

LA-UR. 96-0241

Title: Development of Experimental Techniques to Study Protein and Nucleic Acid Structures

RECEIVED
APR 01 1996
OSTI

Author(s): Jill Trehwella, CST-4
E. Morton Bradbury, LS
Goutam Gupta, T-10
Brian Imai, LS-2
Rodolfo Martinez, CST-4
Clifford Unkefer, CST-4

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



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.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

pic

MASTER

Form No. 836 R5
ST 2629 1091

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

Development of Experimental Techniques to Study Protein and Nucleic Acid Structures

Jill Trehwella*, E. Morton Bradbury, Goutam Gupta, Brian Imai,
Rodolfo Martinez, and Clifford Unkefer

Abstract

This is the final report of a three-year, Laboratory Directed Research and Development (LDRD) project at the Los Alamos National Laboratory (LANL). This research project sought to develop experimental tools for structural biology, specifically those applicable to three-dimensional, biomolecular-structure analysis. Most biological systems function in solution environments, and the ability to study proteins and polynucleotides under physiologically relevant conditions is of paramount importance. We have therefore adopted a three-pronged approach which involves crystallographic and nuclear magnetic resonance (NMR) spectroscopic methods to study protein and DNA structures at high (atomic) resolution as well as neutron and x-ray scattering techniques to study the complexes they form in solution. Both the NMR and neutron methods benefit from isotope labeling strategies, and all provide experimental data that benefit from the computational and theoretical tools being developed. We have focused on studies of protein-nucleic acid complexes and DNA hairpin structures important for understanding the regulation of gene expression, as well as the fundamental interactions that allow these complexes to form.

1. Background and Research Objectives

The development of both experimental and theoretical capabilities is important for the study of biomolecular structure and function for a broad range of applications, including the utilization of information derived from the human genome, which is important in understanding the molecular basis of disease, rational drug design, the development of biomolecular materials, and bioremediation. The goal of this project was to enhance experimental structural biology capabilities in the areas of NMR spectroscopy as well as both neutron and x-ray

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

scattering and diffraction methods through applications to protein and DNA/protein structure studies.

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 the Department of Energy's Office of Health and Environmental Research (DOE/OHER). The initiative seeks to build up the unique facilities of the national laboratories as structural biology centers for external and internal users and to strengthen the in-house structural biology research programs to address the missions of the DOE. Stated missions are the structural problems associated with the Human Genome Project, with DNA damage and repair, and the application of unique capabilities to important biological problems. This work complements ongoing LANL programs in structural biology. This project also lays the foundations for structural biology applications in areas of strategic interest to the Laboratory and to the nation including health applications, technology development (e.g. biomolecular materials), and the environment (bioremediation). A major benefit will be an understanding of how genetic information is translated (transcription), which will be a major factor in interpreting and using the information now being obtained from the Human Genome Project.

3. Scientific Approach and Results

Our experimental approach is to use methods applicable to three-dimensional, biomolecular-structure analysis, including structures in solution. Most biological systems function in solution environments, and the ability to study proteins and polynucleotides under physiologically relevant conditions is of paramount importance. We have, therefore, adopted a three-pronged approach, which involves capabilities in crystallographic and NMR methods, to study protein and DNA structures at high (atomic) resolution as well as neutron and x-ray scattering techniques to study the complexes they form in solution. Both the NMR and neutron methods benefit from isotope-labeling strategies, and all provide experimental data that benefit from the computational and theoretical tools developed at the Laboratory

For several decades crystallography was the only method for obtaining complete three-dimensional high-resolution structures of biomolecules. It remains one of the most powerful approaches, and recent developments in synchrotron facilities and Laue diffraction methods

have opened new horizons for time-resolved studies of biomolecules. Over the past decade, there has also been a revolution in NMR applications to biomolecular structure analysis. During the last three years uniform carbon-13 and nitrogen-15 labeling coupled with three- and four-dimensional NMR methods have made possible one-to-one chemical shift assignments for the thousands of proton resonances in the ^1H -NMR spectrum of a protein or DNA molecule. This development has made NMR spectroscopy competitive with x-ray crystallography for studying the high resolution structures of small biological molecules, as well as facilitating detailed comparisons between crystal and solution forms, thus giving important insights into biochemical function.

Biomolecular complexes often function in large assemblies that are not suited to either crystallographic or NMR approaches. Neutron scattering data from biomolecular complexes in solution give information on the overall structures and relative dispositions of their components. This capability arises from the radically different neutron scattering properties of ^1H and ^2H , and depends upon obtaining neutron scattering data from a complex whose components have different mean neutron scattering densities. DNA and proteins have naturally different neutron scattering densities, and this difference can be enhanced by specific ^2H labeling which increases the quality of the structural data on the individual components.

In this project crystallographic, NMR and neutron scattering techniques were developed and applied to the study of proteins, DNA and their interactions. Specifically, we focused on studies of protein-nucleic acid complexes and DNA hairpin structures important for understanding the regulation of gene expression, as well as the fundamental interactions that allow these complexes to form. An understanding of how genetic information is translated (transcription) will be a major factor in interpreting and using the information now being obtained from the Human Genome Project, and a close interaction with the Genome project is important for this work.

3.1 Structural Studies of Protein/DNA Interactions

The nucleosome core particle is the fundamental repeating unit of chromosomes in higher organisms, and is made up of a core of protein subunits (the histone octamer) surrounded by double-stranded DNA containing 146 base pairs. The histone proteins are subject to chemical modification in the form of reversible N-acetylation of specific lysine residues located in their N-terminal tail regions. This acetylation has long been associated with transcriptional activity. Scattering techniques have been employed to study the effects of the histone tails and their acetylation on nucleosome core particles.

Neutron scattering experiments have been completed on trypsinized nucleosome core particles that revealed a reduction in the radius of gyration of the histone octamer component, supporting the idea that the histone tails contribute significantly to the radius of gyration of the

histone octamer in the intact core particle. At the same time the shape of the DNA component seems to be altered by the loss of the histone tails. Neutron and x-ray scattering experiments were performed at the National Institute of Standards and Technology (NIST) and at the Stanford Synchrotron Research Laboratory (SSRL), respectively, on nucleosomes reconstituted onto a 195-base-pair nucleosome-positioning sequence with either fully- or non-acetylated histones and a minor variant histone. The scattering curves obtained for the acetylated and non-acetylated samples indicate a significant difference in the shapes of the particles. These data indicate that the main interactions of the histone tails may be in the regions of DNA beyond the 146 base pair associated with the histone core and acetylation may modulate that interaction. Neutron experiments were also performed at NIST on nucleosomes reconstituted with a minor variant histone, H2B0, associated with human sperm chromatin. Very interestingly, there was a substantial difference in the structure of the particle with the substitution of this minor variant H2B0 histone. This has significant implications with regard to the role of the minor variant histones in chromatin function.

Identifying sequences of human DNA that are unusually regulated or especially prone to breakage will be one of the many payoffs from the Human Genome Project. The so-called "fragile X" repeat is a three-nucleotide sequence that is associated with the *fmr-1* gene and is involved in certain mental disorders. This repeat has the potential to expand, and once expanded it is fragile and can lead to DNA breakage. Once expanded it can also be methylated, which turns off the *fmr-1* gene and leads to mental disorders. We have determined the structures formed by the fragile X-repeat using multidimensional NMR techniques, and find that they form hairpin structures. These hairpin structures suggest a mechanism of how these repeats can extend during replication, leading to DNA breakage, DNA methylation, and ultimately, to mental disorders. We further showed that chromomycin-A3, a nonintercalative drug specific for G-C base pairs, only binds the Watson-Crick DNA duplex and not the hairpin structures adopted independently by the G- and C-rich strands. Therefore, this drug is capable of blocking the transition from the Watson-Crick hetero-duplex to the self-assembled hairpin structures in the fragile X-repeat. In addition, the Watson-Crick DNA duplex of the fragile X-repeat in the methylated form shows stronger binding to Chromomycin-A3 thereby offering a diagnostic tool for discriminating between the methylated and unmethylated forms of this repetitive DNA. The selective preference of this drug for the methylated form of the fragile X-repeat could be utilized for diagnosis of such repeats in the chromosomal DNA.

3.2 NMR and Stable Isotope Labeling Methods for Biomolecular Structure Analysis

Pulse sequences needed for these studies have been established and we have collected our first multidimensional NMR spectra of proteins (ubiquitin, and the muscle protein complex

troponin C/troponin I (96-115) in which the troponin C component was isotopically labeled). Isotopes are key to the full utilization of multidimensional NMR methods for structure analysis. Uniform labeling with carbon-13 and nitrogen-15 have permitted determinations of solution structure of proteins having molecular weights of up to approximately 20 KDa, but beyond that spectral crowding becomes prohibitive. In addition, the quality of the NMR structures is limited by the relatively small number of geometrical constraints available from the NMR observables. New theoretical methods for deriving the best structure from the limited number of geometrical constraints are the subject of intense study. In a parallel approach we are using specific ^2H labeling to minimize spectral crowding problems and maximize the number of geometrical constraints.

The size limitation on biological macromolecules that can be studied using NMR arises from the fact that both the proton resonance line width and the total number of protons increase in proportion with molecular weight resulting in severe spectral overlap problems for most globular proteins with molecular weights greater than 20 KDa. In addition, protons are relaxed primarily by dipole-dipole interactions with other protons which, at the long correlation times associated with large proteins, can yield broad resonances. Because dipolar relaxation is strongly dependent on the inter-proton distance ($1/r^6$), methylene protons that are close in space (1.8\AA) can broaden each other significantly. This relaxation mechanism is removed by stereoselective replacement of one of the methylene protons with deuterium. The remaining proton will be relaxed by protons that are more distant ($2.3\text{-}3.1\text{\AA}$) which will significantly sharpen the resonance. This approach is important in allowing us to study larger proteins and protein/DNA complexes using existing NMR fields, and to capitalize on the increased spectral resolution available at very high magnetic fields for which line-broadening effects increase in proportion to the field.

Geometrical constraints derived from NMR data are of two types. The nuclear Overhauser effect (NOE) can be detected between proton pairs that are 5\AA or closer in space. In addition, the \cos^2 -dependence (Karplus) of the spin-spin coupling interaction can be used to define the dihedral angle between protons attached to adjacent carbons. The dihedral angle between the alpha-proton and the beta-methylene protons is particularly important in defining the side chain conformation in sixteen of the twenty naturally occurring amino acids. Because the alpha-carbon in amino acids is a stereochemical center, the beta-methylene protons are inequivalent and give rise to separate NMR signals. However, there is no physical basis for assigning these signals, so the dihedral angle defined by the coupling constant must be ignored. Again this problem can also be solved by specific deuterium substitution of one of the methylene protons.

We have developed three approaches for stereospecific deuteration at the beta-methylene position of amino acids. The first involves the use of a chiral auxiliary derived from camphor. This auxiliary directs the stereochemical course of the reduction of a *trans*-alpha,beta-unsaturated acid. Reduction with a deuterated metal hydride reagent yields a stereospecifically deuterated acid which is easily converted to the corresponding amino acid. In addition, we are using a chiral reducing agent which is used to hydrogenate a double bond in a blocked alpha,beta-unsaturated amino acid. This approach requires synthesis of the deuterated amino acid precursor, which has been accomplished.

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.