions between DNA and positively charged polypeptides ultimately led to a decision to focus on chromatin, the nuclear complex consisting primarily of DNA and histones. The histones are basic proteins that help to compact the DNA; it is now well understood that they are also critical participants in the mechanisms that regulate the

biological function of DNA. Not surprisingly, it was suggested quite early that bound histones might serve to block transcription, thus allowing only an appropriate subset of genes to be expressed in a given cell type, or under specific conditions. In early experiments we searched for extended regions of histone-free DNA that might correspond, in this model, to expressed genes. We could not detect such regions and suggested that gene activation might require only that the gene's promoter be free of histones.

Subsequent biophysical studies in our laboratory have addressed questions of how the histones could both compact the DNA and yet make it available to transcription factors and transcribing polymerases. The discovery of the nucleosome¹⁴ made it possible to apply physical methods to well defined structures. The nucleosome 'core particle' consists of about 146 base pairs of DNA wrapped around an octamer of histones, containing two each of histones H2A, H2B, H3 and H4. In a series of studies we assessed the relative contributions of each of the histones to the compaction of DNA and showed that a core composed of histones H3 and H4 played a central role in nucleosome organization. Complexes of H3/H4 with DNA formed 'kernel' structures with many of the properties of the complete nucleosome.¹⁵⁻¹⁸ Even though an octamer composed solely of H3 and H4 was sufficient to reproduce the compaction and topological properties of an intact nucleosome, it was the assembly of a histone H3/H4 tetramer on DNA to form a 7.5 S species that turned out to be biologically important and that addition of the H2A/H2B pair to this species then resulted in formation of a complete nucleosome.^{15,17,18}

The nucleosome is a very stable particle, but thermal denaturation profiles are consistent with initial release from the histone octamer surface, as the temperature is raised, of the duplex DNA at the ends of the core particle.^{19,20} These observations, in agreement with later studies on the restriction enzyme accessibility of DNA within the nucleosome²¹ demonstrate that the DNA ends are relatively free to detach at least transiently from the histone octamer surface. As we next turned our attention to the physical role of chromatin in the transcription process, this property provided a basis for possible mechanisms by which a transcribing RNA polymerase could negotiate passage on a chromatin template. We found that both SP6 RNA polymerase and yeast RNA polymerase III can transcribe through a nucleosome barrier.²²⁻²⁶ We proposed a spooling mechanism through which the approaching RNA polymerase induces the dissociation of the proximal DNA end from the nucleosome core exposing part of the octamer surface. The exposed surface captures the DNA behind the polymerase to form a loop. As the polymerase transcribes through the DNA the exposed DNA is re-bound onto the octamer eventually reforming the nucleosome behind the polymerase. Again the importance of the histone H3/H4 core is highlighted as RNA polymerase pausing is no longer observed once the H3/H4 contacts are cleared.²³ Subsequent studies in other laboratories have shown that although both the phage SP6 polymerase and the yeast RNA polymerase III can transcribe through an intact nucleosome, eukaryotic RNA polymerase II cannot, but requires co-factors that partly disrupt the histone octamer. However some histone remodeling enzymes and chaperones, such as FACT, have been shown to facilitate transcription and nucleosome survival²⁷. In the case of the chaperone FACT, this has recently been shown to depend on its interaction with the H2A/H2B dimer²⁸.

There are also more general questions about the accessibility to nuclear proteins of DNA packaged as chromatin. How do sequence-specific DNA binding proteins recognize their specific sites on DNA? This is a problem that continues to be studied in many laboratories. To address this question with the tools and knowledge available at the time, we examined the interaction of the transcription factor GATA-1 with a mononucleosome reconstituted on a DNA fragment constructed to contain six canonical GATA-1 binding sites. We found that this nucleosome was in fact able to bind the full complement of six GATA-1 protein molecules. The interesting observation, however, was that such binding occurred without the displacement of the histone proteins from their attachment to DNA, even though the nucleosomal structure was almost completely disrupted as judged by accessibility to nucleases.²⁹ Only the central ~50 base pairs around the nucleosome dyad axis were protected against digestion, but removal of GATA-1 led to re-formation of the intact nucleosome. Much more is now known about these series of events, including histone modification and nucleosome remodeling that are typically involved in formation of an active promoter or enhancer. It must be kept in mind, however, that the initial step requires binding of a transcription factor with the sequence recognition properties that confer specificity. How each transcription factor manages its encounter with chromatin remains an important question. Our interest in nucleosome stability also led to studies of the role that histone variants play in vivo, and how they influence chromatin structure. In a series of experiments studying the genome wide distribution of nucleosome core particles containing the histone variants H3.3 and H2AZ, we found that nucleosomes containing both these variants occupied promoter, enhancer and insulator regions previously thought to be nucleosome free.³⁰⁻³² It became evident that nucleosomes containing both these variant histone proteins are intrinsically less stable and unusually sensitive to salt dependent. We therefore proposed that such nucleosomes not only serve as place holders to prevent the region from occupancy with stable nucleosomes, but also allow for facile displacement by transcription factors.

A large part of our more recent research has been concerned with the relationship between transcriptional activation and chromatin structure. We focused initially on the chicken β -globin locus, where we had previously identified GATA-1 as an important regulatory factor, and characterized both promoter and enhancer elements³³⁻³⁴. This led to a detailed examination of the pattern of histone modifications over the active locus, as well as an adjacent upstream silent, heterochromatic region about 16 kb in length.^{35,36} We wondered how the border between the active and inactive loci was maintained and focused on a nuclease hypersensitive site that lay between them, which we thought might constitute a boundary or 'insulator' element. Over a period of several years, we have identified within this element binding sites for factors that block advancing RNA polymerase, as well as factors that recruit multiple histone modification enzyme complexes. These complexes deposit 'activating' marks on adjacent histones, blocking the advance of silencing modifications. We found that the insulator element also recruits the protein CTCF, which we serves to block activation of a promoter by a distal enhancer when placed between them. That led to the discovery of an important role for CTCF in imprinted gene function at the *Igf2/H19* locus^{37,38,39}, and subsequently to the demonstration in many laboratories of a major role for CTCF in large scale organization of chromatin within the nucleus (see below).

Earlier, we had carried out studies of the structure of chromatin released from the nucleus, which had been shown by electron microscopy to form fibers with a diameter of about 30 nm. We used electric dichroism measurements on similar released fragments to demonstrate that the nucleosome core particles within the chromatin fiber are oriented with their flat faces nearly parallel to the fiber axis and inclined at an angle of 20° - 33°. This arrangement is apparently independent of the species and spacer length⁴⁰. Furthermore, field-free relaxation times supported an arrangement of 6 nucleosomes per turn, consistent with hydrodynamic, scattering and electron microscopic studies. These results of course reflected the properties of a heterogeneous population of chromatin fragments. The survey of the chicken β -globin locus described above, and particularly the identification of the 16 KB heterochromatic domain, provided us with a well defined region of the genome that proved amenable to analysis by classical methods of solution physical chemistry. We released the entire 16 kb fragment from the nucleus by restriction enzyme digestion, then used PCR to detect that fragment in subsequent experiments. Sedimentation coefficients were measured on sucrose gradients that had been calibrated precisely by comparison with results in the analytical ultracentrifuge. In other experiments formaldehyde crosslinked fragments were banded in a cesium chloride gradient. The combination of sedimentation coefficient and buoyant density values allowed us to calculate the frictional coefficient of the particle and model its shape, which is consistent with that obtained for bulk material. In later experiments we applied similar methods to the β -globin locus itself, and demonstrated that this fragment had a similar structure, although the heterochromatin fragment had a slightly higher protein content⁴¹; see also our recent review⁴² for a discussion of chromatin structure within the intact nucleus.

We observed that the 16 kb heterochromatin fragment was quite resistant to nuclease digestion in vivo, reflecting its highly compacted state, and made use of this property subsequently to identify the mechanism through which it is kept constitutively heterochromatic. This mechanism, which has similarities to the mechanism of heterochromatin formation identified in *S. pombe*, requires the processing of low level RNA transcripts from the region into small siRNAs by the enzyme Dicer, and the subsequent targeting to these chromatin sites by Argonaute 2 through the siRNAs. This results in recruitment of the histone deacetylases and histone methyltransferases that mark the chromatin as silent.⁴³

Conclusion

During the many years over which the studies described here have been carried out, the increasing understanding of biological mechanisms has made it more difficult to devise realistic systems in which to study the polymer properties of the macromolecules involved. DNA functions in combination with histones and with sequence-specific binding proteins, and these proteins in turn serve in most cases to recruit hosts of co-factors, all of them in some way required for biological activity. Thanks to great advances in structural analysis by X-ray crystallography and NMR we now know a great deal about the principles that govern protein-DNA interaction. We know less about the functional role of unusual DNA and RNA structures. There is now a considerable literature on the function of Z-DNA, but research on triple stranded nucleic acid structures as well as G-tetraplexes is just beginning to reveal

such roles. More recently, attention has been focused on non-coding RNAs, which can have specific secondary structures that confer properties previously associated only with proteins, including the ability to interact with proteins and target sites on DNA, and with regulatory consequences. The elucidation of such secondary structures is a challenging task.

While the major focus of our research has been on chromatin structure at the level of nucleosomes and individual genes, the studies that have sprung from the identification of CTCF as an important boundary/insulator element have revealed that this protein plays a major role in large scale organization of the eukaryotic genome within the nucleus. More generally, genomes appear to be organized into domains that tend to favor interactions within a domain over those between domains. The methods for analyzing these patterns of interaction, Chromatin Conformation Capture (3C) and its more sophisticated variants, have allowed high resolution mapping of long range interactions within the nucleus. Those interactions are of considerable importance for biological function (see for example ⁴⁴⁻⁴⁶). The analysis of DNA and chromatin at this level of organization will require increasingly sophisticated theoretical and experimental polymer chemistry, as is already clear in recent publications ⁴⁷ physical chemists will have plenty of interesting problems in the coming years.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.

References

1. Felsenfeld G, Rich A. *Biochimica et biophysica acta*. 1957; 26:457–468. [PubMed: 13499402]
2. Felsenfeld GD, D R, Rich A. *J Am Chem Soc*. 1957; 79:2023–2024.
3. Wang AHJ, Quigley GJ, Kolpak FJ, Crawford JL, Vanboom JH, Vandermarel G, Rich A. *Nature*. 1979; 282:680–686. [PubMed: 514347]
4. Huang SL, Felsenfeld G. *Nature*. 1960; 188:301–302. [PubMed: 13716401]
5. Leng M, Felsenfeld G. *Proceedings of the National Academy of Sciences of the United States of America*. 1966; 56:1325–1332. [PubMed: 5339293]
6. Gellert M, Smith CE, Neville D, Felsenfeld G. *Journal of molecular biology*. 1965; 11:445–457. [PubMed: 14267267]
7. Leng M, Felsenfeld G. *Journal of molecular biology*. 1966; 15:455–466. [PubMed: 5915177]
8. Eisenberg H, Felsenfeld G. *Journal of molecular biology*. 1967; 30:17–37. [PubMed: 6077933]
9. Flory, PJ. *Principles of Polymer Chemistry*. Cornell University Press; Ithaca: 1953.
10. Stannard BS, Felsenfeld G. *Biopolymers*. 1975; 14:299–307. [PubMed: 240456]
11. Inners LD, Felsenfeld G. *Journal of molecular biology*. 1970; 50:373–389. [PubMed: 5476918]
12. Achter EK, Felsenfeld G. *Biopolymers*. 1971; 10:1625–1634. [PubMed: 5126129]
13. Flory PJ, J RL. *Journal of Chemical Physics*. 1965; 42:3509–3519.
14. Kornberg RD, Thomas JO. *Science*. 1974; 184:865–868. [PubMed: 4825888]
15. Simon RH, Camerini-Otero RD, Felsenfeld G. *Nucleic acids research*. 1978; 5:4805–4818. [PubMed: 745994]
16. Camerini-Otero RD, Felsenfeld G. *Nucleic acids research*. 1977; 4:1159–1181. [PubMed: 331250]
17. Camerini-Otero RD, Sollner-Webb B, Felsenfeld G. *Cell*. 1976; 8:333–347. [PubMed: 986252]
18. Sollner-Webb B, Camerini-Otero RD, Felsenfeld G. *Cell*. 1976; 9:179–193. [PubMed: 987855]
19. McGhee JD, Felsenfeld G. *Nucleic acids research*. 1980; 8:2751–2769. [PubMed: 7433093]

20. McGhee JD, Felsenfeld G, Eisenberg H. Biophysical journal. 1980; 32:261–270. [PubMed: 6264987]
21. Tims HS, Gurunathan K, Levitus M, Widom J. Journal of molecular biology. 2011; 411:430–448. [PubMed: 21669206]
22. Clark DJ, Felsenfeld G. Cell. 1992; 71:11–22. [PubMed: 1394427]
23. Studitsky VM, Kassavetis GA, Geiduschek EP, Felsenfeld G. Science. 1997; 278:1960–1963. [PubMed: 9395401]
24. Bednar J, Studitsky VM, Grigoryev SA, Felsenfeld G, Woodcock CL. Molecular cell. 1999; 4:377–386. [PubMed: 10518218]
25. Studitsky VM, Clark DJ, Felsenfeld G. Cell. 1995; 83:19–27. [PubMed: 7553869]
26. Studitsky VM, Clark DJ, Felsenfeld G. Cell. 1994; 76:371–382. [PubMed: 8293470]
27. Walter W, Kashlev M, Studitsky VM. Method Enzymol. 2004; 377:445–460.
28. Hsieh FK, K O, Patel SS, Dyer PN, Luger K, Reinberg D, Studitsky VM. Proc Natl Acad Sci USA. 2013
29. Boyes J, Omichinski J, Clark D, Pikaart M, Felsenfeld G. Journal of molecular biology. 1998; 279:529–544. [PubMed: 9641976]
30. Jin C, Felsenfeld G. Genes & development. 2007; 21:1519–1529. [PubMed: 17575053]
31. Jin C, Felsenfeld G. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:574–579. [PubMed: 16407103]
32. Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. Nature genetics. 2009; 41:941–945. [PubMed: 19633671]
33. Evans T, Felsenfeld G. Cell. 1989; 58:877–885. [PubMed: 2776214]
34. Hesse JE, Nickol JM, Lieber MR, Felsenfeld G. Proceedings of the National Academy of Sciences of the United States of America. 1986; 83:4312–4316. [PubMed: 3459175]
35. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Science. 2001; 293:2453–2455. [PubMed: 11498546]
36. Litt MD, Simpson M, Recillas-Targa F, Prioleau MN, Felsenfeld G. The EMBO journal. 2001; 20:2224–2235. [PubMed: 11331588]
37. Bell AC, Felsenfeld G. Nature. 2000; 405:482–485. [PubMed: 10839546]
38. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. Nature. 2000; 405:486–489. [PubMed: 10839547]
39. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenko VV. Current biology : CB. 2000; 10:853–856. [PubMed: 10899010]
40. McGhee JD, Nickol JM, Felsenfeld G, Rau DC. Cell. 1983; 33:831–841. [PubMed: 6871995]
41. Ghirlando R, Felsenfeld G. Journal of molecular biology. 2008; 376:1417–1425. [PubMed: 18234217]
42. Ghirlando R, Felsenfeld G. Biopolymers. 2013; 99:225–232. [PubMed: 23348669]
43. Giles KE, Ghirlando R, Felsenfeld G. Nature cell biology. 2010; 12:94–99. sup pp 91-96. [PubMed: 20010811]
44. Dekker J, Rippe K, Dekker M, Kleckner N. Science. 2002; 295:1306–1311. [PubMed: 11847345]
45. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoiyannopoulos J, Mirny LA, Lander ES, Dekker J. Science. 2009; 326:289–293. [PubMed: 19815776]
46. Byrd K, Corces VG. The Journal of cell biology. 2003; 162:565–574. [PubMed: 12925706]
47. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoiyannopoulos J, Mirny LA, Lander ES, Dekker J. Science. 2009; 326:289–293. [PubMed: 19815776]

Fig. 1.

Dependence of intrinsic viscosity and sedimentation coefficient on chain length for polyriboadenylic acid samples as a function of temperature, under near ideal solvent conditions. At ideal solvent conditions the exponents α_η and α_s have identical values of ~ 0.5 showing that this polyelectrolyte behaves as an ideal random coil, as confirmed by radius of gyration measurements. Redrawn from Eisenberg and Felsenfeld.⁸

Fig. 2.

Proposed model of GATA-1 binding to a nucleosome core particle. The binding of GATA-1 to the reconstitute with six GATA-1 binding sites generates a stable complex in solution in the presence of non-specific competitor DNA. Although the contacts between DNA and histone octamer in the complex are perturbed as judged by susceptibility to nuclease, removal of GATA-1 results in quantitative regeneration of the nucleosome core particle. From Boyes et al. ²⁹

Fig. 3.

Hydrodynamic studies of restriction fragments released from the chicken β -globin locus (see text). (A) A plot of the sedimentation coefficient as a function of the ionic strength for both the 15.51 Kbp heterochromatin fragment (green) and the 16.17 kb β -globin fragment (red) with data obtained from calibrated sucrose gradients. Data are the average of at least two (16.17 kb) or three (15.51 kb) experiments. (B) Banding profiles of the crosslinked 15.51 kb heterochromatin fragment (green) and the 16.17 kb β -globin fragment (red) in a TLS-55 CsCl gradient as analyzed by quantitative PCR. (C) Predicted relationship based on the theory of Broersma between the rod length and frictional coefficient for rod like particles having diameters of either 33 (orange line) or 45 nm (blue line). The experimental frictional coefficients in 100 mM NaCl for the 15.51 kb heterochromatin fragment (green) and the 16.17 kb β -globin fragment (red) are indicated by solid colored lines. The expected rod lengths correspond to chromatin fibers having approximately six to seven nucleosomes per 11 nm turn. From Ghirlando and Felsenfeld⁴¹.

Fig. 4.

Two modes of interaction of the insulator protein CTCF. If CTCF and associated factors interact to form a loop in which the enhancer is excluded from the loop containing the gene and its promoter, expression of the gene will be downregulated (Top). If loop formation brings enhancer and promoter together in the same loop gene expression may be stimulated (Bottom).