LA-UR- 96 - 49

Title: **Neutron Scattering and Nuclear Magnetic Resonance Spectroscopy Structural Studies of Protein-DNA Complexes**

Author(s): E. Morton Bradbury, LS
Paolo Catasti, LS-8
Xian Chen, LS-8
Goutam Gupta, T-10
Brian Imai, LS-2
Robert Moyzis, CHGS
Robert Ratliff, LS-3
Santhana Velupillai, T-10

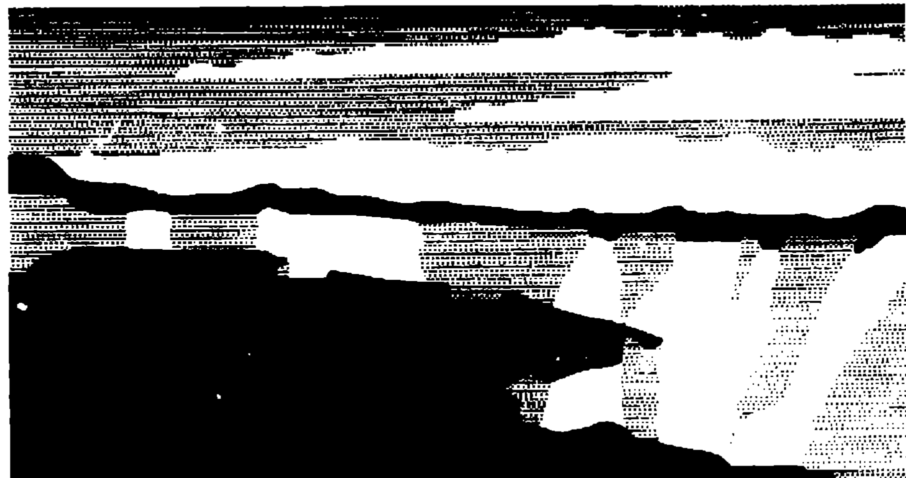
RECEIVED
MAR 13 1996
OSTI

Submitted to: DOE Office of Scientific and Technical Information (OSTI)

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED



Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

MASTER

Form No. 836 R5
ST 2629 10/91

Neutron Scattering and Nuclear Magnetic Resonance Spectroscopy Structural Studies of Protein-DNA Complexes

E. Morton Bradbury*, Paolo Catasti, Xian Chen, Goutam Gupta, Bryan Imai, Robert Moyzis, Robert Ratliff, and Santhana Velupillai

Abstract

This is the final report of a one-year, Laboratory-Directed Research and Development (LDRD) project at the Los Alamos National Laboratory (LANL). The project sought to employ advanced biophysical measurements to study the structure of nucleosomes and the structure of origins of DNA replication. The fundamental repeating unit of human chromosomes is the nucleosome, which contains about 200 base pairs of DNA and 9 histone proteins. Genome replication is strictly associated with the reversible acetylations of histones that unfold chromatin to allow access of factors to origins of DNA replications. We have studied two major structural problems: 1) the effects of histone acetylation on nucleosome structure, and 2) the structure of DNA origins of replication. We have recently completed preliminary x-ray scattering experiments at Stanford on positioned nucleosomes with defined DNA sequence and length, histone composition and level of acetylation. These experiments have shown that lengths of the DNA and acetylations of the histone H4 result in nucleosome structural changes. To understand internucleosomal interactions and the roles of histone H1 we have made preliminary x-ray scatter studies on native dinucleosomes that have demonstrated the feasibility of these experiments. The DNA sequence of the yeast replication origin has been synthesized for structure determination by multi-dimensional NMR spectroscopy.

1. Background and Research Objectives

The nucleosome core particle is the fundamental repeating unit found in eukaryotic chromatin and represents the lowest level of structural organization in the chromosome. Small-angle neutron scattering and x-ray crystallography have shown that the core particle is composed of 146 base pairs (bp) of DNA wrapped in 1.75 left-handed, superhelical turns

* Principal investigator, e-mail: emb@lanl.gov

around a protein octameric core of two each of the histones H2A, H2B, H3 and H4. The overall dimensions of the disc-shaped core particle have been determined to be 55 Å high by 110 Å in diameter. The reversible ε-N-acetylation of lysines in the amino-terminal tails of the core histones has long been associated with gene activation, DNA replication and spermatogenesis. This has led to the theory that acetylation results in a more open and accessible structure of chromatin. Because of the paucity of neutrons at this time, we performed a small-angle x-ray scattering (SAXS) solution study at the Stanford Synchrotron Radiation Laboratory on nucleosomes reconstituted onto a 195 bp DNA molecule that contained a precise nucleosome positioning sequence with either fully unacetylated or acetylated H4 and unacetylated H3.3, H2A.1 and H2B.

In the human genome, 5-10 percent of the chromosomal DNA encodes for proteins. The rest of the DNA was thought to be "junk." However, the notion of "junk" DNA is contradicted by two important observations. First, in higher eukaryotes long stretches of non-coding DNA sequences are conserved in the functional loci such as telomeres, centromeres and replication origins. Slight changes in the DNA sequences of these loci affect their functions. Second, tandem repeats or clusters of DNA sequences present outside the coding region of many human genes have been shown to play crucial roles in gene regulation. For example, the human insulin gene, insulin-like growth factors and their receptor genes and the breast cancer tumor antigen gene, mucin, are under such regulatory controls. These observations can be explained by one of the two mechanisms. Either these non-coding DNA sequences provide specific targets for various gene regulatory proteins or they act as regulatory switches by adopting unusual DNA structures with or without protein factors. Intense research by us on the functional roles of the unusual DNA structures strongly support the second mechanism.

2. Importance to LANL's Science and Technology Base and National R&D Needs

This project supports the Laboratory's core competency of bioscience and biotechnology. This work also enhances LANL's ability to respond to a major initiative in structural biology that has been instituted by DOE's Office of Health and Environmental Research (OHER). The development of a biological neutron scattering program is critical to the long-term DOE support for neutron biological scattering at the Los Alamos Neutron Scattering Center (LANSCE). In addition this project has successfully utilized discoveries made by the Human Genome Project and answered questions of significance in human genetics and disease.

3. Scientific Approach and Results to Date

3.1 Small-Angle X-Ray Scattering Experiments

Histones were purified from HeLa cell nuclei treated with sodium butyrate that inhibits histone deacetylase activity and results in enhanced levels of histone acetylation. The histones were purified to homogeneity by a combination of high-performance liquid chromatography (HPLC) and fast protein liquid chromatography (FPLC) techniques. In order to ensure that any differences seen between the particles were due to differences in acetylation, the same pool of histones which differed only in the level of acetylation of H4 (either zero- or tetra-acetylated) were used to reconstitute nucleosomes onto a 195 bp nucleosome positioning sequence from the Lytechinus 5S ribosomal gene. This particular sequence results in about 18 bp of "extra" DNA beyond the 5' end of the positioning sequence and 31 bp beyond the 3' end of the sequence.

Small-angle x-ray scattering is dependent upon the size, shape and composition of the scattering molecule and the wavelength of the incident x-ray beam. The parameter Q is proportional to the angle of scattering and is defined as $Q = 4psin(q)/l$ where 2q is the angle of scattering from the incident beam and l is the wavelength of the incident light. Figure 1 shows the scattering curves for the tetra-acetylated and nonacetylated particles and the best fit to the data produced by GNOM, a scattering curve-fitting and analysis program. The open triangles are the data points for the tetra-acetylated nucleosomes, while the open circles are the data points for the nonacetylated nucleosomes. The lines represent the best fit to the data as produced by the fitting program GNOM. The abscissa is proportional to the scattering intensity and is an arbitrary scale. The curve for the nonacetylated particle is very similar to previously published data. The undulations in the curve are caused primarily by interference between the superhelical coils of DNA which wrap around the histone core of the nucleosome and to a lesser degree by interference with the histone protein core. The scattering is due more to the DNA coils, as compared to the protein core, because of the higher electron density of DNA relative to protein. The effect of the protein core is to somewhat flatten and slightly shift the overall scattering profile. The tetra-acetylated and nonacetylated curves clearly diverge from each other at higher Q. Additionally, although the peaks and valleys occur at similar values of Q, the features in the tetra-acetylated curve are more flattened and less distinct.

Fourier transformation extracts the pair distance distribution information from the scattering curves and yields the pair distance distribution function or P(r). The P(r) is the probability of finding a pair of scattering centers within the molecule a distance r apart. Additionally, the P(r) is weighted to go to zero when the distance between the scattering centers is zero. Figure 2 shows the P(r) functions for the tetra-acetylated and nonacetylated particles.

Superimposed on the measured P(r) curves are the best fit P(r) curves based on modeling. The top graph shows the nonacetylated model and data. The bottom graph shows the tetra-acetylated H4 model and data. The open symbols are the modeled data points while the solid lines represent the measured data. The most significant information that can be obtained from the P(r) curves is the D_{max} or maximum dimension of the particle which is indicated by the point at which the curve crosses the x-axis. The tetra-acetylated particle clearly has a longer D_{max} than the nonacetylated particle (160 Å vs 140 Å), which indicates that the tetra-acetylated particle has a more extended structure than the nonacetylated particle.

Model fitting was performed on the P(r) curves using Biomod, a Monte Carlo-based solid-modeling and P(r) curve-generating program. This program was modified by the addition of a nucleosome model generation program that models nucleosomes as a series of cylindrical segments. The modeled P(r) curves (Fig. 2) are quite similar to their corresponding measured P(r) curves. The best-fit non-acetylated model had 2.0 turns of DNA wrapped around the histone core with two DNA extensions of 16 and 12 bp extending tangentially from the core. The best-fit tetra acetylated model had 1.83 turns of DNA wrapped around the histone core with DNA extensions of 19 and 23 bp extending tangentially from the core. The modeled and measured curves diverge mostly about the central peak at about 60 Å and above 100 Å. This is best explained by an inability to accurately model the location of the histones tails, which account for up to 20 percent of the total protein mass, due to limitations in the modeling software. The tail mass was incorporated into the central protein cylinder. However, if the tail mass was associated with the DNA extensions, mass would be shifted to the periphery of the model and pair distance lengths would be shifted from the medium range to the longest range. This would be consistent with the interaction of the histone tails with the DNA extensions beyond the 146 bp core region.

Representations of these models are shown in Figure 3. Side and top views of the nonacetylated (left) and tetra-acetylated H4 (right) 195 bp nucleosome models are shown. The models have a DNA superhelical pitch of 28 Å per turn, exterior radius of 55 Å, and a protein core cylinder of 35 Å radius by 55 Å in height. The best-fit non-acetylated model (left) had 2.0 turns of DNA wrapped around the histone core with two DNA extensions of 16 and 12 bp extending tangentially from the core. The best-fit tetra-acetylated H4 model had 1.83 turns of DNA wrapped around the histone core with DNA extensions of 19 and 23 bp extending tangentially from the core.

These results and models are consistent with the decrease in linking number change with acetylation on nucleosomes reconstituted onto closed circular tandem repeats of the Lytechinus nucleosome positioning sequence. This earlier work suggested that reversible histone acetylation could act as a mechanism in eukaryotic cells to introduce supercoils into

chromosomal domains. Our results suggest the molecular mechanism by which this is accomplished. A nonacetylated nucleosome is capable of winding two turns of DNA around the histone core through interactions of the histone N-terminal tails with the DNA extensions beyond 146 bp. Acetylation of the H4 N-terminal tail weakens this interaction sufficiently to cause an unwrapping of 0.17 turns of the DNA from the nucleosome.

3.2 Nuclear Magnetic Resonance (NMR) Spectroscopy Experiments

Here we report the progress of our quantitative research on the functional roles of unusual DNA structures. We focus on three areas of research where unusual DNA structures are expected to play important functional roles: (i) length polymorphism and micro-satellite instability, (ii) DNA methylation, and (iii) tandem repeats or sequence clusters as conformational switches in gene regulations.

i) Micro-satellites are tandem repeats of di-, tri-, tetra-nucleotides, or longer sequences that are interspersed throughout the genome. The micro-satellites often exhibit length polymorphisms, i.e., variation in the repeat number. These length polymorphisms serve as genetic markers in many diseases like fragile X syndrome (FRAX), myotonic dystrophy (DM), Huntington's disease (HD), and breast cancers. In all genetically inherited neuromuscular disorders (i.e., FRAX, DM, HD, etc.) the massive expansion of the triplet repeats coincides with the onset of the disease. Therefore, it is important to ask: what is so special about these tandem repeats that naturally causes length polymorphisms and in some cases massive expansions? We unambiguously demonstrate that the single strands of these disease related triplets tend to form hairpin structures during replication leading to slippage and expansion of the repeat.

ii) Many human genes are silenced by DNA methylation in which the enzyme, human methyltransferase, specifically attaches a methyl group at position 5' of the cytosine residues in the CpG steps. In many cases methylation-dependent gene expression directly corresponds to the initiation or progression of a disease. For example, in FRAX, methylation of the 5' CpG island silences the FMR-1 gene and signals the onset of the syndrome. In one third of the breast cancer patients, the estrogen receptor (ER) gene is silenced by methylation of the 5' CpG sites. These observations raise the question: is there something special about these CpG-rich sequences on the 5' side of a gene that makes them more susceptible to methylation? Indeed, we have shown that the preferred substrate of methylation is a hairpin not a Watson-Crick duplex. The gene regulatory sequence tends to form a hairpin with CC pairs (not GC pairs) at the target CpG site of methylation. The presence of CC pairs in the hairpins stimulate methylation.

iii) Special DNA sequences are located on the 5' control side of many growth factors and their receptors. These sequences, like the telomeres, consist of tandem repeats or clusters of G- or C-stretch followed by a spacer. This prompts us to ask: do the sequences of the G- and C-rich strands also form telomere-like structures? And if so, what is the mechanism by which these unusual structures regulate transcription? Here again we have shown that both the G- and C-rich sequences of these regulatory sequences form unusual hairpin structures that facilitate transcription by creating nuclease hyper-sensitive sites and negative super-helicity.

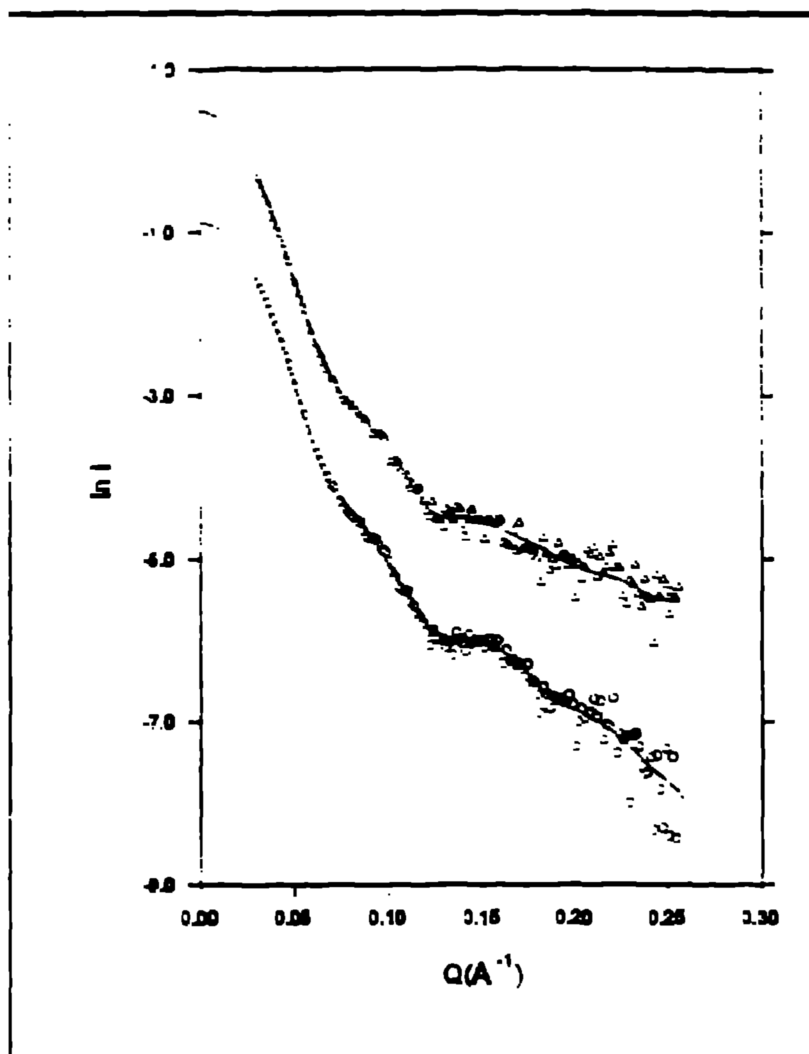


Figure 1: Small-angle scattering data and corresponding best-fit curves for nonacetylated and tetra-acetylated 195 bp nucleosomes.

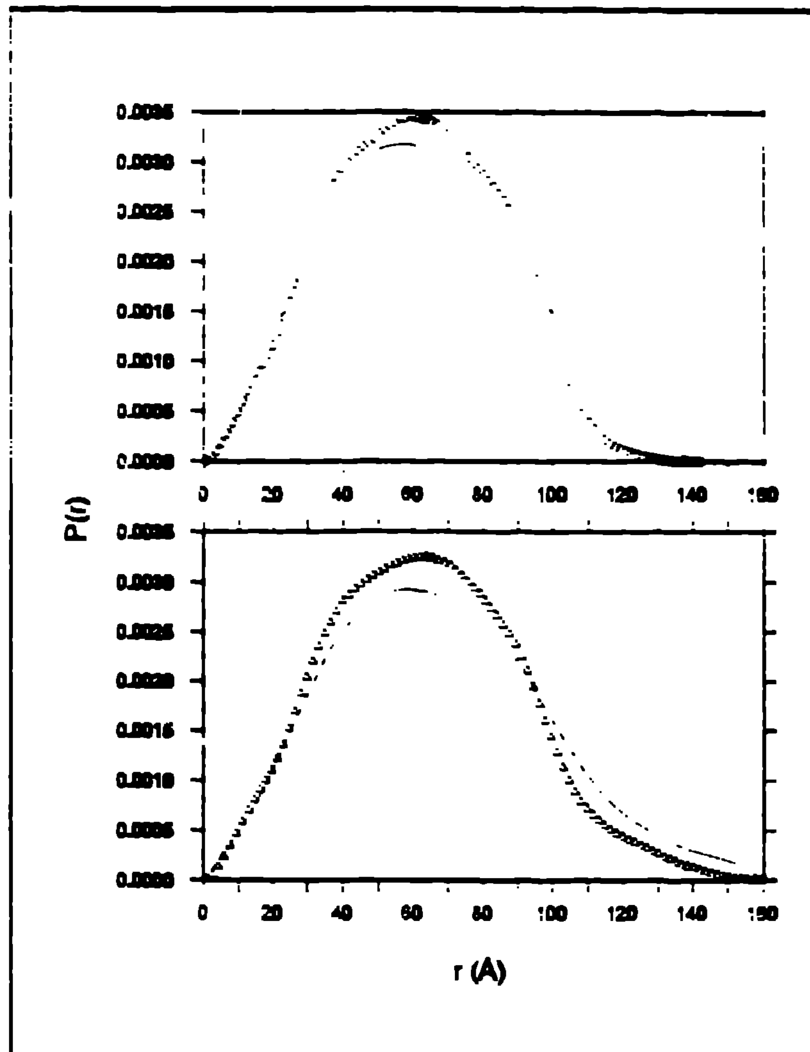


Figure 2: Pair-distance distribution functions derived from small-angle scattering data and best-fit models.

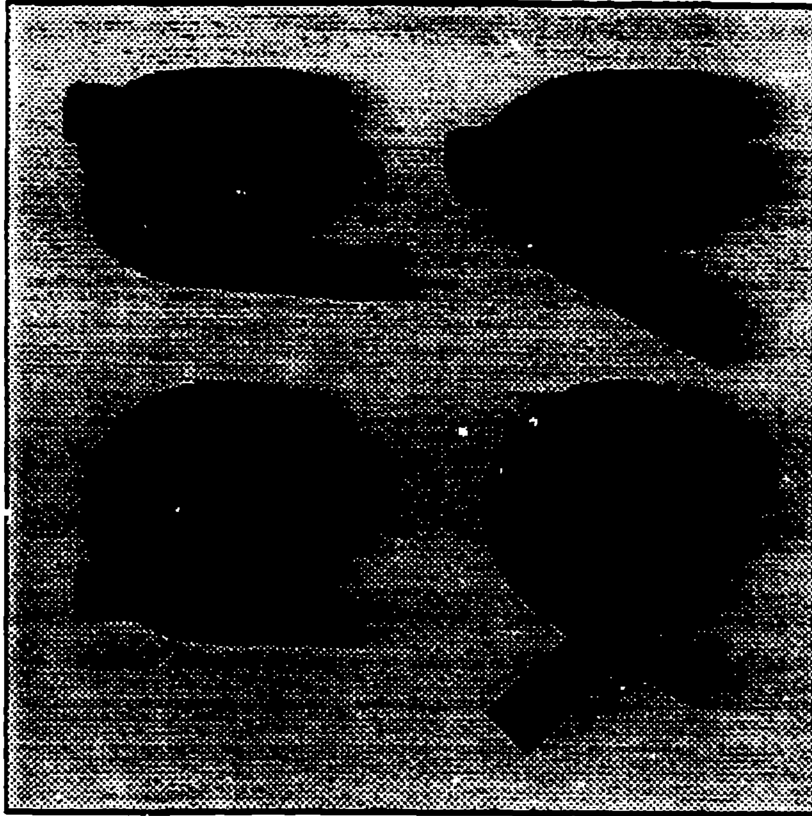


Figure 3: Three-dimensional representations of the model used to generate the best-fit P(r) functions to the measured P(r) functions.