AD_____

GRANT NUMBER DAMD17-94-J-4420

TITLE: Crystallization and Structure Determination of the Human Estrogen Receptor by X-ray Diffraction

PRINCIPAL INVESTIGATOR: Stephen C. Harrison, Ph.D.

CONTRACTING ORGANIZATION: Children's Hospital
Boston, MA 02115

REPORT DATE: October 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

19970117 116

DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Annual (15 Sep 95 - 14 Sep 96)

4. TITLE AND SUBTITLE Crystallization and Structure Determination of the Human Estrogen Receptor by X-Ray Diffraction

5. FUNDING NUMBERS
DAMD17-94-J-4420

6. AUTHOR(S)
Stephen C. Harrison, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Children's Hospital
Boston, MA 02115

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

This project is a postdoctoral training fellowship for Dr. Robert Nolte, to support X-ray crystallographic studies of cancer-related macromolecular complexes. We have prepared crystals of a complex between a six Zn-finger segment of the transcription factor TFIIIA and a 31 BP DNA site. The determination of this structure is almost complete at an intermediate resolution, with the position of the DNA and protein elements determined. A novel interaction between one of the zinc fingers and the minor groove has been observed. The higher resolution details of the structure, which will be determined in the next several months, should reveal significant new aspects of DNA recognition by Zn-finger proteins. These interactions are responsible for the regulation of many loci important for cell growth. During the final project year of this training grant, Dr. Nolte will conclude the TFIIIA studies and focus his attention on structural studies on another protein ZAP-70, a T-cell tyrosine kinase which plays a crucial role in signal transduction.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	18
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

X In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

X In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature

10/11/96
Date

Table of Contents

Introduction

The goal of this project is postdoctoral training for Dr. Robert Nolte in the area of structural biology as applied to problems in cancer and to breast cancer in particular. The report below corresponds to a revised SOW, formally requested in a submission to Dr. Patricia Modrow in a letter dated August 8, 1996. The revised SOW had been described and justified in a revised Progress Report for 1994-1995 submitted in January, 1996; approval received in June, 1996 .

The Problem: Cancer involves anomalies in cellular regulation. Two key regulatory steps are the transduction of signals from the cell surface and the control of specific gene expression. The estrogen receptor, on which the original SOW for this grant focused, participates in both these steps. It receives hormonal signals and in response activates defined genes. The recently discovered BCRA1 gene appears to encode a nuclear protein that may

regulate gene expression (Miki et al, 1994). Progress in structural studies of tyrosine-kinase mediated signaling pathways and protein/DNA complexes that regulate gene expression makes the possibility of structure-based drug discovery and development a real one, but many basic principles remain to be discovered.

Previous work. (i) A number of structural studies of protein/DNA complexes involved in transcriptional regulation have provided the guidelines of how to think about specific recognition (see, for example, Harrison, 1991 and Steitz, 1990). An important notion to emerge both from the structural studies and from studies of transcriptional regulation at many complex promoter/enhancer sites is that large multi-protein complexes bound to DNA are critical for the sort of combinatorial control seen in cells of all higher eukaryotes (and in human cells in particular). Work in our laboratory on the polymerase III transcription factor known as TFIIIA has led to crystallization of a complex between a part of the factor containing six "zinc fingers" (each finger is a distinct DNA-binding domain) and a 31 BP DNA site. This is a larger complex than any so far studied from this class of transcriptional regulatory proteins. TFIIIA is unusual among transcription factors in that it binds not only to sites that regulate gene expression but also to the RNA transcript. A similar double specificity has recently been observed for the product of the Wilms tumor suppressor gene, WT1 (Caricasole et al, 1996). (ii) Protein tyrosine kinases are responsible for many key signaling events in normal and cancer cells. The kinases related to Src have been especially widely studied, in part because of the early importance of Src in revealing the existence of oncogenes and oncogene products. We have determined the structure of the Src-homology 2 (SH2) and Src-homology 3 (SH3) domains of the Lck kinase, and we are engaged in an effort to crystallize the auto-inhibited form of the full Lck and Src proteins. An important adjunct to this effort, which could spawn a number of drug-design strategies,

will be to apply the expression approaches developed for Src/Lck to the protein that is "next" in the Lck-initiated pathway of T-cell signaling, ZAP-70.

Purpose of present work. The broad goal is to enhance our understanding of specific protein-protein and protein-DNA interactions in gene regulation and in signal transduction. In particular, we are concluding the structure determination of a complex between a portion of the DNA-binding segment from the polymerase III regulatory factor TFIIIA and its cognate site. We are also building upon our own recent success in structural studies of regulatory domains from Src-like protein kinases to establish a broader program. Thus, Dr. Nolte will first complete the TFIIIA/DNA structure determination (in collaboration with Dr. Ray Brown of this laboratory), a task that should be completed in the next few months. He will then focus his efforts on expanding the preliminary ZAP-70 studies The cancer relevance of these studies is described in the body of the report.

Methods. All of the projects described here involve expression of proteins, or of specific protein domains, in order to crystallize them for investigation by X-ray diffraction. X-ray methods are the best way to determine the detailed atomic structure of a protein, a necessary pre-requisite for molecular analysis of function and drug-targeting strategies.

Body of Narrative

TFIIIA:

The transcription factor IIIA (TFIIIA) from Xenopus laevis binds both DNA and RNA. The discovery of multiple zinc finger motifs in TFIIIA pointed to a complex mode of protein-DNA interaction. Each of these motifs is an independently folded protein domain that recognizes a specific DNA sequence of three base pairs. Nine tandem zinc fingers are present in TFIIIA, allowing the protein to bind to an extensive region of the 5S RNA gene promoter.

The first six zinc fingers of TFIIIA have been expressed in E.coli and reconstituted with a duplex of two 31 base-pair synthetic DNA strands (shown in **Figure 1a,b)** whose sequence derives from the Xenopus oocyte 5S RNA gene promoter . Crystals of the complex were grown using "native" DNA and one of two different substituted DNAs, in which three thymidine nucleotides were modified to contain iodine at specific locations. Six complete data-sets were collected on two trips to the National Synchrotron Light Source at Brookhaven Laboratory during the last year and the diffraction statistics of these data sets are summarized in **Table 1.** These data-sets were collected from either native or derivative crystals with the data in some cases extending to 3.15Å. Anisotropy in the data limits the effective resolution to approximately 3.5Å. Difference Patterson maps have been solved, and the position of the iodine atoms determined. From these iodine positions, low resolution phases have been calculated, and Dr. Nolte has been working to extend these phases to the resolution limit of data. The initial and current phasing statistics are summarized in **Table 2.** The current phases allow the calculation of an electron density map in which the DNA can clearly be seen. Dr. Nolte has built the 31 base pairs of DNA into the density as illustrated in **Figure 2.** Density is also visible for the protein, as shown in **Figure 3,** and he has tentatively built a Cα trace of 183 of the 190 residues contained in this construct. An overview of the current structure at approximately 4.5Å resolution shows the arrangement of the 6 fingers on the DNA (**Figure 4**). A novel interaction between one of the fingers and the minor groove of the DNA is apparent. This interaction may give us insight into the mechanism by which TFIIIA is able to bind to DNA. The molecular details describing the interaction between individual atoms of the protein and DNA will be available in the next few months as the higher resolution phases are determined and the final model is refined.

Crystals will be grown for an additional data collection trip to Brookhaven, which will occur in late October 1996, with the intention of extending the

resolution of this structure so that the finer details of the DNA / protein interaction can be determined.

Many genes with key roles in cancer are regulated by zinc-finger containing proteins. Understanding the principles of DNA recognition by members of this family will be important for establishing a picture of their transcriptional control. In particular, the feasibility of efforts to find drugs that target gene-specific transcriptional regulation will depend on such understanding. The TFIIIA structure, when completed, will contain the longest strand of DNA and more zinc fingers than has been observed in any other structure to date. It will greatly expand the database of interactions, and it has already suggested a novel binding interaction to the minor groove of DNA, not previously observed.

Src-family kinases.

We have determined the structures of the regulatory domains of the lymphocyte tyrosine kinase, Lck (Eck, et al, 1993;Eck, et al, 1994), and members of the laboratory are currently engaged in efforts to crystallize the auto-inhibited state of the full-length, multi-domain molecule of Lck and Src. Src-like kinases of this sort are important in a variety of tumors, and they appear to modify proteins that have a role in breast cancer (see, for example, Arnold et al, 1995). Upon completion of the TFIIIA project, Dr. Nolte will build upon the work on Lck/Src kinases by examining the structure of the immediate downstream target of Lck - another tyrosine kinase known as ZAP-70, which contains two SH2 domains and a kinase domain. Preliminary studies involving the expression and purification of ZAP-70 have been carried out, and a small amount of protein has been purified to demonstrate the feasibility of this project. ZAP-70 kinase was expressed and purified using an Sf-9/Baculovirus expression system as follows: A fragment encoding residues 1 to 619 of human ZAP-70 tyrosine kinase with an N-terminal polyhistidine tag was subcloned into the pVL1392 and pBlueBac4 (Invitrogen) transfer vectors (Pst I and EcoRI restriction sites) using standard PCR-based

methods. Recombinant Baculovirus was obtained by cotransforming Sf-9 cells with linearized viral DNA (Baculogold, Pharmingen) and the transfer vector. High-titer viral stocks (~109) were prepared from plaque-purified recombinant virus. For protein production, Sf-9 cells were cultured in 2 L Spinner flasks in TNM/FH media supplemented with 10 % serum. Cells were infected with 25 ml/liter viral stocks and harvested 64-72 hours post-infection by centrifugation. Cells were lysed by sonication and the lysates cleared by high speed centrifugation. The resulting supernatant was loaded onto a Nickel-NTA column, washed and eluted with 200mM imidazol. The ZAP-70 containing fractions were further purified by ion exchange on SP-Sepharose column, followed by gel filtration on Superdex-200 column to achieve > 95% pure protein. Fractions containing ZAP-70 were identified by SDS-PAGE and Western Blotting (anti-human ZAP-70) during the purification.

Conclusion:

The crystallographic study of a complex between (TFIIIA fingers 1-6) and cognate DNA will, when completed, yield significant new information about modes of specific DNA recognition by this family of transcription factors. Other work in our laboratory on gene regulation at complex promoter/enhancers will then be poised to capitalize on the results. Dr. Nolte will himself complete this period of training and prepare to launch independent directions, by turning his attention to structural analysis of signal transduction pathways. Parallel efforts in our laboratory have been successful at expression and crystallization of domains of the Lck and Src kinases, and we will expand this effort to include the downstream "effector" of Lck, ZAP-70.

References

Arnold, S.F., Obourn, J.D., Jaffe, H., and Notides. AC. "Phosphorylation of the human estrogen receptor on tyrosine 537 in vivo and by src family kinases in vitro" Molecular Endocrinology 9, 24-33 (1995)

Caricasole, A. et al "RNA binding by the Wilms tumor suppressor zinc finger proteins" Proceedings of the National Academy of Sciences, USA 93: 7562-7566 (1996).

Cowtan, K. "DM: an automated procedure for phase improvement by density modification" Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 31, 34-38 (1994).

Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography Acta Cryst. D50, 760-776 (1994).

Eck, M.J., Atwell, S.K., Shoelson, S.E. & Harrison, S.C. "Structure of the regulatory domains of the Src-family tyrosine kinase p56lck" Nature, 368, 764-769 (1994).

Eck, M.J. Shoelson, S.E., Harrison, S.C. "Recognition of a high affinity phosphotyrosyl peptide by the Src homology domain of p56lck" Nature, 362, 87-91 (1993).

Harrison, S. C. "A structural taxonomy of DNA-binding domains" Nature, 353, 715-719 (1991)

Miki, Y. , Swensen, J., Shattuck-Eidens, D., Futreal,P.A., Harshman,K., Tavtigian,S. et al. 1994, "A strong candidate for the breast and ovarian cancer susceptibility gene" Science 266, 66-71 (1994)

Otwinowski, Z. "Maximun likelihood refinement of heavy atom parameters" Daresbury Study Weekend Preceedings (1991).

Steitz, T. A., "Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding" Quarterly Reviews of Biophysics, 23, 205-280 (1990)

FIGURE 1

A) Sequence of the two 31 base pair oligonucleotides contained in the TFIIIA crystal structure

```
1                 10                  20                  31
|                  |                   |                   |
A C G G G C C T G G T T A G T A C C T C G A T G G G A G A C C

  G C C C G G A C C A A T C A T G G A C C T A C C C T C T G G T
  |     |                 |                   |             |
  63    60                50                  40            33
```

B) Sequence of the N-terminal 190 Amino acids contained in the TFIIIA crystal structure

N-Terminal Sequence :

1- M G E K A L P V V Y K R -12

Finger 1 :

13- Y I C S F A D C G A A Y N K N W K L Q A H L C K H T G E K P

Finger 2 :

43- F P C K E E G C E K G F T S L H H L T R H S L T H T G E K N

Finger 3 :

73- F T C D S D G C D L R F T T K A N M K K H F N R F H N I K I C V

Finger 4 :

105- Y V C H R E N C G K A F K K H N Q L K V H Q F S H T Q Q L P

Finger 5 :

135- Y E C P H E G C D K R F S L P S R L K R H E K V H A G

Finger 6 :

162- Y P C K K D D S C S F V G K T W T L Y K H V A E C H Q D

Figure 2
Electron Density Map of TFIIIA : DNA Structure

Figure 3
Electron Density of TFIIIA : Protein Structure

Figure 4
TFIIIA Structure : DNA / Protein Organization

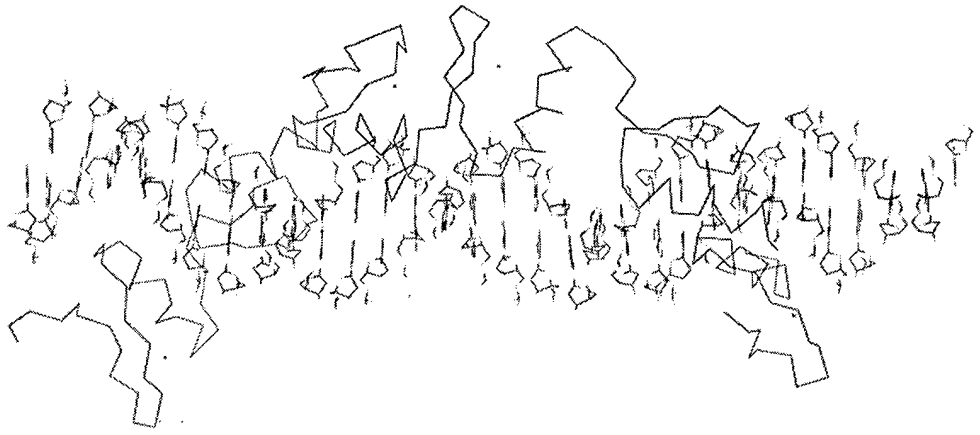

Table 1 : Summary of X-ray Diffraction Statistics for TFIIIA Data Sets

Name	N23	N13	N1	RB26	RB27	RBI3	RB57
Derivative	Native	Native	Native	BCF	BCF	BCE	BCE
Wavelength (Å)	1.283	1.283	1.542	1.283	1.283	1.542	1.283
Resolution (High)	3.10 Å	3.16Å	3.10Å	3.15Å	3.20Å	3.20Å	3.20Å
R-Factor (High)	5.1%	14.2%	10.8%	10.3%	9.2%	8.4%	8.1%
Resolution (Nominal)	3.20Å	3.50Å	3.65Å	3.46Å	3.49Å	3.55Å	3.60Å
R-Factor (Nominal)	4.97%	10.9%	5.9%	7.0%	8.8%	9.4%	7.6%
I / Sigma (Average)	15.3	8.0	10.7	14.4	9.5	12.2	10.1
Completeness	92.0%	59.5%	97.0%	83.3%	92.8%	94.0%	86.3%
Frames Collected	157	140	333	210	216	293	140
Frame Width (deg)	1.25	1.50	1.0	1.5	2.5	1.0	1.5
Dataset Multiplicity	2.04	1.69	3.06	3.01	2.6	2.17	1.82

Table 2 : TFIIIA Phasing Statistics

	Initial MIR Phases	Current Density Modified Phases
Resolution	20-4.0 Å	25-3.1 Å
Figure of Merit	0.390	0.645
Phasing Power	1.61	-
Cullis R-Factor	0.84	-

Note :

Initial phases : Calculated using the MLPHARE program in the CCP4 package (Otwinowski, 1991)

Current Phases : Calculated with the DM program (Cowtan, 1994)