Structural and Computational Studies of Two Oligonucleotide Modifying Enzymes: I-*Ppo*I and T4 Polynucleotide Kinase

Eric A. Galburt

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2002

Program Authorized to Offer Degree: Department of Biochemistry

UMI Number: 3062944

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Eric A. Gallant

Date 8/14/02

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Eric A. Galburt

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Co-Chairs of Supervisory Committee:

Barry L. Stoddard

William Parson
William W. Parson

Reading Committee:

Barry L. Stoddard

Roland K. Strong

William Parson
William W. Parson

Date: August 9, 2002

University of Washington

Abstract

Structural and Computational Studies of Two Oligonucleotide Modifying Enzymes: I-*Ppo*I and T4 Polynucleotide Kinase

Eric A. Galburt

Co-Chairpersons of the Supervisory Committee:
Affiliate Professor Barry L. Stoddard,
Department of Biochemistry, Member, Fred Hutchinson Cancer Research Center
Professor William W. Parson, Department of Biochemistry

Endonucleases are a large class of enzymes that catalyze the formation of site-specific double strand breaks in DNA. Homing endonucleases are expressed from open reading frames within intervening sequences (introns or inteins) and promote the lateral transfer of their host intron to a cognate intronless allele. A novel mechanism of DNA endonucleolytic cleavage, exhibited by the homing endonuclease I-*Ppo*I and characterized by a histidine general base, has been visualized at high resolution by intermediate trapping and studied computationally. The structures of three separate catalytic species were determined by X-ray crystallography and are presented here. The nudged elastic band method was used to calculate the minimum energy path for the catalyzed reaction and the predicted rate is compared to experiment. In addition, the structure of I-*Ppo*I bound to its DNA target site shows that bound DNA is severely bent, resulting in significant deformations of the minor and major grooves near the scissile phosphates. To study the role of conformational changes within the protein catalyst and the DNA substrate, the structures of both the apo enzyme and the L116A mutant protein/DNA complex with severely decreased binding affinity were determined. The wild-type protein displays a rigid-body rotation of its individual subunits upon DNA binding. Homing site DNA is not detectably bent in the absence of protein, but is sharply bent in both the WT and L116A complexes, indicating that binding involves a large distortion of the DNA and a smaller change in protein conformation. The critical function of leucine 116 is also described.

T4 phage polynucleotide kinase (PNK) was identified over 35 years ago and has become a staple reagent for molecular biologists. The enzyme displays 5'-hydroxyl kinase, 3'-phosphatase, and 2',3'-cyclic phosphodiesterase activities against a wide range of substrates. These activities modify the ends of a nicked tRNA generated by a bacterial response to infection and facilitate their repair by T4 RNA ligase. DNA repair enzymes that share conserved motifs with PNK have been identified in eukaryotes. PNK contains two functionally distinct structural domains and forms a homotetramer. The C-terminal phosphatase domain is homologous to the L-2 haloacid dehalogenase family and the N-terminal kinase domain is homologous to adenylate kinase. The active sites have been characterized through structural homology analyses and visualization of bound substrate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3'-AMP	adenosine 3'-monophosphate
Å	angstroms
ADP	adenosine diphosphate
ADK	adenylate kinase
ALS	advanced light source
AMoRe	automated molecular replacement
ATP	adenosine triphosphate
bHLH	basic helix-loop-helix
bp	base pair
BSA	bovine serum albumin
CAP	catabolite activating protein
CAPS	3-cyclohexylamino-1-propane sulfonic acid
CHES	2-(N-cyclohexylamino) ethanesulfonic acid
CCD	charged coupled device
C-terminal	carboxyl-terminal
CNS	crystallography and NMR system
CV	column volume
DFT	density functional theory
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
ds	double-stranded
DSB	double strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EP	enzyme-product
EPMR	evolutionary program for molecular replacement
ES	enzyme-substrate
EVB	empirical valence bond

FEL	free electron laser
GMS	gel mobility shift
GTP	guanine triphosphate
HAD	L-2-haloacid dehalogenase
HF	Hartree-Fock
hPNK	human polynucleotide kinase
IPTG	isopropyl-thio-β-D-galactosidase
kDa	kilo Daltdons
keV	kilo electron volts
KIE	kinetic isotope effect
MAD	multiple anomalous dispersion
MEP	minimum energy path
MES	2-(4-morpholino)-ethane sulfonic acid
MM	molecular mechanics
N-terminal	amino-terminal
NEB	nudged elastic band
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PFK/FBP	6-phosphofructo-2-kinase/ fructose-2,6-bisphosphatase
PNK	polynucleotide kinase
PSP	phosphoserine phosphatase
QM	quantum mechanics
rmsd	root mean square deviation
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RT	room temperature
SAD	single anomalous dispersion
SDS	sodium dodecyl sulfate

ss	single-stranded
TBP	TATA-box binding protein
Tris	tris hydroxymethyl aminomethane
tRNA	transfer ribonucleic acid
WT	wild-type

ACKNOWLEDGEMENTS

There are many people who need to be thanked for helping me during my graduate school career...

Barry Stoddard – For guidance; scientific and otherwise
Roland Strong, William Parson, Hannes Jonsson – For helpful advice and ideas

Rachel Friedman – For strength through good and bad
My Family – For the unconditional support of all my endeavors

Jill Bolduc, Chris Bond, Brett Chevalier, Dave Erho, Greg Ireton, Melissa Jurica, Betty Shen, Clint Spiegel, Django Sussman – For being easily the best colleagues one could hope for both as scientists and friends

Meg Chadsey, Graeme Hennkelman, Ray Monnat, John Pelletier, Jakyoung Song, Geoff Wilson – For fruitful collaborations and intellectual stimulation

DEDICATION

To my parents

Chapter 1

X-ray Crystallography

Introduction

Over the past century, the science of X-ray crystallography has progressed from the serendipitous discovery of X-rays in 1895 to the routine structure determination of complex macromolecules. A useful timeline of the development of X-ray crystallography is provided by a selection of Nobel prizes over the last 100 years: the discovery of X-rays (W. C. Röntgen, 1905); X-ray diffraction by crystals (M. Von Laue, 1914); X-ray crystal structure determination (W. H. and W. L. Bragg, 1915), the crystallization of enzymes (J. B. Sumner, 1946); the structures of the globular proteins haemoglobin and myoglobin (J. C. Kendrew and M. Perutz, 1962); the structure of Vitamin B12 (D. Hodgkin, 1964); the development of direct methods for the determination of crystal structures (H. Hauptmann and J. Karle, 1985); the three-dimensional structure of the photosynthetic reaction center (J. Deisenhofer, R. Huber, and H. Michel, 1988); the mitochondrial F_1-ATPase structure (J. Walker, 1997). X-ray crystallographic determination of protein structures has made it possible to study biology at atomic resolution and the rate of structural determinations of proteins has been increasing rapidly over the last 2 decades.

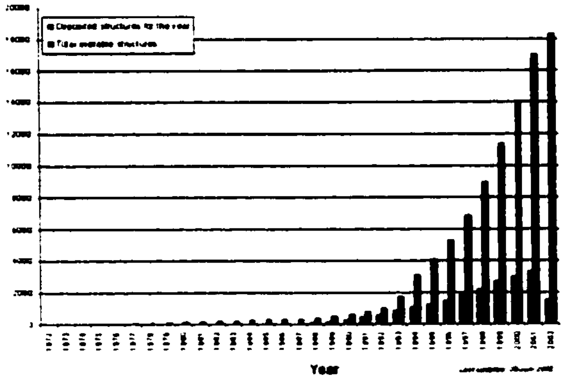


Figure 1.1: The growth of the Protein Data Bank (PDB) **(Berman et al., 2000)**; http://www.pdb.org

In the following sections, a brief overview of the method is provided followed by a review of current attempts to use crystallography in a time-resolved manner to elucidate the structures of macromolecules in different functional states.

The two main components of the X-ray crystallographic experiment are, quite obviously, X-rays and crystals. X-rays are electromagnetic waves with wavelengths on the order of the separation of atoms in molecules. They interact with the electron clouds of atoms and are scattered either elastically or inelastically depending on the precise wavelength of the X-ray and the nature of the atoms involved. Protein crystals are grown in a variety of ways and consist of repeating blocks known as unit cells. The unit cell is made up of a number of asymmetric units that are related to one another by a series of symmetry operations that define the space group of that particular crystal. The combination of the repeating molecular units of the crystal and X-rays (Figure 1.2) results in the phenomenon of lattice diffraction (Figure 1.3).

The diffraction pattern that results when X-rays are shot through a protein crystal can be analyzed through a variety of techniques to back-calculate the structure within the crystal. These techniques are all based on the fundamental relationship between the diffraction pattern (reciprocal space) and the electron density (real space) within the crystal. The mathematical formalism of the Fourier transform describes this fundamental relationship and is used to go between reciprocal and real spaces (Figure 1.4). Just as real space consists of an electron density at each point in space, reciprocal space consists of the amplitudes of the reflections that are measured directly in the diffraction experiment and their associated phases. The fact that the phases of the reflections cannot currently be measured leads to the well known "phase problem" and requires the application of additional techniques such as the F^2-series, or Patterson function. The Patterson function calculates a map of density-weighted interatomic vectors when only reflection amplitudes are known (Figure 1.5). Combined with the scattering of incorporated heavy atoms (isomorphous replacement) or anomalously (inelastic) scattering atoms (anomalous dispersion), the Patterson function and Fourier transform allow for the calculation of the real space structure from measurements of reflection intensities only.

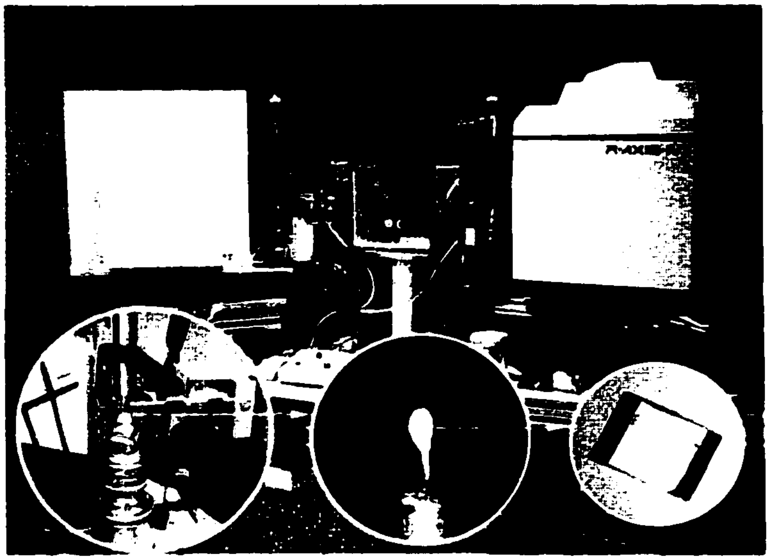


Figure 1.2: Typical instrumentation used for measuring X-ray diffraction of macromolecular crystals. The large image shows two X-ray detectors and the inset images are progressive close-ups of the setup. The left image shows the goniometer on which the crystal is positioned. The X-ray collimator (horizontal, pointing left), cooling nozzle (verical, pointing down), and centering microscope (horizontal, pointing back) can all be seen. The black piece of metal affixed to the collimator is the beam stop. The crystal is harvested from its growth environment with the fiber loop shown in the center image and plunged into liquid nitrogen for rapid cooling. The loop is then placed on the goniometer and positioned so that the crystal is centered in the X-ray beam. The rightmost image shows a typical protein crystal (B. Chevalier).

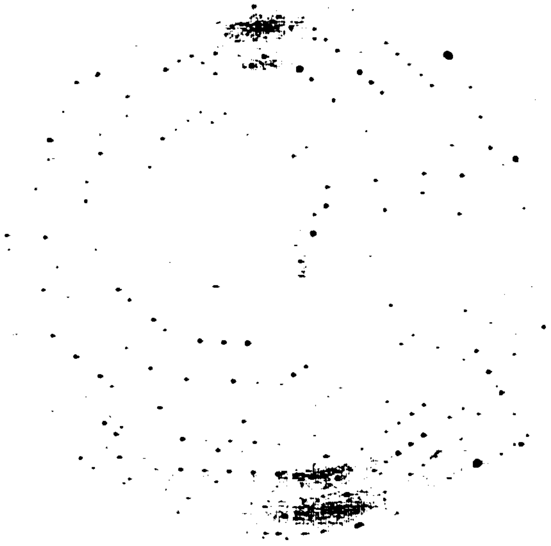


Figure 1.3: A typical protein crystal diffraction pattern

$$\text{(a)}\quad \vec{F}(hkl) = V \int_{x=0}^{1}\int_{y=0}^{1}\int_{z=0}^{1} \rho(xyz)\cdot e^{2\pi i(hx+ky+lz)}\,dxdydz$$

$$\text{(b)}\quad \rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F(hkl)\right|\cdot e^{-2\pi i(hx+ky+lz)+i\phi(hkl)}$$

Figure 1.4: The Fourier transform relating (a) the amplitude of reciprocal space reflections, F(hkl), to real space electron density, ρ(xyz), and (b) vice-versa. V is the volume of the unit cell, ϕ(hkl) is the phase of the hklth reflection, hkl are the indices of individual reciprocal space vectors, and xyz are the real space coordinates.

$$\int_{x=0}^{1}\int_{y=0}^{1}\int_{z=0}^{1} \rho(xyz)\cdot \rho(x+u, y+v, z+w)\,dxdydz =$$

$$\frac{2}{V}\sum_{h}\sum_{k}\sum_{l}\left|F(hkl)\right|^{2}\cos\left\{2\pi(hu+kv+lw)\right\}$$

Figure 1.5: The Patterson function (below) calculates a map of density-weighted interatomic vectors (uvw) (above) from measured amplitudes of reflections, F(hkl), and the volume of the unit cell, V.

Time-Resolved Macromolecular Crystallography

The processes of life that take place within a cell are all dependent on an extensive array of chemical transformations. Sugar must be metabolized into useful chemical energy stores, raw materials acquired from the environment must be converted into a wide array of different chemical moieties, and the functions of the cell must be carried out. To regulate these chemical transitions and to speed up reactions that are intrinsically slow, every cell has an arsenal of protein molecules with specific functions. These large molecules (1000's of atoms) are incredible catalysts, in some instances increasing reaction rates by as much as 17 orders of magnitude. Protein molecules are hetero-polymers consisting of different sequences of 20 different amino acid building blocks that fold up into a well defined three dimensional structure. X-ray crystallography allows for the determination of their structure, but knowing the structure does not necessarily reveal how the protein functions. One would like to have structures of the protein at different points along the path of the reaction that it catalyzes, for example the protein in complex with reactants, with products, and with any chemical intermediates that occur in between. Time-resolved crystallography experiments address these questions. In addition to increasing our knowledge of the physical chemistry of catalysis, there are many possible practical applications of these studies. By learning how proteins from disease-causing organisms function, researchers can learn how they might perturb these functions to help treat illness. Furthermore, by understanding the mechanisms of natural proteins, one might be able to design novel proteins that catalyze new transformations that could make industrial processes more efficient (Walsh, 2001).

The techniques of X-ray crystallography allow investigators to build a detailed molecular model corresponding to the electron density, which results in a single, static image of the molecule. As mentioned earlier, proteins often proceed through a series of intermediate structural states as they perform their biochemical functions. These functions include the binding of specific molecules, the catalysis of specific chemical reactions, and structural changes in response to excitation by light. A number of techniques, collectively referred to as "time-resolved crystallography", can be used to determine the structures of these discrete intermediates which lead to a better understanding of the chemistry that underlies the protein's function.

Reactions are possible within many macromolecular crystals because the crystals have a significant water content. Approximately 40% to 60% of the volume within most protein and nucleic acid crystals is composed of solvent; this water is found in large channels and ordered hydration shells around the macromolecules. As a result, small reactant molecules (usually less than 100 atoms in size) can diffuse throughout a crystal of an enzyme (a protein that catalyzes a specific chemical reaction), bind to its active site (the reactive center of the enzyme), and be turned over (undergo a catalyzed transformation). Similarly, non-catalytic processes (such as binding reactions, conformational changes, and electron transfer steps) can be induced in crystals as well.

Reactions in the crystalline environment may exhibit kinetic rates similar to those measured in solution, or may be significantly slower due to restraints imposed by the crystal lattice. In either case, it is often possible to accumulate a rate-limited intermediate and to collect X-ray diffraction data during the lifetime of that species. A key reason that such experiments are possible is that efficient methods have been developed over the past 15 years for collecting X-ray diffraction data at almost any speed. When data are collected using the polychromatic "Laue" diffraction method at a modern synchrotron X-ray source, X-ray exposure times can be as short as 10 to 100 picoseconds and still provide useful signal-to-noise ratios.

For time-resolved experiments, the investigator must address three critical parameters. A method is chosen to initiate the reaction that allows a specific intermediate to accumulate throughout the crystal. If the desired intermediate is not inherently rate-limited or its half-life is too short to allow its visualization, then a strategy is employed to stabilize or trap that intermediate. Finally, a method of data collection is chosen that is sufficiently rapid to match the lifetime of both the transient reaction intermediate and of the crystal specimen itself. At the fastest end of the time-resolved experimental spectrum is a "single-turnover" experiment, in which all the data are collected during a single reaction event. This is usually done either because the reaction can only occur once (i.e. the reaction does not cycle repeatedly and is irreversible) or because the intermediate only accumulates uniformly throughout the crystal during an initial reaction event. In such experiments, the reaction is usually triggered by a photolytic event, and the resulting accumulation and decay of the desired reaction intermediate species occurs quickly (nanoseconds to milliseconds). This necessitates rapid data collection, often using the polychromatic Laue diffraction method as mentioned above. At the slowest experimental extreme, reactions can be initiated by slow diffusion of substrate into the crystal. In this type of protocol the intermediate is often trapped by using chemical and physical conditions such as reduced temperature or a non-optimal solvent or pH for the reaction. Such experiments are appropriate for slow reactions and long-lived intermediates. The exact experimental strategies that are employed for time-resolved crystallographic experiments depend on the kinetics of the reaction, the conformational changes exhibited by the molecule during the reaction, and the size, symmetry, stability, and solvent content of the crystals.

Reaction initiation

The concept of conducting time-resolved crystallographic studies relies on the possibility that a reaction can be triggered and driven throughout the population of molecules in such a manner that a unique reactive species accumulates in the crystal. In practice, the process of initiating a reaction is equivalent to adding an additional step in the reaction, with rate-constant k_{start}, that converts a non-reactive precursor complex (A^*) into a fully productive reaction complex (A). That complex can then proceed to convert substrate (A) into product (P), often by way of one or more transient intermediates (I_n):

$$\mathbf{A^* \rightarrow A \rightarrow I_n \rightarrow P}$$

The faster the reaction and the shorter the half-life of the targeted intermediate, the faster must be the triggering rate for the experiment.

The simplest method of reaction initiation is to diffuse a reactant species into the crystal. This method can be used either to trigger a single reaction event, or to drive a reaction cycle continuously. The former strategy (collecting all the diffraction data during a single reaction event) has been used by several groups to study a wide variety of molecular reactions. For example, Bill Scott's laboratory at the University of Santa Cruz has determined the mechanism of self-cleavage for the hammerhead ribozyme (Scott and Murray, 2000) and Janos Hajdu's group at Uppsala University has determined the structures of intermediates formed during the reactions catalyzed by enzymes that oxidize and reduce small molecule substrates such as cytochrome C peroxidase and catalase (Fulop et al., 1994; Gouet et al., 1996). The latter strategy, which is often called a "steady-state" experiment, has been used by a large number of other investigators. For example, our group at the Fred Hutchinson Cancer Research Center has visualized catalytic intermediates formed by isocitrate dehydrogenase during steady-state turnover; similar studies have been reported on a variety of enzymes. In an especially impressive recent study, Ilme Schlichting at the Max Planck Institute and her collaborators determined the structures of a series of intermediates formed by the cytochrome P450cam protein as it catalyzes the oxygenation of camphor (Schlichting et al., 2000).

Regardless of the exact nature of the experiment, diffusion is far too slow for the uniform accumulation of anything other than exceptionally long-lived intermediates. During a single reaction event, accumulation of a unique rate-limited species throughout the crystal can be achieved only for systems with overall rates less than 10^{-2} reactions per second per molecule. In a steady-state experiment, the reaction is driven in the crystal using a continuous flow of a high concentration of reactants that guarantees that the majority of the reaction site in the crystal are continuously engaged in the reaction. The reaction rate that can be successfully accommodated in that type of experiment is somewhat higher, up to approximately one reaction per second per molecule.

Many biochemical reactions must be initiated using faster methods than diffusion. The use of photoreactive (caged) compounds has enabled investigators to trigger the rapid release of biologically important effectors. An increasingly common use of such compounds is the release of a reactant within a macromolecular crystal for time-resolved studies. In such experiments, the X-ray exposure is timed to match the accumulation and lifetime of a transient, rate-limited species during the initial reaction event. In recent years, extraordinarily efficient photoreactive caging groups have been created that can be incorporated onto hydroxyl, carboxyl, phosphoryl, and amide groups of a variety of compounds. Almost all are capable of sub-microsecond product release.

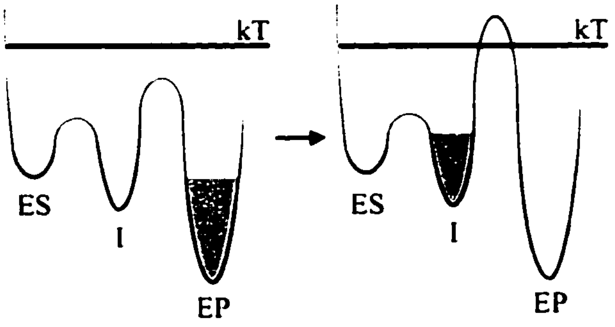
Extremely rapid X-ray studies that rely on photolysis of a reactive chemical moiety can also be conducted on intermediates formed by photoreaction centers. These molecules possess the inherent advantage of having the photoexcitable molecule (a reactive chromophore) bound to the protein or nucleic acid at virtually 100% occupancy prior to reaction initiation. These studies are greatly facilitated by that molecule's photochemical efficiency and by the characteristic spectroscopic signals often exhibited by each reaction intermediate. This can allow nanosecond time resolution of early events in these reactions, far faster than the time resolution that can be achieved for any system that requires a binding event after the initial triggering step. Detailed time-resolved studies have been reported by Elizabeth Getzoff's group at Scripps Institute and Keith Moffat's group at the University of Chicago for the reaction exhibited by the photoreactive yellow protein (PYP) (Brudler et al., 2001; Ren et al., 2001). These studies have provided snapshots of several intermediates formed by this protein and its associated chromophore during their photocycle reaction. Similarly, a number of investigators, including Joel Berendzen at Los Alamos, the Schlichting group, and the Moffat group have visualized the photo-dissociation and rebinding of carbon monoxide to myoglobin using photolytic reaction initiation and a variety of methods to trap the reaction and collect X-ray data (Teng et al., 1997; Chu et al., 2000; Schlichting and Chu, 2000).

Time-resolved diffraction experiments using photolysis to trigger the reaction have also been employed on a variety of systems, particularly membrane-bound receptor proteins, using electron crystallography. Studies by Richard Henderson at the Medical Research Council on bacteriorhodopsin have allowed the direct visualization of transient conformational and reaction intermediates (Subramaniam et al., 1999). These experiments are greatly facilitated by the nature of 2-D electron diffraction experiments, in which the protein is arrayed as a crystalline monolayer across a sample grid. This allows rapid and efficient reaction triggering either by flash-photolysis (bacteriorhodopsin) or by spray-mixing of ligand on the receptor molecules, followed by flash-cooling of the sample. Because the sample thickness corresponds to a single unit cell width, the efficiencies both of reaction initation and of cooling and trapping are higher than for three-dimensional crystals.

Stabilization and trapping of intermediates

The techniques used to extend the lifetime of an intermediate species fall into two broad, overlapping categories: physical and chemical trapping (Figure 1.6). Both methods can be used to reduce the rate of one or more individual steps in a biological reaction and to extend the lifetime of a reaction intermediate for X-ray diffraction analysis. It is always desirable to maximize the half-life of the intermediate of interest in time-resolved experiments so that the simplest method of X-ray data collection can be used, and to increase the signal-to-noise ratio through longer exposure times.

(a) chemical trapping



(b) physical trapping

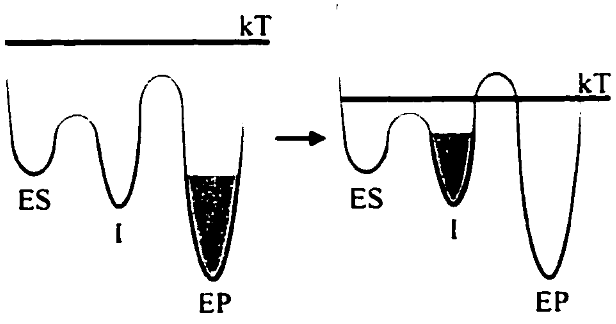


Figure 1.6: (a) Chemical and (b) physical trapping illustrated using hypothetical free-energy diagrams. In chemical trapping, the chemical environment is manipulated such that the energy barrier between the intermediate (I) and product (EP) increases. In physical trapping, the barrier remains constant, but the thermal energy (kT) is lowered so that the system is trapped once is progresses from substrate (ES) to intermediate.

Physical trapping

The use of low temperature has become increasingly common for time-resolved studies. By varying reaction conditions, cooling rates and final temperatures, it is possible, in separate experiments, to trap species that represent different, sometimes closely related reaction intermediates. For systems that display well-characterized reaction pathways, with several intermediates that predominate at different temperatures, the exploitation of unique trapping protocols allows crystallographers to perform detailed mechanistic studies. The rate constant for a single kinetic step in a reaction varies with the temperature and depends on the activation energy for the step. For a reaction step with an activation energy of 4 kcal/mol, a reduction in temperature from 20° C to -80° C decreases the rate of that step by a factor of more than 30. For an energy barrier of 10 kcal/mol, the rate constant is reduced by over three orders of magnitude. Hence, by judicious lowering of the temperature of the crystal during turnover, an investigator should be able to trap and observe the structures of intermediate states.

Before cooling, crystals are generally transferred to mixed solvent systems (cryosolvents) with a high viscosity and depressed freezing point to avoid destruction of the sample by the formation of ice. However, several difficulties can arise at very low temperatures in crystallographic studies of enzyme-ligand complexes. The physical constants of various chemical groups in enzyme catalysts show a large temperature-dependence, including the pK_a's of ionizable groups that are often involved as critical residues in catalysis. The magnitude of the shift in pK_a as a function of temperature is itself dependent on the enthalpy of ionization for the group. This implies that the catalytic mechanism may be altered in an unexpected manner by cooling. In addition, the dielectric constant of the reaction medium can be affected by the change in temperature, as well as by the change in medium composition when cryosolvents are used.

Methods of flash-freezing, now commonly used to stabilize crystals for data collection, can also be applied for intermediate trapping experiments. In theory, the reaction profile is less susceptible to the problems discussed above since the reaction proceeds at ambient temperature prior to extremely rapid cooling. Only the presence of cryoprotective agents (glycerol, sucrose, etc.) are possible confounding factors in the kinetic and chemical mechanism of the reaction. It should be possible to combine photolytic reaction triggering and subsequent flash-freezing for time-resolved experiments.

Chemical trapping

A different strategy also may be considered, either separately or in conjunction with low temperature. The relative occupancy of a specific catalytic intermediate may be elevated, and its structure determined, by adjusting the chemical nature of reaction components so that a particular intermediate accumulates. In essence, the kinetic profile of the catalytic reaction is altered to impose a new rate-limit or a thermodynamic dead-end. Such techniques can be used to isolate an intermediate within the context of a single-turnover experiment, or as a steady-state complex. These experiments may include a significant change in the pH of the reaction or a perturbed solvent environment. Alternatively, enzymes that catalyze single-substrate/single-product reactions (or that proceed through separable half-reactions) may be studied under conditions of thermodynamic equilibrium that favor a single predominant catalytic species. Additional methods of chemical trapping are the use of site-directed mutagenesis or chemical modification of the substrate or a cofactor to extend the lifetime of a specific catalytic intermediate.

Extremely rapid X-ray studies

Currently, the time resolution afforded by rapid diffraction methods is as fast as 10 to 100 ps. However, recent research has begun to address the possibility of examining reaction processes at time resolutions three orders of magnitude faster. At these time scales, atomic vibrations can be resolved and studied. X-ray pulses as short as 100 femtoseconds have been generated at the Advanced Light Source in Berkeley that in the future could allow researchers to study how atomic motions contribute to various biochemical reactions. Recent developments in free-electron lasers (FEL) could also revolutionize both the time scales and required geometry in which macromolecular structure data is obtained. Janos Hajdu and co-workers have studied the way that single-molecule diffraction could be used with short, high flux X-ray bursts from an FEL to generate structural information (Hajdu, 2000). Not insignificantly, this would remove the requirement of crystallizing a sample of interest. In contrast to lattice diffraction, single molecule diffraction is results in a continuous diffraction pattern that must be analyzed with a different set of mathematical tools (Neutze et al., 2000). It would also provide a molecular snapshot along with important information regarding the ultrafast structural dynamics that are due to atomic vibrations and that likely affect reaction rates. There are still major experimental hurdles to overcome, such as how to position the sample in the beam, and to prevent the intense radiation from destroying the sample. However, in the future, these experimental techniques could greatly increase the range of time scales available to crystallographers interested in how macromolecular systems change in time.

Chapter 2

Catalytic Mechanisms of Restriction and Homing Endonucleases

Introduction

Enzymes have long been appreciated for their ability to catalyze specific chemical transformations. One of the most important chemical transformations catalyzed in nature is phosphoryl transfer. "Phosphate esters and anhydrides dominate the living world" (Westheimer, 1987) as phosphoryl transfer plays crucial roles in most basic cellular functions including metabolism, signal transduction, DNA replication, and transcription. Phosphoryl transfer reactions also allow for the manipulation of nucleic acids and form the basis for modern molecular biology. Kinases catalyze the transfer of the γ-phosphate from ATP to a variety of molecules, while phosphatases catalyze the removal of phosphoryl groups. Mutases transfer phosphoryl groups between two atoms on a single molecule. Polymerases catalyze the extension of nucleic acid polymers by transferring the α-phosphate of nucleotide triphosphates from pyrophosphate to the terminal 3' hydroxyl group, and endonucleases cleave phosphodiesters by transferring a phosphoryl group from a bridging oxygen atom to an activated water molecule.

Both homing and restriction endonucleases have been extensively studied biochemically and structurally. A great deal of information regarding the specific structural mechanisms of these enzymes has been obtained and yet, there are still unresolved issues. Here we discuss the chemistry of phosphoryl transfer reactions, similarities and differences between the catalytic mechanisms of homing endonucleases and type II restriction endonucleases, and potential future studies that may shed more light on the details of these mechanisms.

Phosphoryl Transfer

Formally, the direct transfer of a phosphoryl group can take place according to a variety of mechanisms depending on the position of the attacking nucleophile relative to the leaving group and the degree of bond formation at different points along the reaction coordinate (Fersht, 1999). The transition state geometry of phosphoryl transfers is a trigonal bipyramid with three equatorial and two axial oxygens. Oxygen atoms entering or leaving this configuration must always be axial. The extremes with respect to bond cleavage and formation are associative mechanisms (S_N2), in which the nucleophile is completely bonded to the phosphorus atom before the leaving group is eliminated, and dissociative mechanisms (S_N1), in which the elimination of the leaving group occurs first and the nucleophile attacks an unstable metaphosphate intermediate (PO_3) (Figure 2.1).

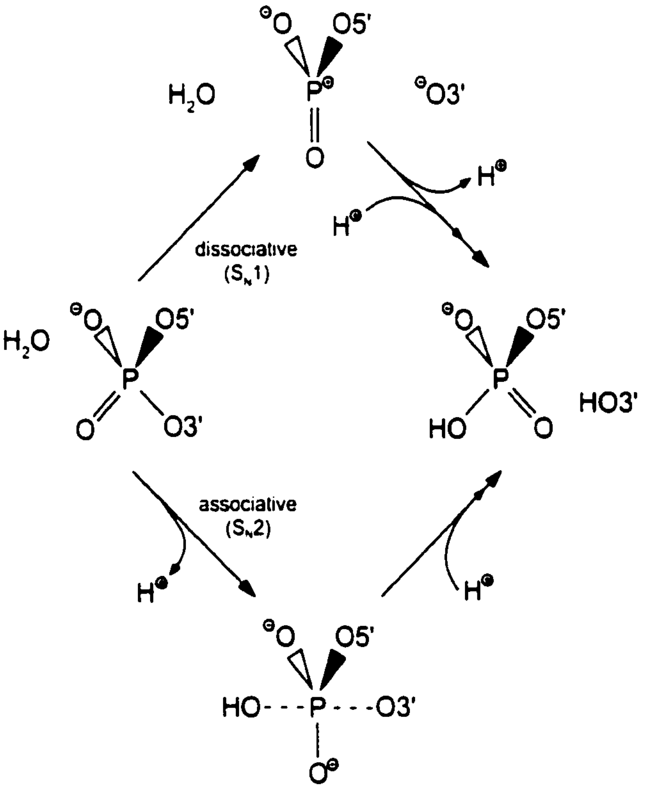


Figure 2.1: A schematic illustrating associative and dissociative extremes for the mechanism of phosphoryl transfer.

In solution, S_N1 reactions result in the racemization of products, but an enzyme that does not release metaphosphate constrains the position and orientation of the intermediate such that chirality is either maintained or inverted. Associative mechanisms can be classified as in-line or adjacent. When the nucleophile and leaving group are both apical at the same time along the reaction coordinate, the mechanism is termed in-line and the reaction results in an inversion of configuration. In an adjacent mechanism, the leaving group is equatorial when the nucleophile attacks and the nucleophile is equatorial when the leaving group is released. Thus, adjacent mechanisms are also characterized by a *pseudorotation* somewhere along the reaction coordinate that reorganizes the bypyramid so that oxygen atoms switch from apical to equatorial and vice versa. It has been shown experimentally that phosphoryl transfer enzymes use in-line associative mechanisms and no evidence has been seen for adjacent or dissociative mechanisms. The products of some transfers do retain configuration relative to substrate, but this is accounted for by a two-step reaction through a phospho-enzyme intermediate. This leads to a retention of configuration through an even number of associative transfers rather than a single dissociative transfer.

At physiological pH, phosphate ester bonds have large barriers to cleavage even though they are thermodynamically unstable. This is mostly due to the fact that around neutral pH, the phosphoryl group is negatively charged and electrostatically repels potential attacking nucleophiles. This duality of kinetic stability and thermodynamic instability can explain nature's choice of phosphate for such a variety of roles (Westheimer, 1987). To efficiently catalyze the cleavage of phosphate esters, certain chemical elements are required. These include a nucleophilic group to transfer the phosphoryl group to, a basic moiety to activate and position the nucleophile, a general acid to protonate the leaving group, and one or more positively charged groups to stabilize the phosphoanion transition state. The diversity of chemical groups available to proteins has made it possible for evolution to arrive at many strategies that satisfy the above requirements. A common feature of many nuclease catalysts (and other phosphoryl transfer enzymes), including restriction and homing endonucleases, exonucleases, and some ribozymes, is the use of bound metal ions as cofactors. Endonucleases with one, two, and three metals bound in active complexes have been observed crystallographically. Metal ions can act as Lewis acids to lower the pK_a of coordinating water molecules. A resulting hydroxide may then serve either as the nucleophile or the general base. Alternatively, an acidic, metal-bound water molecule can protonate the leaving group. Perhaps most importantly, the positive charge of the divalent metal ion can stabilize the -2 charge that accumulates at the phosphoanion transition state relative to the -1 charge of the ground state. All three of these functions could lower the free energy of the transition state and accelerate the reaction. It has long been understood that divalent metals are required for endonuclease function and that even a high concentration of monovalent cations does not restore activity. General one- and two-metal mechanisms are illustrated in Figure 2.2.

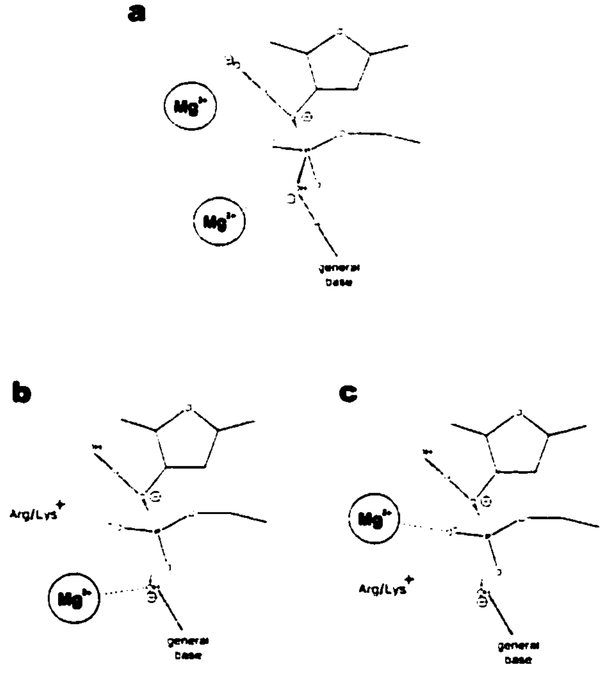


Figure 2.2: The canonical transition states of one and two metal endonucleolytic mechanisms. In each mechanism the metal ions stabilize the negative charge that develops on the dianionic transition state. (a) The two-metal mechanism uses metals to position both the nucleophile and the proton donor water molecules. (b) A one-metal mechanism observed in restriction endonucleases in which the metal ion is responsible for positioning and activating the nucleophile. (c) A one-metal mechanism observed in I-*Ppo*I in which the metal ion positions the proton donor, but does not make any contacts to the nucleophilic water. Both one-metal mechanisms include additional positive charge from protein side-chains to complement the divalent metal ion and stabilize the dianionic transition state.

Exceptions are ribonucleases like barnase and ribonunclease A that cleave phosphodiester bonds in RNA. Rather than using metal ions as cofactors, these enzymes provide the necessary positive charge by positioning a lysine residue in the active site. Likewise, the hairpin and hammerhead ribozymes do not have a divalent metal in the active site. In this case, metals are needed for the proper folding of the ribozyme, but do not seem to play a direct role in the catalytic mechanism (Murray and Scott, 2000; Rupert and Ferre-D'Amare, 2001). A possible explanation for the lack of metals in some ribonucleolytic active sites could be the relative instability of RNA phosphodiesters compared to DNA phosphodiesters. These enzymes have an easier job than deoxyribonucleases due to the presence of 2' hydroxyls that can be positioned and activated for nucleophilic attack and formation of a 2',3'-cyclic phosphodiester.

The catalytic mechanims of type II restriction endonucleases have been extensively studied. More recently, our understanding of homing endonuclease catalysis has also been growing rapidly. Here, we summarize both what is known and what is yet to be determined about the catalytic mechanisms of both type II restriction and homing endonucleases.

Restriction Endonucleases

There are more than 3000 known type II restriction endonucleases. Restriction enzymes combined with methyltransferase enzymes form the basis of RM (restriction-modification) systems of bacteria and other prokaryotes that protect host organisms against the invasion of foreign DNA. Restriction endonucleases are used ubiquitously as reagents for the manipulation of DNA in the laboratory, but have long been a field of study unto themselves in which the fundamental biochemical, structural, kinetic, and catalytic features of protein-DNA interactions are considered. The first crystal structure of *Eco*RI bound to its cognate DNA site was published in 1986 at a resolution of 3 Å (McClarin et al., 1986). Since then, the protein/DNA complex structures of 11 other members of this family have been determined and the resolution for many of the complexes extends well past 2 Å (Winkler et al., 1993; Cheng et al., 1994; Newman et al., 1994; Bozic et al., 1996; Wah et al., 1997; Newman et al., 1998; Deibert et al., 1999; Deibert et al., 2000; Lukacs et al., 2000; Huai et al., 2001; van der Woerd et al., 2001). These high-resolution structures have allowed a detailed examination of both DNA binding and catalysis. The resulting wealth of information is well summarized by a recent review (Pingoud and Jeltsch, 2001).

Despite the abundance of biochemical and structural data, the exact mechanisms used by type II restriction endonucleases are still a matter of debate. Here, we will review the general structure of the type II active site and discuss different mechanistic proposals.

Type II Restriction Endonuclease Active Site Architecture

The active sites of type II restriction endonucleases are defined by the conserved PD...(D/E)XK motif, where the gap ranges between 8 (*Nae*I) and 25 residues (*Bgl*I) and X represents any hydrophobic amino acid. Different positions within the motif are conserved to varying amounts (Table 2.1).

Table 2.1: Summary of conserved active site motifs of type II restriction endonucleases.

enzyme (year)	2^{nd} site	$P_{x-1}D_x$	...	$(D/E)_{x-2}Z_{x-1}{}^{\$}K_x$	metals
*Eco*RI (1986)	D_{59}	PD_{91}	X_{19}	EAK_{113}	1
*Eco*RV(1993)	E_{45}	PD_{74}	X_{15}	DIK_{92}	2
*Bam*HI (1994)	E_{77}	D_{94}	X_{16}	EFE	2
*Pvu*II (1994)	[illegible]	[illegible]	X_9	ELK_{70}	2
*Fok*I (1997,IIS)		PD_{450}	X_{16}	DTK_{469}	0
*Bgl*I (1998)	E_{87}	PD_{116}	X_{25}	DIK_{144}	2
*Cfr*10I (1998,IIF)	E_{71}	PD_{134}	X_{53}*	SVK_{190}(E [illegible]	0
*Mun*I (1999)		PD_{83}	X_{14}	EIK_{100}	0
*Bgl*II (2000)	N_{69}	D_{84}	X_9	DVQ	1
*Ngo*MIV (2000,IIF)	E_{70}	PD_{140}	X_{44}	SCK_{187}(E [illegible]	2
*Bso*BI$^{\%}$ (2001)		D_{212}	X_{27}	ELK_{242}	0
*Nae*I (2001,IIE)	E_{70}	D_{86}	X_8	DCK_{97}	0
*Bse*634I (2002)	E_{80}	PD_{146}	X_{49}	GLK_{198}(E [illegible]	$1^{\#}$

red text denotes variations from the "conserved" sequence motif
$^{\$}$Z represents any hydrophobic group
*the sequence $P_{133}D_{134}X_{11}E_{146}L_{147}K_{148}$ also exists, but 146ELK was observed to point away from the active site in the protein/DNA complex
$^{\%}$BsoBI contains a $D_{251}E_{252}H_{253}$ motif and H253 is believed to be the general base
$^{\#}$the observed metal ion is Gd^{+3} and was used as a heavy atom derivitive

The first aspartic acid is the only residue to be absolutely conserved among the enzymes for which structural data are available. The is sometimes replaced by isoleucine, valine, asparagine, or threonine; the second acidic residue, by serine; the hydrophobic residue, by cysteine; and the lysine by aspartic acid. The motif serves as a general guideline, but does not guarantee the existence of an endonuclease active site. For example, the motif was observed in the sequence of the His-Cys box homing endonuclease (Wittmayer and Raines, 1996), I-*Ppo*I, but the structure revealed that the motif residues do not constitute the active site (Flick et al., 1998).

The active sites of type II restrictions endonucleases have been observed crystallographically in many different metal bound states. The extreme case is the 13 different crystal structures that have been described for *Eco*RV. There appear to be two main varieties of type II active sites with regard to metal-binding capabilities. The enzymes *Bgl*II and *Eco*RI have been observed to contain a single metal-binding site per subunit (Lukacs et al., 2000). Importantly, the *Eco*RI structure (1QPS) is an active complex, demonstrating that a single metal is sufficient for DNA cleavage. In contrast, the enzymes *Bam*HI, *Bgl*I, *Pvu*II, and *Ngo*MIV all contain two metal binding sites per subunit (Newman et al., 1998; Viadiu and Aggarwal, 1998; Deibert et al., 2000; Horton and Cheng, 2000; Grazulis et al., 2002) (Figure 2.3). *Bam*HI has been solved in both substrate and product complexes indicating a two-metal mechanism is responsible for endonuclease activity (Viadiu and Aggarwal, 1998). One of the two metal-binding sites is homologous to the single site observed in *Bgl*II and *Eco*RI (Lukacs et al., 2000) (M1, Figure 2.3). This site is formed by the two conserved acidic residues, the backbone carbonyl of the hydrophobic residue (PD...(D/E)XK), and a non-bridging oxygen from the scissile phosphate. The second metal site (M2) is created by the first conserved aspartic acid, a substrate non-bridging oxygen, and a complement of other acidic residues depending on the specific enzyme. In *Pvu*II, the backbone carbonyl of G56 completes the second metal site, while in *Bam*HI, it is the side chain of E77.

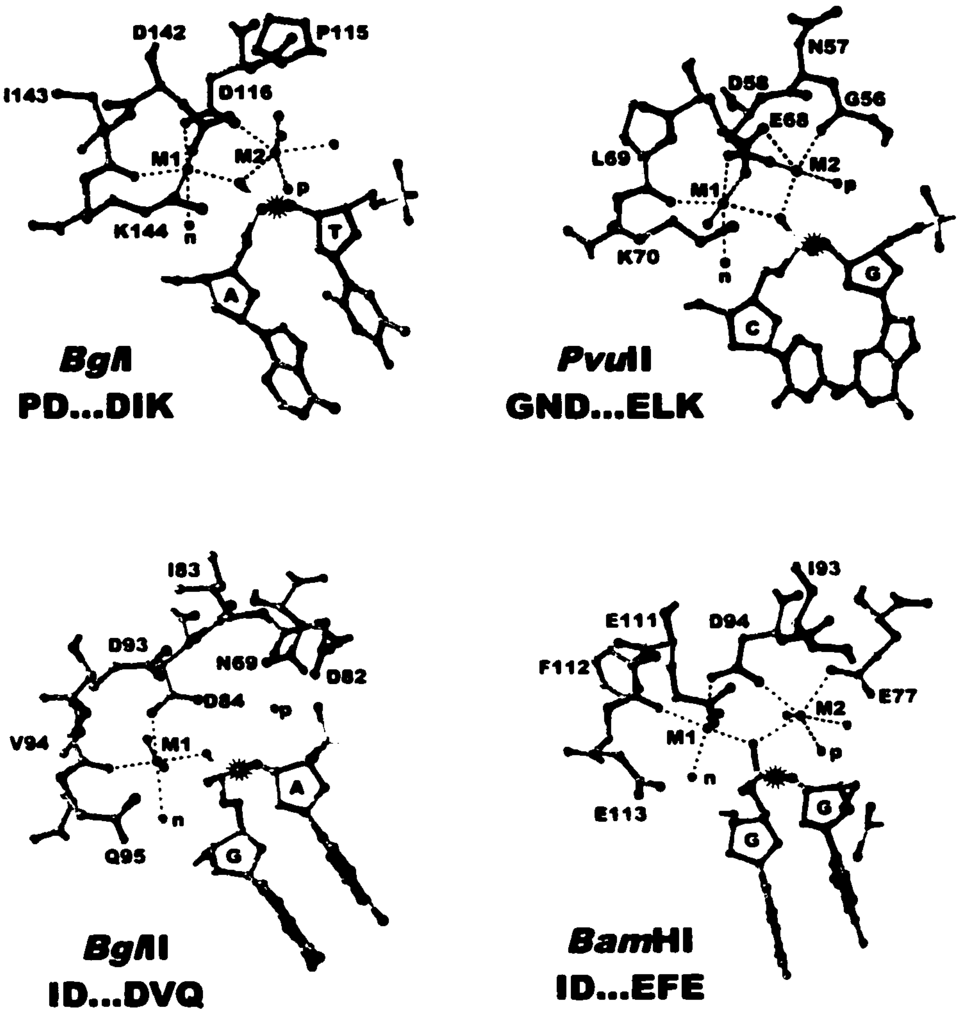


Figure 2.3: Ball and stick representations of the active sites of four restriction endonuclease enzyme-substrate complexes. Each active site is labeled along with its sequence motif and is shown in the same orientation. The metal ion (green) site that is conserved in all four structures is labeled M1 and the second site is labeled M2. Metal ligands are indicated with green dotted lines. Water molecules are labeled according to their proposed function: nucleophile (n) or proton donor (p). The scissile bonds are indicated with asterisks.

Lying outside the clearly defined one-metal and two-metal categories is *Eco*RV. *Eco*RV has been solved in a variety of different states with many divalent cations (Mg^{2+}, Mn^{2+}, Co^{2+}, Ca^{2+}, Mg^{2+}/Ca^{2+}, and Mn^{2+}/Ca^{2+}) by both co-crystallization and soaking experiments (Kostrewa and Winkler, 1995; Perona and Martin, 1997; Horton et al., 1998;

Martin et al., 1999; Horton and Perona, 2000). The data available indicate that *Eco*RV uses 2-3 metal ions to cleave DNA, but the number of metal ions involved in the catalytic mechanism is still debated. Clearly, the precise mechanism by which these enzymes catalyze DNA cleavage depends on the minimal number and positions of metal ions in the active site. For example, the proposed catalytic mechanism of the LAGLIDADG homing endonuclease, I-*Cre*I, was revised after a high-resolution crystal structure of a metal-bound product complex was determined to have three metals bound per protein dimer instead of two (Chevalier et al., 2001). The following sections describe two potential catalytic mechanisms that have been proposed for type II restriction endonucleases.

The one-metal mechanism

The one-metal mechanism involves a metal bound in site 1 as described above (Figure 2.2c). A metal-coordinated water molecule is positioned and activated for in-line nucleophilic attack of the electrophilic phosphorous atom. In *Bgl*II, the nucleophilic water molecule also makes hydrogen bonds to Q95, both bridging and non-bridging oxygen atoms from the scissile phosphate, and a water molecule bound to a non-bridging oxygen atom from the 3'-adjacent phosphate (Lukacs et al., 2000). A second metal bound water molecule is positioned by N69, D82, and D84, and makes a hydrogen bond to the leaving group oxygen. This water molecule appears to be well positioned to donate a proton to stabilize the leaving group. In *Eco*RI, the nucleophilic water is not contacted by a protein residue, but does make hydrogen bonds to both a 3'-phosphate bound water and a non-bridging oxygen from the scissile phosphate (Pingoud and Jeltsch, 2001). A putative leaving group protonating water molecule is positioned by R145. The single metal mechanism uses an in-line, metal-bound nucleophile with a pK_a lowered through metal association. However, neither structure has an obvious general base, which would be required to abstract the proton from the nucleophilic water molecule in an associative nucleophilic attack. Possible features that could complete a single metal mechanism include substrate-assisted catalysis where the general base is a non-bridging phosphate oxygen (Jeltsch et al., 1993), a bulk solvent supplied metal bound hydroxide ion that does not need to be deprotonated, or a dissociative transition state. Any of these possibilities would abrogate the requirement for the protein to supply an efficient general base.

The two-metal mechanism

The two-metal mechanism is similar to the one-metal mechanism in that the metal ion in site 1 is responsible for positioning the nucleophilic water for in-line attack. In addition, a lysine residue in both *Pvu*II (K70) and *Bgl*I (K144) contacts the nucleophile (Figure 2.2a,b) (Newman et al., 1998; Horton and Cheng, 2000). The second metal is located on the other side of the scissile phosphate from the nucleophilic water molecule and is well positioned to activate a water molecule for protonation of the leaving group oxygen and to help stabilize the negative charge that develops in an associative transition state (Figure 2.1). Intriguingly, a calcium bound, pre-reactive complex of *Bam*HI indicates that E113 makes a hydrogen bond to the nucleophilic water molecule (2.9 Å) and represents a possible general base (Figure 2.3) (Viadiu and Aggarwal, 1998). However, neither *Pvu*II nor *Bgl*I has a potential general base positioned close to the nucleophilic water which results in mechanistic issues similar to those in the single-metal mechanism. Furthermore, a mutation of E113K in *Bam*HI kills the enzyme despite the fact that E113 is located in the lysine position of the conserved active site motif (Xu and Schildkraut, 1991). This divergence of enzyme active sites even between enzymes that belong to a putative (two-metal) sub-grouping of type II enzymes illustrates a crucial point. At our current level of understanding, it seems completely possible, if not likely, for type II enzymes to use several different catalytic mechanisms. Even enzymes that have both been observed to bind two metal ions might differ in their reaction mechanisms and reaction coordinates. For example, *Bam*HI (with a putative general base) might proceed along a fully associative path while *Pvu*II (lacking a putative general base) could catalyze the reaction through a more dissociative path.

Homing Endonucleases

Homing endonucleases are proteins encoded by open reading frames contained within mobile introns (Belfort and Perlman, 1995; Belfort and Roberts, 1997; Chevalier and Stoddard, 2001). The first homing endonuclease was discovered in yeast, but they are now known to exist in all branches of life. Homologous endonucleases are also found as optional, independently folded domains in self-splicing protein introns termed 'inteins'. The catalytic activity of these endonucleases promotes the lateral transfer of the intervening sequences by a targeted transposition mechanism termed homing (Figure 2.4). These enzymes loosely recognize 14-40 DNA base pairs in contrast to the strict recognition of 4-8 base pairs exhibited by restriction endonucleases. While type II restriction enzymes are generally intolerant of base substitutions in their recognition sequence, homing endonucleases can accommodate a variety of polymorphisms in their target sites, and in some cases, have regions of their recognition site that are completely degenerate. There are four group I, intron-encoded, homing endonuclease families that are characterized by well-conserved sequence motifs. They are the LAGLIDADG, His-Cys box, H-N-H, and GIY-YIG families. In the following section, we will focus on the catalytic mechanism of each of these enzyme families.

LAGLIDADG

The LAGLIDADG enzymes comprise the largest homing endonuclease family, with well over 200 known members, and can contain either one or two copies of the conserved motif. Enzymes that contain one LAGLIDADG sequence, such as I-*Cre*I and I-*Ceu*I, are ~18-25 kDa in mass and form obligate homodimers. Those with two copies in the same chain, such as like PI-*Sce*I and I-*Dmo*I, have about twice the mass of their homodimeric cousins and fold into pseudo-symmetric monomers. The first 7 residues of the motif form the core of an α-helix that comprises the center of the dimer (or pseudo-dimer) protein interface, and are fairly well conserved (Table 2.2).

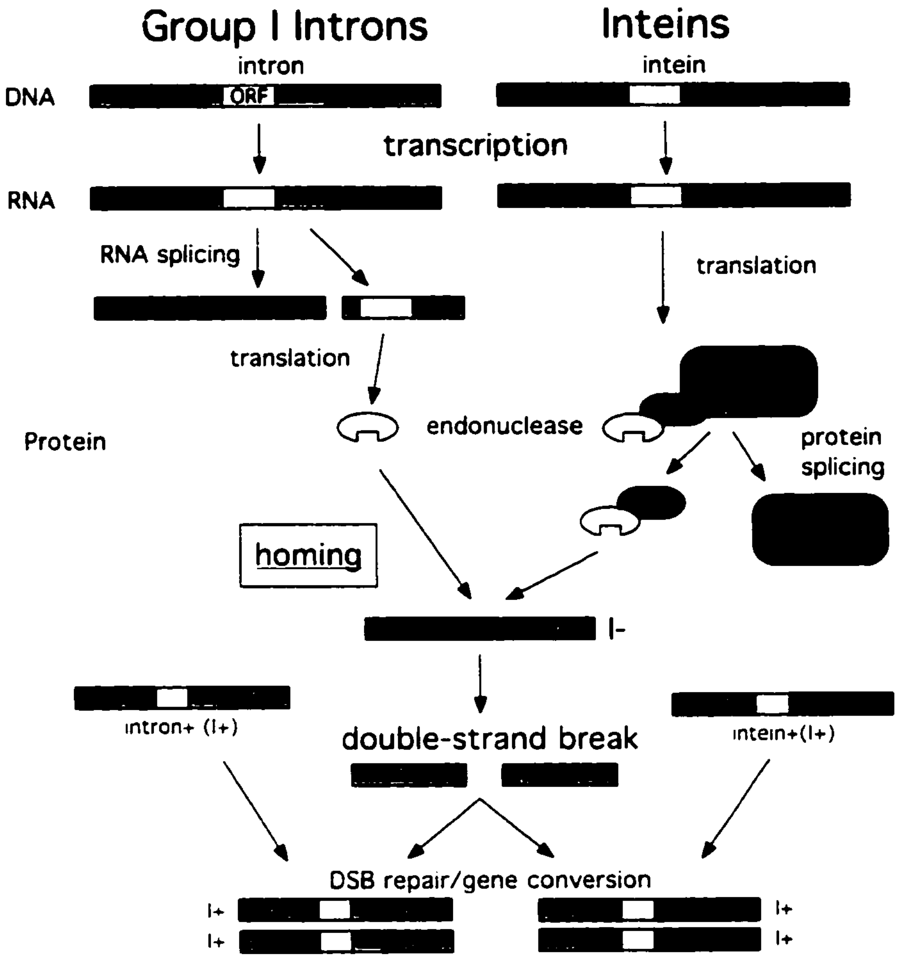


Figure 2.4: The process of homing. Homing endonucleases are encoded by open reading frames (ORF) within either introns (left) or inteins (right). After RNA or protein splicing, the endonuclease introduces a double-strand break in cognate alleles lacking the intron or intein. Double-strand break repair (DSB) is initiated to repair the lesion, resulting in the conversion of the allele from intron minus to intron plus. Thus, the end result of the homing process is the duplication and transfer of both the intervening sequence and the endonuclease ORF.

The last three residues are almost completely conserved and are composed of the metal-binding catalytic aspartic acid (sometimes a glutamic acid) flanked by glycine residues which facilitate a tight β-turn structure and terminate the LAGLIDADG helix. Outside this conserved sequence signature, the LAGLIDADG enzymes diverge widely in sequence. Most of the crystal structures that have been solved have been of apo enzymes (Duan et al., 1997; Heath et al., 1997; Silva et al., 1999; Ichiyanagi et al., 2000), so the following description of the catalytic mechanism for LAGLIDADG enzymes is based on the recent high resolution co-crystal structures of I-*Cre*I and its DNA homing site in both substrate and product complexes and the superposition of other known structures using I-*Cre*I's DNA binding mode (Chevalier et al., 2001).

Table 2.2: LAGLIDADG conserved motifs

Enzyme	LAGLIDAD$_n$G	Active site
I-*Cre*I	LAGFVDGD$_{20}$G	K$_{98}$ R$_{51}$ E$_{47}$
I-*Dmo*I	LLGLIIGD$_{21}$G IKGLYVAE$_{117}$G	K$_{120}$ K$_{43}$ Q$_{42}$ - K$_{130}$ N$_{129}$
PI-*Sce*I	LLGLWIGD$_{218}$G LAGLIDSD$_{326}$G	K$_{301}$ R$_{231}$ D$_{229}$ K$_{403}$ H$_{343}$ T$_{341}$
I-*Pfu*I	LAGFIAGD$_{149}$G IAGLFDAE$_{250}$G	L$_{220}$ - D$_{173}$ K$_{322}$ - M$_{263}$
I-*Ceu*I	LAGFLEGE$_{66}$A	K$_{116}$ Q$_{93}$
I-*Mso*I	IAGFLDGD$_{??}$G	?
I-*Ani*I	LVGLFEGD$_{26}$G LVGFIEAE$_{158}$G	?

The I-*Cre*I active site appears to conform to a canonical two-metal mechanism similar to those described above. One metal positions and activates the nucleophilic water and the second metal stabilizes both the transition state phospoanion and the oxygen leaving group. The most striking and unusual feature of this structure is that the second metal is used jointly by the active sites of both subunits of the homodimer. This results in three metals shared between two active sites in a relatively small area (Figure 2.5). They are coordinated by the two aspartic acids, the oxygens on the scissile phosphates, the conserved glycine's backbone carbonyl, and water molecules (two in the substrate complex and one in the product complex). In the substrate structure, a metal-bound water (hydroxyl) appears well positioned for nucleophilic attack; however, there are no direct protein contacts to this water, which once again begs the question of what functions as the general base. Instead of a protein contact, there is a well-ordered network of water molecules such that hydrogen-bonded water molecules link the nucleophilic water to the leaving group oxygen or to bulk solvent. This raises the interesting possibility that the reaction proceeds by concerted transfer of hydrogen atoms along this chain of water molecules and that these exchanges are responsible for both nucleophilic activation and leaving group protonation. Another possibility that arose in the discussion of type II restriction endonucleases is that catalysis is substrate-assisted. The substrate-assisted mechanism uses a non-bridging phosphate oxygen as the general base. The pK_a for the first ionization of phosphoric acid is around 2 so it does not initially seem like a good candidate, but it might be a better base than nearby water molecules because the pK_a of H_3O^+ is -1.7. If the general base oxygen is on the scissile phosphate, a further advantage for the substrate-assisted mechanism is that protonation of the phosphate oxygen neutralizes the charge repulsion between the phosphate and the negatively charged nucleophilic water.

The $p21^{ras}$ GTPase is similar to I-*Cre*I in that the general base has not been located by mutagenesis studies and is not clearly defined by the structure of the active site. Molecular dynamics simulations (Schweins et al., 1994) and linear free energy relationships (Schweins et al., 1995) have indicated that substrate-assisted catalysis is used by $p21^{ras}$ and perhaps other GTPases.

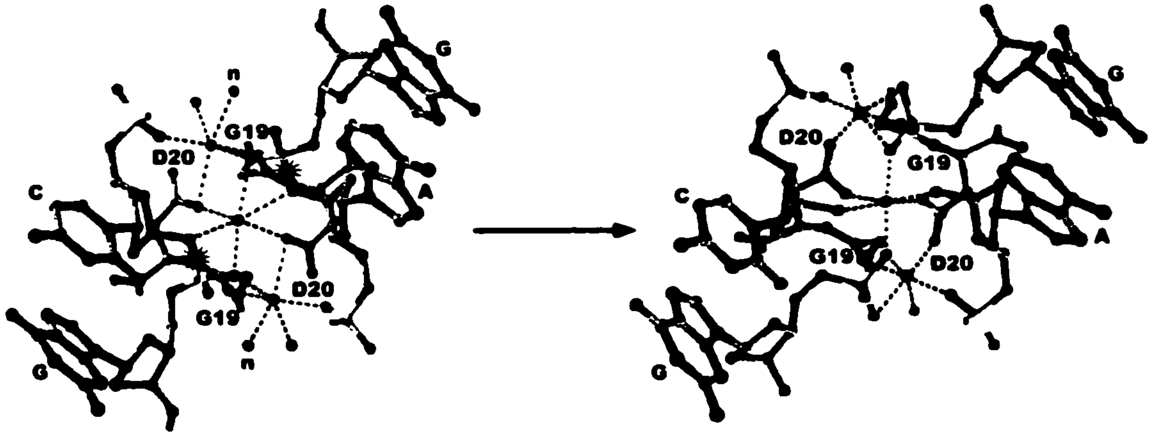


Figure 2.5: The active sites of I-*Cre*I in substrate (left) and product (right) forms. The green atoms are metal ions, the nucleophilic waters are labeled (n), and the scissile bonds are indicated by stars (left).

In either the water network or the substrate-assisted mechanism, the environment created by the protein/DNA complex must stabilize any high-energy intermediates along the reaction coordinate. This stabilization is likely contributed to by three other residues in I-*Cre*I that have been shown to be important for catalysis: lysine 98, arginine 51, and glutamine 47 (Seligman et al., 1997). The corresponding residues in I-*Ceu*I are also required for activity (Turmel et al., 1997). However, these residues are only moderately conserved in the other LAGLIDADG enzymes (Chevalier and Stoddard, 2001) (Table 2). The two copies of K98 in I-*Cre*I are K120 and nothing in I-*Dmo*I, and L220 and K322 in PI-*Pfu*I. The two copies of R51 in I-*Cre*I are nonexistent in PI-*Pfu*I, and R231 and H343 in PI-*Sce*I. Finally, the two copies of Q47 in I-*Cre*I are D173 and M263 in PI-*Pfu*I, and D229 and T341 in PI-*Sce*I. One aspect of these residues that is well conserved is that most have the capacity to either donate or accept one or more hydrogen bonds. Another conserved aspect is the presence of basic residues. In the complex between I-*Cre*I and DNA, these residues appear to be responsible for ordering the network of water molecules that surrounds the scissile bond. It is possible that these peripheral active-site residues are responsible for positioning and polarizing this water network in an optimal catalytic configuration. The lack of strong conservation of these residues among family members could indicate that many different combinations and configurations of chemical groups are able to fill this role. However, it is still not clear that all LAGLIDADG enzymes utilize the same endonucleolytic mechanism. Different LAGLIDADG enzymes might use different catalytic strategies for which I-*Cre*I is not a good model. I-*Cre*I is a perfect example of an enzyme for which there are biochemical data on active-site mutants and both substrate and product complexes have been solved to high resolution, yet the precise chemical mechanism is still a matter of debate. We will revisit this point at the end of the review where we discuss possible future directions in the study of endonuclease mechanism.

His-Cys Box

The His-Cys box homing endonuclease family is defined by 2 clusters of histidine and cysteine that form structural zinc-binding sites. Zinc is required for protein folding, but does not directly play a role in catalysis. I-*Nja*I (Elde et al., 1999), I-*Nan*I, and I-*Nit*I (Elde et al., 2000) are isoschizomeric members of this family, but the most well studied member of the His-Cys box family is I-*Ppo*I (Ellison and Vogt, 1993; Argast et al., 1998;

Flick et al., 1998; Wittmayer et al., 1998). The catalytic mechanism of I-*Ppo*I has been studied structurally (Galburt et al., 1999) and biochemically (Friedhoff et al., 1999; Mannino et al., 1999) and both studies support similar conclusions. I-*Ppo*I uses a single metal ion mechanism in which a magnesium ion accelerates the reaction by stabilizing the phosphoanion transition state, positioning and activating a water molecule to donate a proton to the leaving-group oxygen, and possibly introducing strain in the substrate complex that is relieved in the product complex. The bound metal is coordinated by asparagine 119, the leaving group oxygen, a non-bridging oxygen on the scissile phosphate, and three water molecules in the substrate complex (Figure 2.6).

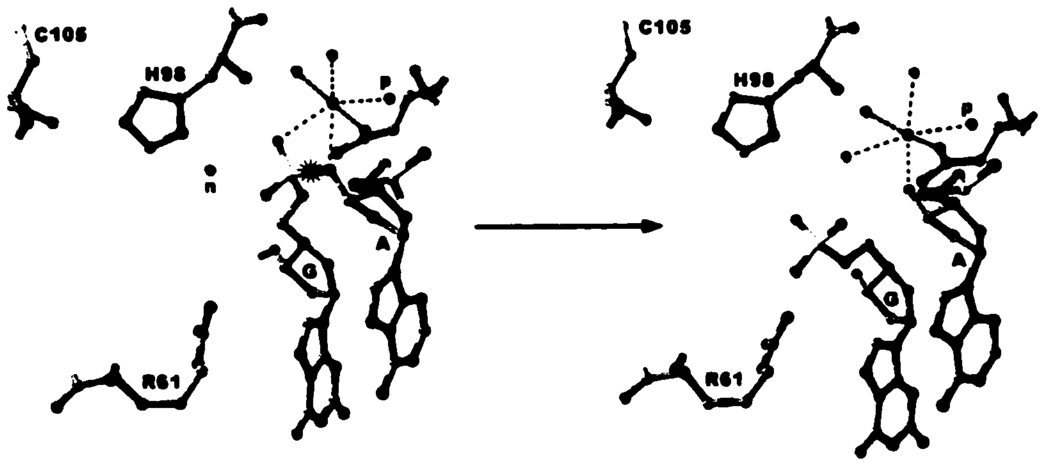


Figure 2.6: The active site of I-*Ppo*I in substrate (left) and product (right) forms. The green atoms are metal ions, the nucleophilic water is labeled (n), the proton donating water molecule is labeled (p), and the scissile bond is indicated with a star (left).

In the product complex, the now terminal 5'-phosphate moves away from the metal and towards arginine 61. The leaving group is stabilized by the metal and a fourth water molecule now replaces the non-bridging oxygen as the sixth metal ligand. The nucleophilic water has been observed in the crystal structures and is positioned nicely for in-line attack. Histidine 98 is the general base and a hydrogen bond is observed between the nucleophile and the δN of H98. The backbone carbonyl of a zinc-coordinating cysteine (C105) stabilizes the correct tautomer of H98 by hydrogen bonding to the εNH. Mutational analysis has indicated that H98A eliminates nearly all activity, N119A and R61A depress activity by 3-4 orders of magnitude, while H78A and H101A decrease activity by about a factor of 10 (Mannino et al., 1999).

This class of enzymes is novel in their use of a histidine residue as the general base in a DNA endonuclease reaction. Furthermore, it is noteworthy that the nucleophilic water is not a metal ligand and so is not being activated directly by the metal ion. The interaction between the scissile phosphate and R61 increases significantly in the product so it is hypothesized that R61 acts to stabilize the transition state relative to the substrate to accelerate the forward reaction and the product relative to the transition state to decelerate the reverse reaction, but it is not known what this interaction is like in the transition state. Also, it is not clear what the precise roles of H78 and H101 are. Interestingly, the nonspecific *Serratia* nuclease uses the exact same active site as I-*Ppo*I (Friedhoff et al., 1996; Friedhoff et al., 1999; Miller et al., 1999). *Serratia* nuclease is able to cleave both DNA and RNA and is the most well-studied member of a family of nonspecific nucleases of which NucA is another structurally characterized member. Homology modeling has shown that NucA shares the same active site architecture as I-*Ppo*I, but the arginine (R61) that is conserved in *Serratia* (R57) is an aspartic acid (D95) in NucA (Meiss et al., 2000). The substitution of the positively charged arginine with a negatively charged aspartate is in conflict with the proposed mechanism and the authors have proposed the asparate binds a second metal ion, thus reconstituting the electrophilic nature of that position in the active site in NucA.

H-N-H

The H-N-H family of endonucleases is characterized by a defining motif that spans 25 residues of well-conserved sequence. The homing endonucleases I-*Cmoe*I, I-*Tev*III, I-*Hmu*I, and I-*Hmu*II (Eddy and Gold, 1991; Goodrich-Blair and Shub, 1996) are members of the H-N-H family and I-*Cmoe*I has been well characterized biochemically (Drouin et al., 2000). No H-N-H homing endonuclease structure has been solved. However, the H-N-H motif has been structurally characterized by studies on the bacterial colicins E7and E9. Bacterial colicins are nonspecific endonucleolytic weapons used by bacteria to compete with one another for limited resources. The endonucleases are expressed by a bacterium that also has a gene for an "immunity protein" that binds to the colicin and inactivates it. Once the endonuclease is outside the cell it loses the immunity protein as it is actively transported through a target cell's membrane and becomes active. The structures of colicins E7 and E9 have both been solved in complex with their respective immunity proteins (Im7 and Im9) (Kleanthous et al., 1999; Ko et al., 1999). Interestingly, the immunity proteins do not inactivate the colicins by directly occluding the active sites, but probably prevent DNA from binding productively to the colicin (Kleanthous et al., 1999). The H-N-H motif folds into a conserved structural motif that has been called the $\beta\beta\alpha$-Me motif (Kuhlmann et al., 1999). It is composed of two antiparallel β-strands followed by an α-helix and is stabilized by the binding of a metal ion between the first strand and the helix. This active site structural motif is also observed in the His-Cys box family, *Serratia* nuclease family, and the T4 endonuclease VII (Raaijmakers et al., 2001). The His-Cys box enzymes, T4 endonuclease, and colicins have very different tertiary structures overall, so the re-occurrence of the $\beta\beta\alpha$-Me motif is probably best explained by convergent evolution of enzyme active sites in evolutionarily unrelated families. The conservation of the $\beta\beta\alpha$-Me motif also suggests that the catalytic mechanisms for these endonucleases might be related.

Although none of the H-N-H enzymes have been studied crystallographically in DNA complexes, the combination of biochemical results and the structures of the active sites under several different conditions suggest hypotheses regarding the mechanism of phosphodiester hydrolysis used by the H-N-H family of enzymes. The colicin structures have been solved under three different conditions. The active site of colicin E9 has been solved with nickel bound and with no metal bound (Figure 2.7) while the active site of colicin E7 has been visualized with a bound zinc ion (Kleanthous et al., 1999; Ko et al., 1999). In both of the structures with bound metal, the cations are tetrahedrally coordinated. The zinc structure clearly shows the metal coordinated by three histidine residues (H544, H569, and H573) and a water molecule. The nickel structure revealed the metal coordinated by two histidine residues (H102 and H127) and an oxygen from a bound phosphate molecule. NMR experiments verified the fourth ligand as a third histidine residue (H131) as seen in the zinc structure (Hannan et al., 1999). The phosphate in the nickel structure is also contacted by a fourth histidine residue (H103). R96, E100, and H127 have all been shown by mutagenesis experiments to be important for catalysis (Garinot-Schneider et al., 1996). These residues form a triad in both colicin structures, with E100 in the middle forming a salt bridge with R96 and accepting a hydrogen bond from H127. Activity assays have revealed that while zinc is able to stabilize the colicins against proteolytic digestion, the enzymes are not active against DNA (Pommer et al., 2001). The colicins display activity against double stranded DNA, single stranded DNA, and single stranded RNA (Pommer et al., 2001). Magnesium is preferred for dsDNA cleavage while nickel is favored over magnesium for ssDNA cleavage. Interestingly, the colicins cleave ssRNA with no requirement for metal cofactors although nickel does accelerate the reaction. In the metal-free active site, putative catalytic residues are observed to rearrange and take on different conformations. This brings up the intriging possibility that the basic H-N-H catalytic motif is highly adaptable and capable of taking on slightly different configurations depending on the enzyme and substrate (Pommer et al., 2001).

Based on these experimental observations, researchers have proposed two possible mechanisms (Pommer et al., 2001) (Figure 2.7). They both describe the activation of the nucleophilic water by the general base H103 and the stabilization of the negative charge of the transition state by a bound metal ion. In the 'magnesium' mechanism, magnesium is coordinated by H127 and H102, which are stabilized by the E100/R96 salt bridge. The metal-activated H102 then serves as a proton donor for the leaving group. In the 'nickel' mechanism, a nickel-bound water molecule functions as the proton donor. These mechanisms have a great deal in common with the well-characterized mechanisms for I-*Ppo*I and *Serratia* nuclease described above. One significant difference between the experimentally observed colicin structures and those of I-*Ppo*I or *Serratia* nuclease is that the catalytic metal coordination is tetrahedral instead of octahedral, although one large unsettled point is how magnesium binds in the active site. Since magnesium prefers octahedral coordination spheres, it is possible that in an active complex with magnesium and DNA bound, that the active site reorganizes to afford such a coordination. Magnesium also prefers oxygen ligands, so it will be interesting to see if the histidine coordination observed in the current structures is maintained in a magnesium complex. In the structure of T4 endonuclease VII, a calcium ion is predicted to be coordinated tetrahedrally by the oxygen atoms of N62, D40, and two oxygens from a bound phosphate molecule (Raaijmakers et al., 2001). While the structures of DNA-bound complexes of H-N-H and other ββα-Me enzymes are required to verify these conclusions, it seems that not only is the H-N-H motif adaptable, but that the ββα-Me scaffolding itself is adaptable and is used in a variety of ways by different classes of nucleases.

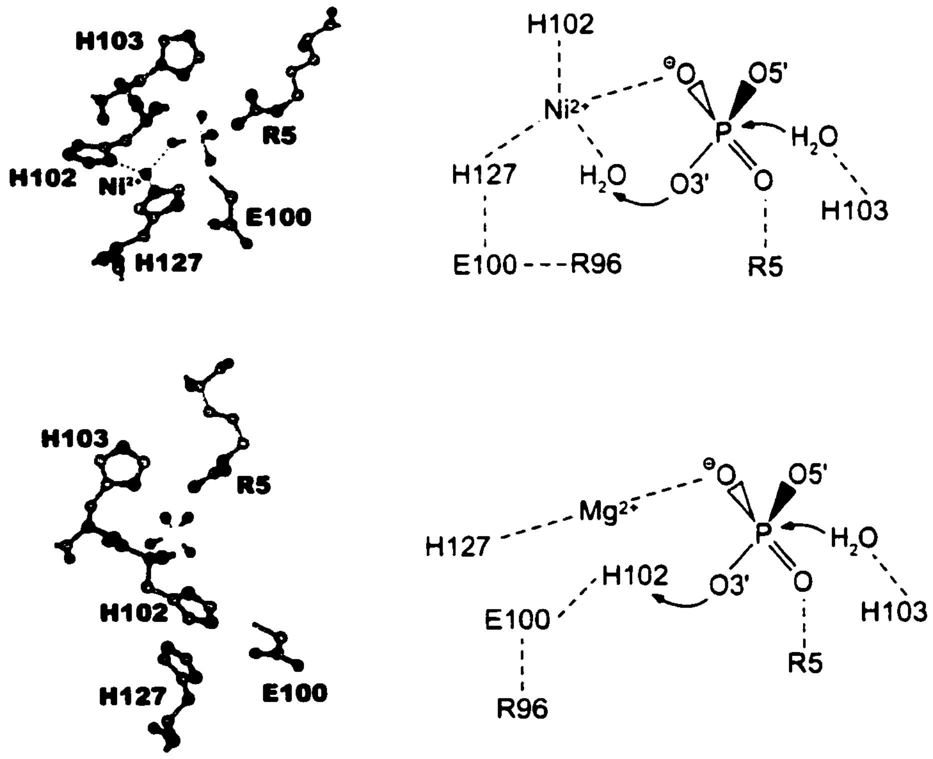


Figure 2.7: Ball and stick representations of two active-site structures from bacterial colicin E9 (left). The top structure shows the active site with a bound nickel ion and phosphate (PDB ID: 1BXI) while the lower structure contains only a bound phosphate (PDB ID: 1EVW). Potential catalytic mechanisms are also shown in schematic form (right). The top scheme is proposed for nickel-supported catalysis and is based on the nickel/phosphate structure while the bottom scheme is proposed for magnesium-supported catalysis and is based on the structure of the active site with no metal bound.

GIY-YIG

The fourth family of homing endonucleases is the GIY-YIG family represented by I-*Tev*I (Derbyshire et al., 1997; Kowalski et al., 1999; Van Roey et al., 2001) and I-*Bmo*I (Edgell and Shub, 2001). These enzymes have an N-terminal catalytic domain that contains the GIY-YIG sequence motif and a C-terminal DNA binding domain. The binding domain/DNA co-crystal structure of I-*Tev*I has been determined by crystallography (Van Roey et al., 2001) . The binding domain is a combination of helix-turn-helix, α-helix, and Zn-finger DNA binding subdomains although the Zn-finger has recently been shown to have a novel function as a distance determinant for cleavage (Dean et al., 2002). The catalytic domain has been studied by NMR (Van Roey et al., in press) and crystallography (Patrick Van Roey, Marlene Belfort and Victoria Derbyshire personal communication). It consists of conserved GIY and YIG triplets separated by 11 amino acids and folds into a novel α/β fold with a three-stranded anti-parallel β-sheet core. The primary role of the conserved motif appears to be structural, but there is no clear identification of an active site. Residues that have been implicated by mutagenesis to be catalytic (R27 and E75) (Derbyshire et al., 1997) are separated by at least 6.5 Å and do not form a well-defined active site. It is also not clear how a single catalytic domain of I-*Tev*I performs the sequential nicking that generates the double strand break. Perhaps a transient protein dimer is formed on DNA and generates two competent active sites. Alternatively, the catalytic domain of I-*Tev*I could catalyze both strand cleavages in series. Clearly, there are still many unknowns regarding the catalytic mechanism of this family. The structure of the full length enzyme bound to DNA should go a long way towards elucidating what appears to be a novel endonuclease mechanism.

Future prospects

To provide the best possible description of endonuclease catalytic mechanisms, each enzyme must be studied with a combination of crystallographic and biochemical techniques. The structures of a particular enzyme/DNA complex in a product complex and preferably many substrate complexes that have been trapped via different methods (mutation, thiol-substitution, Ca^{2+} ion, etc...) must be solved at resolutions sufficient to accurately determine the positions of metal ions and solvent molecules. Single turnover activity assays must be carried out with active site mutants, metal substitutions, and across a wide range of pH values to assay the effects on the rate of the chemical step. Through the application of these multiple experimental techniques, a particular enzyme's catalytic mechanism can be determined. However, as above, the exact molecular mechanism of catalysis often remains elusive even after gathering a variety of experimental results.

A way to test mechanistic proposals that is beginning to come of age is the computational modeling of reaction paths. These approaches can test the validity of particular mechanism by comparing the experimentally derived rates and crystal structures to those predicted by the calculation of energies along a proposed path (Glennon et al., 2000; Dinner et al., 2001). New techniques under development may eventually be able to determine the best possible reaction coordinate for a catalyzed reaction (without having to define a small subset of important coordinates) given the crystallographically determined substrate and product complex structures (Henkelman et al., 2000). Calculations with modified active site components will give an estimate for the energetic contributions of these components and allow researchers to tease apart the potentially multiple functions of individual groups. Combined with biochemical experiments and the structural determination of trapped substrate, intermediate, and product complexes, these rapidly developing computational methods should bring us closer to a complete understanding of how endonucleases and other enzymes work.

Summary

We have described the catalytic mechanisms of type II restriction and group I intron-encoded homing endonucleases. All of the enzymes discussed must supply the necessary catalytic elements for DNA cleavage: a properly positioned nucleophilic water molecule; a general base to activate the nucleophile; a general acid to protonate the leaving group; and localized positive charge to stabilize the phosphoanion transition state. In some cases, all of these elements have been determined, while in other systems the identity of only a subset are agreed upon. Interestingly, mechanisms of endonucleolytic catalysis do not segregate according to the pathway function of a particular enzyme (i.e. restriction or homing). For example, the apparent catalytic mechanism used by I-*Cre*I is more related to the mechanisms that are used by type II restriction enzymes than to the His-Cys box homing endonuclease, I-*Ppo*I. Much has been learned about the specific catalytic mechanisms of both restriction and homing endonucleases, but there is still much that is unknown. Open issues include the mechanisms of DNA cleavage used by HNH and GIY-GIY homing endonucleases, whether all endonucleases that share active site motifs use the same mechanism, and the quantitative energetic contributions to catalysis made by individual components of an active endonucleolytic complex along the reaction coordinate. The continued study of endonuclease mechanisms by biochemical, structural, and computational methods should reveal answers to these questions.

Chapter 3

Endonucleolytic Catalysis by I-*Ppo*I

Introduction

Any phosphodiester cleavage event requires several important mechanistic elements. These include activation of a nucleophile by a general base, stabilization of the phosphoanion transition state and protonation of the leaving group (Pingoud and Jeltsch, 1997). A common feature of most nucleases is the use of a bound metal ion as a Lewis acid to lower the pK_a of a water molecule, which may then act as a nucleophile or as a base. However, recent studies of catalysts such as homing endonucleases have indicated that the active sites and structural mechanisms of nucleases are more diverse than originally thought (Aggarwal and Wah, 1998; Jurica and Stoddard, 1999). Homing endonucleases are encoded by open reading frames within mobile introns (Belfort and Perlman, 1995; Belfort and Roberts, 1997) and are also found as independently folded domains in self-splicing inteins (Belfort et al., 1995; Pietrokovski, 1998). I-*Ppo*I, a His-Cys box homing endonuclease, utilizes a novel endonucleic mechanism.

I-*Ppo*I is encoded within the third intron of the nuclear 26S rRNA gene of *Physarum polycephalum* (Muscarella and Vogt, 1989; Muscarella et al., 1990). It cleaves a 16 bp homing site to generate four-base, 3' overhangs with a k_{cat}/K_m of 10^8 $M^{-1}s^{-1}$ and is activated by many divalent metal ions(Lowery et al., 1992; Ellison and Vogt, 1993; Wittmayer and Raines, 1996). The structure of I-*Ppo*I in complex with its DNA homing site has been determined at high resolution, both as an uncleaved complex in the absence of bound metal ions (trapped by sulfur substitution of a metal-binding DNA oxygen atom), and as a cleaved product complex that contains a single bound divalent cation per subunit (Flick et al., 1998). The metal is coordinated by a conserved asparagine residue, the 3'-OH of the cleaved DNA and four bound water molecules. It is positioned to interact with the scissile phosphate, and cannot coordinate a water molecule for in-line nucleophilic attack. Instead, a conserved histidine residue (H98) is properly positioned to activate a water molecule.

A nonspecific nuclease from *Serratia marcescens*, which displays a different fold from I-*Ppo*I, also contains conserved, essential asparagine and histidine residues in its active site (Miller et al., 1994; Friedhoff et al., 1996). Superposition of the active site regions of I-*Ppo*I and *Serratia* nuclease indicate that they are very similar and exhibit a similar metal coordination scheme (Friedhoff et al., 1999; Miller et al., 1999). It has been proposed that a similar, novel mechanism of DNA cleavage is used by both enzymes (Friedhoff et al., 1999; Miller et al., 1999). In this mechanism a histidine residue activates a water molecule for in-line attack, and a single metal ion stabilizes the phosphoanion transition state and the 3' leaving group. An asparagine residue is the only metal ligand donated by the protein.

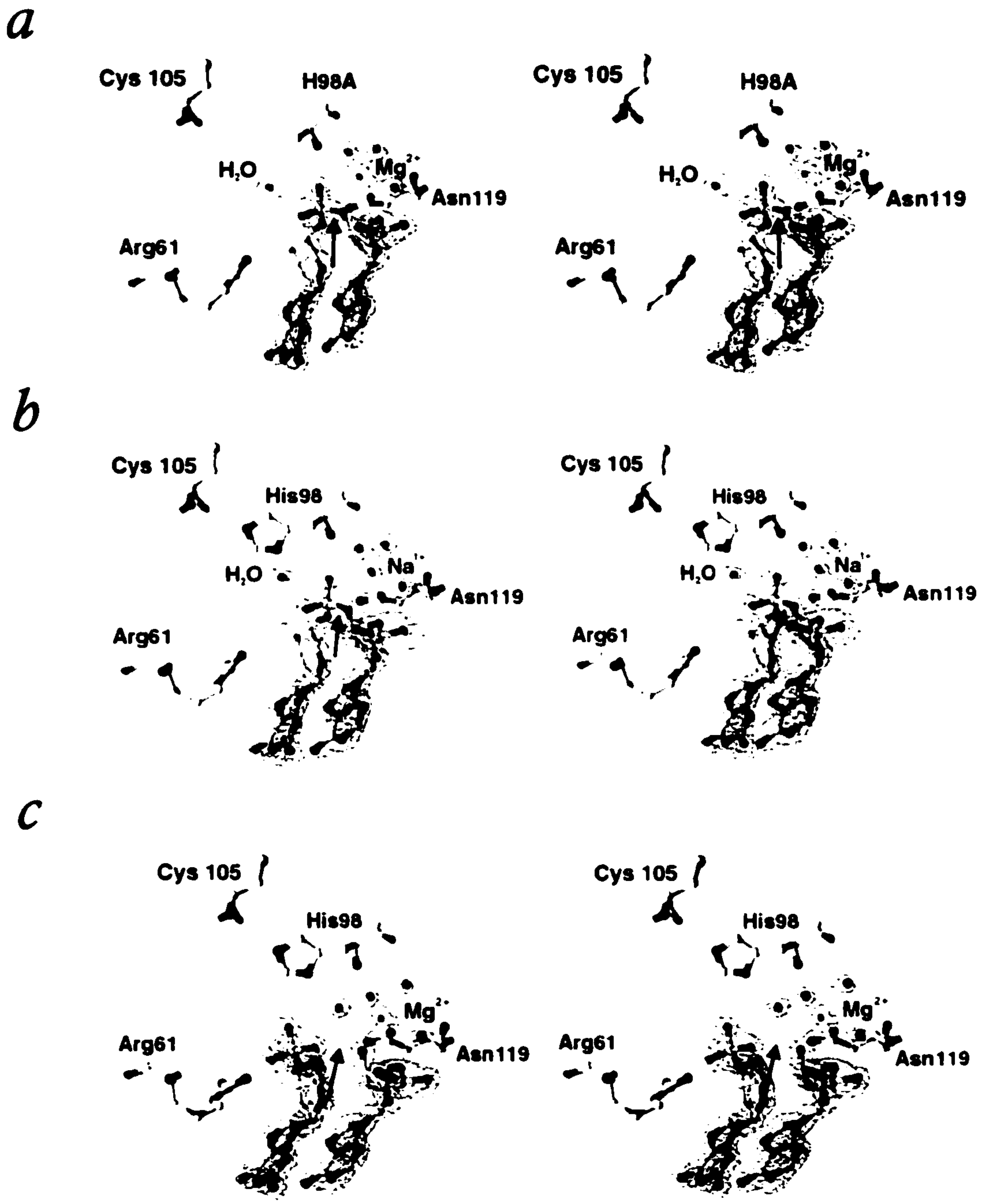
To directly visualize and characterize the mechanism of bond cleavage exhibited by I-*Ppo*I, its structure has been determined as three distinct species on the reaction pathway, using intermediate trapping to capture early states preceding bond cleavage. The enzyme–substrate (ES) complexes were trapped by substitution of a monovalent cation for an activating divalent cation in the active site, or by substitution of alanine for the His 98 general base. Both of these substitutions inhibit DNA cleavage.

Results

Catalytic metal coordination, substrate destabilization

A single monovalent (Na^+) or divalent (Mg^{2+}) cation is bound to the protein–DNA complex in each active site of the protein dimer; this binding site is visible and occupied in all three structures (Figure 3.1). The binding of cations at this site has been verified in separate experiments by soaking crystals with electron-rich manganese (Mn^{2+}) ions and examination of isomorphous and anomalous difference Fourier maps.

Figure 3.1: Electron density difference maps for three catalytic complexes formed by I-*Ppo*I. (a) Complex of H98A I-*Ppo*I, DNA, and magnesium. (b) Complex of wild type I-*Ppo*I, DNA and monovalent sodium ion. (c) Complex of wild type I-*Ppo*I, DNA, and magnesium. In all three maps, Fo-Fc density corresponding to the atoms omitted from the refinement model is displayed at 3σ contour levels. The maps were calculated after a full round of positional and simulated annealing refinement. Substitution of alanine for histidine at position 98 (a) or substitution of an octahedrally coordinated monovalent sodium ion for a similarly coordinated magnesium ion (b) both fully inhibit cleavage in the crystal. In both trapped ES complexes a bound hydrolytic water is clearly visible, as is a bound cation associated with the scissile phosphate. In the complex of wild type enzyme in the presence of magnesium (c), a similar difference map clearly demonstrates complete cleavage and formation of a product complex. The free 5'-phosphate group has undergone a significant conformational motion and clearly exhibits tetrahedral geometry, consistent with nucleophilic attack and formation of a PO_3^{2-} end group. The side chain of Arg 61 moves slightly to facilitate an electrostatic contact with the released phosphate in the product complex. With motion of the 5'-phosphate, the metal acquires an additional water ligand to complete a six-fold octahedral coordination shell.


a
Cys 105
H98A
H_2O
Mg^{2+}
Asn119
Arg61
Cys 105
H98A
H_2O
Mg^{2+}
Asn119
Arg61
b
Cys 105
His98
H_2O
Na^{+}
Asn119
Arg61
Cys 105
His98
H_2O
Na^{+}
Asn119
Arg61
c
Cys 105
His98
Mg^{2+}
Arg61
Asn119
Cys 105
His98
Mg^{2+}
Arg61
Asn119

These experimental results indicate the likelihood that this site is occupied by bound metal cations rather than by a water molecule, in agreement with the recent structural analysis of the *Serratia* nuclease (Miller et al., 1999). In the structure of the H98A enzyme–DNA–Mg^{2+} complex, the bound Mg^{2+} ion is coordinated in a six-fold geometry by the side chain oxygen of Asn 119, the bridging 3' oxygen and a single nonbridging oxygen atom of the scissile phosphate, and three well-ordered water molecules (Figures 3.1*a*, 3.2*a*). The average metal-oxygen ligand bond distance is 2.48 Å (s = 0.09 Å). This distance is slightly longer (by ~0.2 Å) than typical magnesium bond distances observed in a more highly charged metal binding site, and is indicative of the relatively neutral character of the metal ligands in I-*Ppo*I. In an ideal octahedral coordination geometry, all bond angles between non-opposing metal ligands would be near 90°. This geometry is observed in the cleaved product complex. In contrast, both structures of the trapped substrate complex demonstrate that the bond angle from the non-bridging phosphate oxygen, through the bound cation to the 3' bridging oxygen, is significantly strained with a value of 58°. The complementary angle from the same non-bridging phosphate oxygen to 'water 1' is 120°, while all other orthogonal angles through the bound metal are near 90° (σ = 5.6°; Fig. 3.2*a*).

The structure of the complex of wild type enzyme, DNA and Na^+ (Fig. 3.1*b*), trapped by the substitution of monovalent sodium cations for activating divalent magnesium cations, is in agreement with these observations. Electron density in omit maps supports the presence of an uncleaved phosphodiester bond and a bound cation in a strained six-fold octahedral coordination, similar to that described for the substrate complex trapped by the H98A mutation in the presence of Mg^{2+}. This is reasonable, as sodium resembles magnesium in its ionic radius and its number of electrons and can display octahedral coordination both in solution and in crystal lattices (Cotton and Wilkinson, 1980). Substitution of sodium for magnesium in the active site appears to inhibit cleavage by reducing the charge density of the bound cation, leading to inadequate stabilization of the phosphoanion transition state.

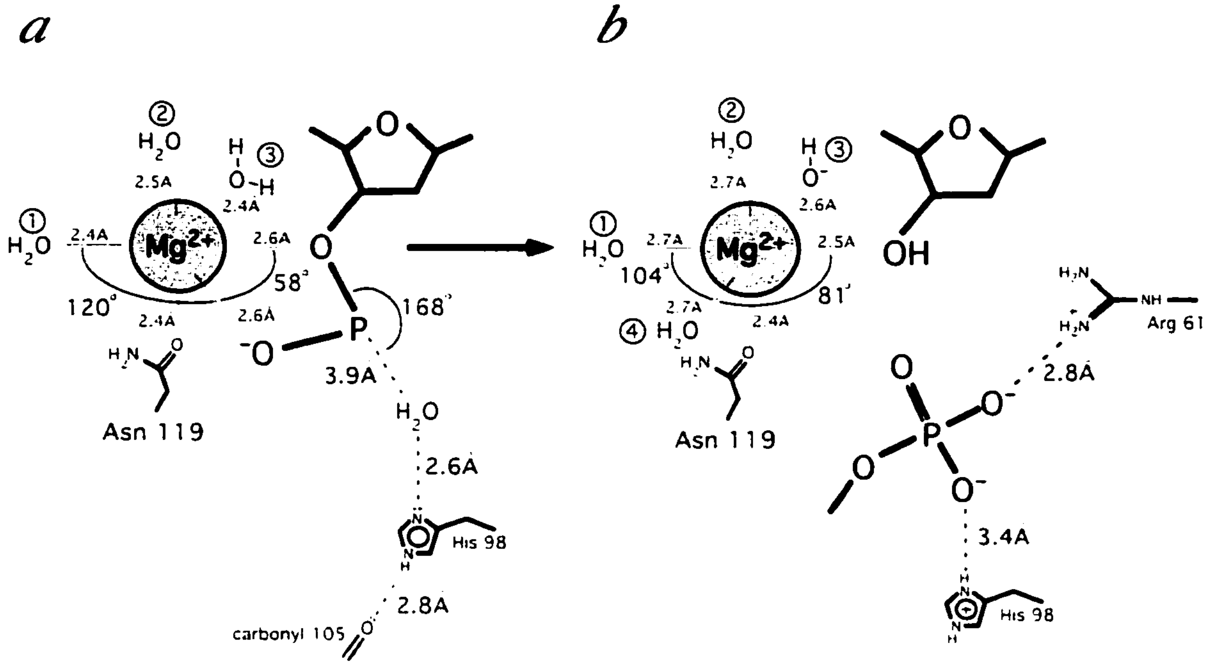


Figure 3.2: Interactions of DNA, solvent, and bound cation in the region of the scissile phosphate. (a) The uncleaved substrate complex. Distances and angles shown are an average for the two ES complexes as described in the text. In each structure. The distances and interactions are identical for the two active sites in the crystallographic asymmetric unit. The angle from the refined water position, through the phosphorous atom, to the 3' oxygen leaving group is 168°, which is reasonable for a direct S_N2 displacement leading to inversion of configuration of the phosphate group. The octahedral coordination of the bound metal is clearly strained, due to the interaction of the metal with adjacent bridging and non-bridging oxygen atoms from the scissile phosphate. (b) the product complex. Movement of the 5'-phosphate away from the bound cation has allowed the binding of an additional inner shell water ligand and optimization of the octahedral coordination of the metal.

The single bound metal ion in I-*Ppo*I appears to serve three distinct roles in catalysis (Figure 3.3).

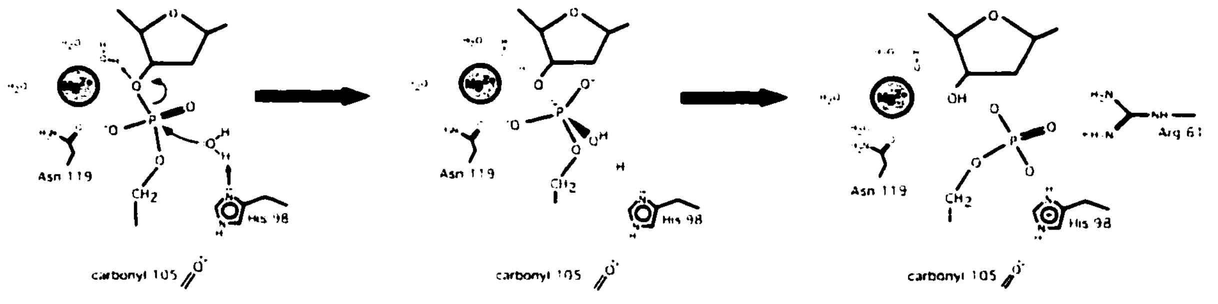


Figure 3.3: Proposed catalytic mechanism for the I-*Ppo*I/*Serratia* endonucleases. The single bound metal participates solely in transition state stabilization and protonation of the 3'-hydroxylate leaving group; it does not activate the hydrolytic water molecule. The general base for the reaction is a histidine imidazole ring, activated by a hydrogen bond between its ε-nitrogen and a backbone carbonyl oxygen. Residue numberings are from I-*Ppo*I.

Direct interaction of the bound metal ion with the scissile phosphate indicates that Mg^{2+} stabilizes the phosphoanion intermediate and the 3' hydroxylate leaving group. Second, a water molecule in the inner coordination sphere of the metal is appropriately positioned to donate a proton to the 3' hydroxylate leaving group. The metal ion decreases the pK_a of this water molecule and accelerates proton transfer. Finally, the bound metal forms a geometrically strained octahedral complex with surrounding protein, DNA and solvent atoms that is relaxed after DNA bond cleavage. It is interesting to note that the I-*Ppo*I endonuclease displays full catalytic function with a wide variety of divalent cations, while most restriction enzymes and the LAGLIDADG homing endonucleases are far more restricted in their use of metals. Promiscuous use of metal ions by an endonuclease might indicate that the bound cation is used predominantly for charge stabilization of the phosphoanion transition state. In contrast, greater sensitivity of cleavage rates with respect to metal species might imply a metal ion is also used to precisely position and activate a nucleophile. Recent studies on the *Eco*RV endonuclease, measuring rates of ES association, hydrolysis and dissociation with different metal ion species, appear to support this conclusion (Baldwin et al., 1999).

Alignment and activation of the hydrolytic water

In both ES complexes we observed an electron density peak corresponding to an ordered water molecule positioned for inline attack on the scissile phosphate (Figures 3.1*a*,*b*, 3.3*a*). In these structures the modeled water molecule is 3.9 Å from the phosphorus atom, and the angle from the water molecule through the phosphorus atom to the 3' oxygen leaving group is 168°. The appearance and position of the bound solvent molecule is not simply a result of the H98A mutation, but is a consistent feature of the uncleaved enzyme–DNA complex. The δN of His 98 appears to be directly hydrogen-bonded to the water molecule, with a distance between the imidazole nitrogen and water oxygen atom of 2.6 Å. This histidine appears to act as a general base by activating the water molecule (Figure 3.3) and may also participate in stabilization of the phosphoanion transition state.

Because the observed nucleophilic water molecule is not associated with a bound cation or any other electrophilic group, its pK_a is likely to be higher than the metal-bound water nucleophiles that are postulated for enzymes such as *Bam*HI or *Eco*RV. Because the pK_a of an uncharged histidine residue is only about 6, it would seem likely that such a side chain must be rendered a stronger base through an interaction with a hydrogen bond acceptor in order to effectively deprotonate this water molecule. In both substrate complexes the backbone carboxylate oxygen of Cys 105 is 2.8 Å from the His 98 εN, and is positioned to form a linear hydrogen bond (Figure 3.3*a*). In the structure of the *Serratia* nuclease19, the putative general base (His 89) displays a similar interaction between its εN and Asn 106. However, there are currently no reported pH dependence studies of the chemical step of the I-*Ppo*I reaction, nor has the importance of this interaction been experimentally tested by mutation of Asn 106 in *Serratia* nuclease or by measurement of the pK_a of His 98 in I-*Ppo*I.

Conformational changes and transition state stabilization

A series of conformational changes are observed in the active site as a result of DNA bond cleavage (Figures 3.1*c*, 3.3*b*). The free 5'-phosphate moves by over 2.5 Å from its position in the substrate complex and forms a 2.8 Å electrostatic bond with a guanido nitrogen of Arg 61, which moves by ~0.5 Å. The movement of the 5'-phosphate disrupts the interaction between its non-bridging oxygen and the bound metal ion. A fourth well-resolved water molecule is added to the inner metal coordination sphere, which assumes a more ideal octahedral geometry. The previously strained metal bond angles relax by ~20°, and the average orthogonal coordination angle for the ligands is 89.6° ($\sigma = 5.3$). The metal ion does not move significantly upon cleavage, and maintains interactions with Asn 119 and the 3' oxygen leaving group of the cleaved phosphodiester bond.

These structures indicate that the phosphoanion transition state is stabilized through contacts with the bound metal ion and the imidazole ring of His 98 (Figure 3.3). This contact exists in the ES complex as a polar interaction with the hydrolytic water molecule and is maintained in the free 5'-PO_3 group of the EP complex. Arg 61 does not appear to play a role in transition state stabilization, because the distance from this side chain to the scissile phosphate before bond cleavage is too long, at 5.5 Å. Arg 61 does, however, appear to stabilize the final product complex, and thus may help to drive the reaction forward by inhibiting re-ligation on the enzyme.

Methods

The overexpression and purification of wild type I-*Ppo*I have been described(Flick et al., 1997). The H98A mutant was made with the Transformer^{TM} Site-Directed Mutagenesis Kit (Clontech) and was expressed and purified identically to wild type. CHELEX 100 Resin (Bio-Rad Laboratories) was used to remove divalent cations in the metal-free crystallization. For this latter experiment, the protein and DNA were dialyzed extensively against buffer containing solid CHELEX resin, as were all crystallization buffers. The relative cleavage activities of the wild type enzyme, the H98A mutant and the wild type enzyme in the absence of divalent cations were determined by digestion of circular plasmid and linearized substrates (M. Chadsey; unpublished results). Both the H98A substitution and the removal of divalent cation inhibit cleavage. The crystallization of the enzyme in complex with the cleaved product DNA and divalent cations has been described(Flick et al., 1997). Under similar conditions cocrystallization of the wild type protein and DNA in the absence of divalent cations, or of the H98A mutant and DNA in the presence of divalent cations yield uncleaved substrate complexes. The double-stranded DNA 21-mer (5'- TTGACTCTCTTAAGAGAGTCA-3') (Oligos Etc.) was used in all three cocrystallizations. The H98A mutant crystals were obtained by the hanging drop method at 4°C. The protein/DNA solution contained I-*Ppo*1 (~4.0 mg ml^{-1} in 100 mM NaCl, 50 mM Tris (pH 8.0), 5% glycerol, 10 mM $MgCl_2$, 5 mM DTT), DNA (~3.0 mg ml^{-1}), 5 mM spermine, and 2.5 mM EDTA. The drop consisted of 2 μl protein/DNA solution mixed with 2 μl of the well solution (100mM Citrate (pH5.6), 24% PEG 3350, 20 mM NaCl, 2 mM EDTA). 200 μl of Al's oil (50% silicon oil/50% paraffin oil; Hampton Research) was added to the top of the well. The metal free crystals were obtained by the sitting drop method at 4°C. To remove trace amounts of Mg^{2+} from the protein solution, 200 μl of I-*Ppo*1 was dialyzed overnight against 50 ml of 100 mM NaCl, 50 mM Tris (pH 8.0), 5% glycerol, 5 mM DTT, 5 mM spermine, 2 g CHELEX® 100 Resin. The dialysis was performed multiple times to insure a metal free solution. Similarly, 20 μl of DNA was dialyzed against ~15 ml of chelexed TE (10 mM Tris (pH 7.4), 1 mM EDTA). Plastic microbridges (Hampton Research) washed in chelexed water were used to avoid trace metals that may accompany siliconized cover slips. The protein/DNA solution contained dialyzed I-*Ppo*1 (~4.0 mg ml^{-1}), dialyzed DNA (~3.0 mg ml^{-1}), 5mM spermine, 2.5 mM EDTA. The sitting drop consisted of 2μl protein/DNA solution mixed with 2μl of the CHELEX® treated well solution (100mM Citrate (pH 5.8), 26% PEG 3350, 20mM NaCl, 2mM EDTA). 200μl of Al's oil was added to the top of the well. All complexes crystallized in $P3_121$ with unit cell dimensions a = b = 113.9 Å, c = 89.0 Å; the asymmetric unit is an enzyme dimer and its bound DNA palindrome. The data for both H98A and the metal-free complex were collected from cryo-cooled crystals, on an in house RAXIS IV imaging plate area detector and at beamline 5.0.2 at the ALS (λ =

0.9792 Å), respectively. Data for Mn^{2+}-substituted structures were collected by soaking wild type crystals in 100 mM $MnCl_2$ for 1 hour and then collecting a data set at beamline 5.0.2 at ALS, using an incident wavelength of 1.89 Å (6.54 keV), corresponding to the anomalous edge of the manganese ion. For this data set, a full quadrant of reciprocal space (90°) was collected to maximize redundancy; and the crystal was oriented with the three-fold crystallographic symmetry axis offset by ~20° from the X-ray axis to increase the percentage of Friedel mates measured on the same exposure.

All data sets were processed using the DENZO/SCALEPACK program suite (Otwinowski and Minor, 1997). For the manganese-soaked data set, isomorphous and anomalous difference maps were calculated after an initial round of refinement with all metal and solvent coordinates removed from the model. The refinements of all structures were performed using the X-PLOR package (Brunger, 1992). Deoxyribonucleotide and phosphate atoms that span the cleavage site and all water and cation atoms were omitted from the model before initial refinement cycles and map calculations. To calculate an R_{free} during refinement, 6% of reflections were set aside before any refinement cycles (Brunger, 1993). Data and refinement statistics are shown for all data sets in Table 3.1.

Coordinates

Coordinates of the sodium substituted (1CZ0) and H98A (1CYQ) uncleaved substrate complexes have been deposited in the Protein Data Base along with the product complex (1A73).

Table 3.1: Crystallographic Data and Refinement Statistics

	H98A/Mg^{2+}	**WT/Na^{+}**	**WT product**
Data			
Source	Anode	5.0.2 ALS	5.0.2 ALS
Resolution (Å)	1.93	2.10	1.80
Wavelength (Å)	1.54	0.98	1.00
Unit cell (Å)	a=b=113.9	a=b=113.9	a=b=114.0
	c=89.0	c=89.0	c=89.0
Measured reflections	171353	206866	362336
Unique reflections	49539	38078	71892
R-merge (last shell) (%)	4.0 (30.2)	6.7 (17.8)	4.9 (26.7)
Completeness (last shell)(%)	98.3 (82.2)	96.9 (97.8)	99.8 (98.1)
Refinement			
R-factor (%)	20.4	19.1	22.6
R-free (6%) (%)	23.5	21.6	24.7
Number of atoms	3702	3744	3751
Number of waters	362	396	403
R.m.s.d. bond length (Å)	0.005	0.006	0.011
R.m.s.d. bond angles (deg.)	1.15	1.18	1.70
R.m.s.d. impropers (deg.)	1.03	1.04	1.04
R.m.s.d. dihedrals (deg.)	27.0	27.3	27.6
Mean B-value, overall (Å^2)	18.6	21.9	19.8
Mean B-value, protein (Å^2)	17.0	20.6	15.5
Mean B-value, DNA (Å^2)	20.6	22.8	17.3
Mean B-value, solvent (Å^2)	25.5	28.1	20.7
Mean B-value, cations (Å^2)	Mg^{+2}, 17.9	Na^{+}, 21.8	Mg^{+2}, 19.4

Chapter 4

Computational Modelling of the I-*Ppo*I Mechanism

Introduction

To fully understand how an enzyme catalyzes the chemical transformation of substrate to product, one needs to know the following: the path of the reaction or the reaction coordinate, the free energy of every state along this path, and the contributions of individual active site components to the energy of each state.

The best description of the reaction coordinate requires knowing the structures of all states on the minimum energy path (MEP) between the enzyme-substrate complex (ES) and the enzyme-product complex (EP). The MEP completely defines the type of reaction that takes place. For example, a nucleophilic substitution reaction is associative (S_N2) if the MEP describes the addition of the nucleophile prior to the departure of the leaving group, but is dissociative (S_N1) if the reverse occurs. Experimentally, one can determine the ES and EP complexes along with the structures of any stable intermediates that might occur in-between through the application of a wide array of techniques (chemical trapping, physical trapping, and rapid Laue data collection) collectively known as time-resolved crystallography. The unstable, high-energy transition states that lie in-between stable states are much more difficult to access experimentally since they have lifetimes on the order of a bond vibration (10^{-13} s). Kinetic isotope effects (KIE) can yield insights into these fleeting structures, but these experiments are rarely carried out due to their difficulty.

Computational techniques can be used to test theories of enzyme catalysis. If we understand an enzyme mechanism, we should be able to predict the rates of the chemical steps for the WT enzyme and the effects of mutations on these rates. If this is not the case, the theory we are using is wrong or incomplete and needs revision. However, if calculations do match well with experiment, additional information about mechanism can be gained. Specifically, the energy of each state (stable or unstable) along the reaction coordinate can be determined and the catalytic roles played by active site (and surrounding) residues can be probed more subtly than is possible with mutational analyses. There are various computational methods for studying enzyme catalysis, but they are all constructed by considering two issues. One issue is the method of searching the energy landscape on which the reaction takes place to determine the reaction coordinate. The other issue deals with the nature of the energy function that defines the landscape itself.

As mentioned before, the most detailed definition of reaction coordinate is the MEP that connects ES and EP. There are two kinds of methods that are used to try to determine such a path computationally. One group of methods slowly changes the system from the initial structure to the final structure along a chosen path while allowing other degrees of freedom to relax. These methods are known as drag methods since the system is dragged along a predetermined path. Different paths are tried and the lowest energy one is picked as the best reaction coordinate. A drawback of this method is that paths that are not thought of before the simulations are never considered and the minimum energy path is not a guaranteed outcome. The other approaches are based on a computational technique known as the nudged elastic band (NEB). With the NEB, initial guesses for paths are made, but then a minimization procedure is used that allows the path to change. If the initial path goes over a large energy barrier and a nearby path is lower in energy, the minimization procedure will move the system towards the lower energy path. When the minimization converges, one is left with the MEP. One potential drawback of this method is that the minimization is performed with respect to an energy landscape and not a free energy landscape. There are theories (harmonic transition state theory; hTST) that one can use within the NEB formalism that allow for an estimate of the entropy based on the shape of the landscape nearby the MEP, but validity of these techniques depends on the degree to which atomic motions orthogonal to the MEP are harmonic.

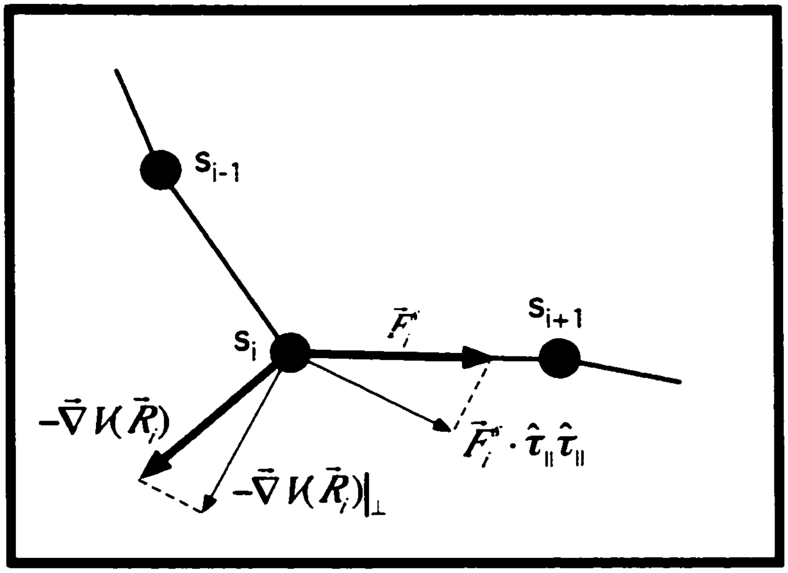
Once the method of determining the path has been selected, an energy function must be chosen. There are three types of energy functions. The first group are classical, molecular mechanical functions like CHARMM (Brooks et al., 1983), AMBER (Cornell et al., 1995), and ENCAD (Levitt et al., 1995). These potentials are often used when the transition of interest does not include changes in bonding. In studies of enzymes, covalent bonds are often being formed and broken and quantum mechanics must be used to accurately describe these events. The most common technique currently is to use hybrid energy functions such as the empirical valence bond (EVB) (Warshel, 1991) or other quantum mechanical/molecular mechanical (QM/MM) potentials. The idea here is to define a small subset of atoms as quantum atoms while treating the rest of the system classically. There are some tricky problems with this technique such as how to treat the bonds between quantum atoms and classical atoms and how to treat the non-bonded interactions of atoms across the QM/MM boundary, but this technique offers a nice compromise between using quantum mechanics and still including a large number of atoms in the calculations. The third type of energy function is a purely quantum mechanical one. The advantage here is that there are no strange boundaries where the nature of the energy functions change. There are several different approaches to quantum calculations ranging from semi-empirical potentials (AM1) to *ab-initio* methods such as Hartree-Fock (HF) and density functional theory (DFT). The drawback is that *ab-initio* quantum mechanical calculations are computationally demanding. This fact limits the size of the system one can consider in a reasonable amount of time. Currently, the system is limited to around 100 non-hydrogen atoms for DFT calculations. Semi-empirical potentials are much less expensive, but are only useful for compounds closely related to the training set that was used to parameterize them.

Examples of quantum mechanical studies on enzymes are growing in number and as computer power continues to increase, they will become even more prevalent. Arieh Warshel's group has been developing and performing calculations with the EVB potential coupled to free energy perturbation for two decades. An example of this work is their study of Ras GTPase (Schweins et al., 1995). The laboratory of Martin Karplus has recently studied the catalytic mechanism of uracil-DNA glycosylase with QM/MM methodologies (Dinner et al., 2001). Both Peter Kollman's and Thomas Bruice's labs have performed QM/MM studies on methyltransferases (Lau and Bruice, 1999; Kollman et al., 2001). Finally, Richard Friesner's group has performed DFT minimizations on the active site substructure of methane monooxygenase to investigate the structures of experimentally determined intermediates in the reaction cycle (Guallar et al., 2002).

The Nudged Elastic Band

The nudged elastic band method allows one to determine the minimum energy path (MEP) between two local minima on a potential energy surface (Jonsson et al., 1998). Harmonic transition state theory (Vineyard, 1957) allows the MEP to be equated with the free energy reaction coordinate and any local maxima along the MEP represent transition states between two stable intermediates in the reaction. The model system that we will use to test this methodology for protein catalysis is the endonuclease reaction catalyzed by the homing endonuclease, I-*Ppo*I. Using time-resolved crystallography, we have solved the structures of both substrate and product forms of the protein-DNA complex to high resolution (Galburt et al., 1999). Presumably, these crystal structures represent two local energy minima on the potential energy surface for this reaction. The NEB method can now be used to find the lowest energy reaction path between these two structures. The first step is to link the two end point structures with intermediate structures that lie on a line in the multidimensional space defined by the degrees of freedom in the reaction. This "chain of states" is then minimized according to the chosen energy potential subject to an additional constraint. This constraint attaches each state to its neighbor with a spring of natural length equal to the initial separation of states and spring constant *k*. This fictional force keeps the states well spaced on the path and prevents all of the states from sliding down the energy landscape into local minima.

This improves the sampling of unstable, high energy, transition states along the path. To make the technique even better at converging on the true MEP, only the spring forces parallel to the current path are considered at each minimization step. This allows the path to relax according to the energy landscape alone while still keeping the states evenly spaced along the pathway (Figure 4.1). The result will be, in one step, the free energy profile along a well defined reaction coordinate for the endonuclease reaction catalyzed by I-*Ppo*I.



$$\underbrace{\vec{F}_i^{nudged}}_{\text{nudged force}} = \underbrace{-\vec{\nabla}V(\vec{R}_i)|_\perp}_{\text{component of potential force perpendicular to path}} + \underbrace{\vec{F}_i^s \cdot \hat{\tau}_\parallel \hat{\tau}_\parallel}_{\text{component of spring force parallel to path}}$$

Figure 4.1: Calculation of the "nudged" force. The force on state s_i is calculated by taking the vector sum of the component of the potential force ($-\nabla V$) perpendicular to the current path (defined by the neighboring states, s_{i-1} and s_{i+1}) and the component of the spring force (F^s) parallel to the path tangent ($\tau_\parallel$).

This methodology will allow us to answer interesting questions about the relative energetic importance of particular features of the I-*Ppo*I active site including the magnesium ion and its strained coordination sphere, the proposed general base histidine, a phosphate-arginine salt bridge, a backbone carbonyl-histidine hydrogen bond, and well ordered water molecules found in the active site. In addition to answering specific questions about active site features, the method provides us with snapshots of unstable states along the reaction coordinate that are simply not observable experimentally. This information is of great importance to our understanding of protein catalysis as these unobservable transition states affect catalytic rates far more than the observable states of a reaction.

Methods

Nudged elastic band calculations were performed for the enzyme, I-*Ppo*I. The crystallographic structures of the substrate and product complexes described above (Chapter 3) were used as the initial and final states of the band respectively. To be able to complete the calculation in a reasonable time, we needed to trim the structures down to around 100 non-hydrogen atoms. This was done by trying to include all residues and atoms that were believed to be important for catalysis. Hydrogen atoms were added to these structures manually, resulting in a cluster size of 164 total atoms. Because crystallography at 2 Å does not allow for the visualization of hydrogen atoms, we were forced to place hydrogens manually and hope that the minimization procedure would correct erroneous placements. In the substrate complex, H98 was left in a neutral state with a protonated Nε forming a hydrogen bond with the backbone carbonyl from C105 and all of the water molecules in the inner coordination sphere of the metal ion were modeled as neutral. In the product complex, a hydrogen from the nucleophilic water was transferred to the Nδ of H98, one of the terminal 5'-phosphate oxygens was protonated, and the leaving group oxygen was left unprotonated. Atoms on the exterior of the cluster were frozen and the endpoint structures were then minimized with DFT. Six intermediate images were created by interpolating linearly between the inital and final states. These images were linked together according to the prescription of the elastic band method. Band calculations were begun with a WT active site, with a sodium ion, and with the H98A mutant strucure. Due to constraints on both time and resources, only the WT and sodium substituted calculations were pursued.

Results To Date

WT band

Depending on our assignment of hydrogen positions on the 5'-phosphate, two different WT paths have been identified. In path 1, a water molecule moves into the magnesium coordination sphere before the nucleophile attacks the phosphorous and the phoshate rotates away from the metal. In path 2, these events are reversed with nucleophilic attack occurring first. Each path was pursued until it was clear that path two was of lower energy and closer to the MEP. Path 2 is close to being fully minimized (forces ~ 0.2 eV/Å) and the following results and conclusions are based on this path..

To date, the WT band calculation reveals an energy barrier of 1.6 eV (37 kcal/mol) (Figure 4.2d). The energy diagram indicates the existence of an intermediate state before the transition state (Figure 4.2b). The structure of the intermediate state is almost identical with that of the substrate state (Figure 4.3). The largest differences between the initial states and image b are found in the positions and orientations of two water molecules. In particular, the oxygen atom of the nucleophilic water molecule moves 0.9 Å into a position more suitable for attack of the phosphorous. The water molecule closest to the nucleophile moves 3 Å away and joins a water network towards the periphery of the cluster. The angle of attack of the nucleophile (leaving group oxygen-phosphorous atom-nucleophilic oxygen) in the original minimized substrate image is 138°, but in the intermediate structure this increases to 153° making attack more favorable. It is important to note that in the crystal structure of the substrate complex, the angle of attack is 168°. The discrepancy between the the experimental and DFT minimized angles is likely another consequence of having to use such a small cluster of atoms in the calculations.

Image d represents the structure of the transition state. The phosphorous atom is 3.5 Å away from the leaving group, indicating that cleavage has already occurred. The now terminal phosphate has moved to within 3.2 Å from R61 (compared to 2.4 Å in the product state). The non-bridging phosphate oxygen that was coordinating the metal ion is now 4.2 Å away and interacts with the now protonated H98. The metal coordination sphere is severely distorted and contains only five members. The water molecule that forms the eventual sixth metal ligand is located 3.5 Å away (compared to 2.0 Å in the product state) and is approximately 45° away from an octahedral geometry.

Lastly, the sixth image of the band finds an energy minimum that is lower than the final state (Figure 4.2f). In this state, a proton from a metal-coordinated water molecule transfers over to the leaving group oxygen, leaving a metal-coordinated hydroxyl. This lends support to our hypothesis regarding the protonation of the leaving group by a metal bound water and allows us to treat the sixth image as the final state (Figure 4.3f).

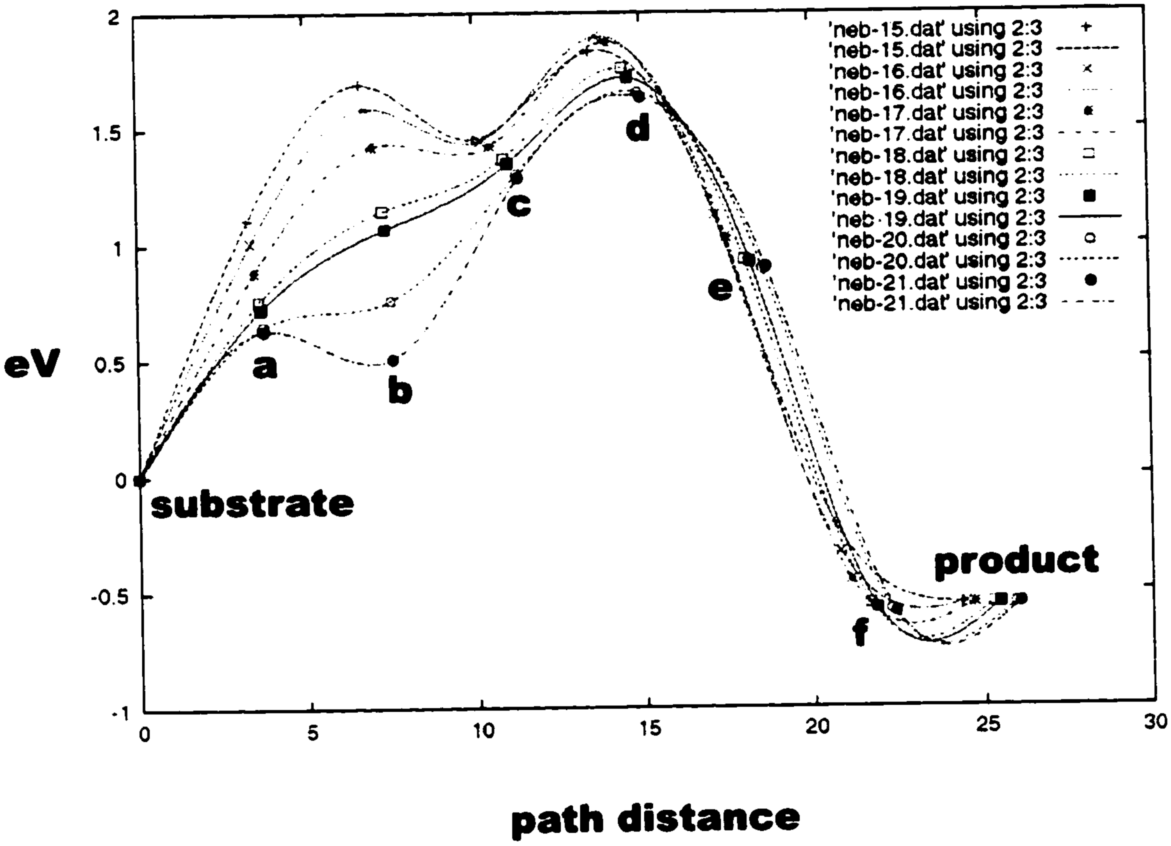


Figure 4.2: The energy diagram calculated for the WT active site. The y-axis is energy (eV) and the x-axis describes the distance along the determined path such that as the images move from left to right, the system moves from substrate to product. The actual curvatures of the traces are determined by considering the forces on each image. The force (the first derivitative of the energy) serves as an estimate for the slope of the energy curve. The different traces represent the path at different times during the minimization. Image b shows the largest change in energy during the procedure. Initially, this image was located at a local maximum transition state, but it currently occupies a local minimum intermediate state. Image d is closest to the transition state and image f turns out to be of lower energy than our originally modeled product state. Minimizing image f outside the context of the band calculation reveal that its energy is significantly lower than that of the initial product state, indicating that it should be considered the true end of the path.

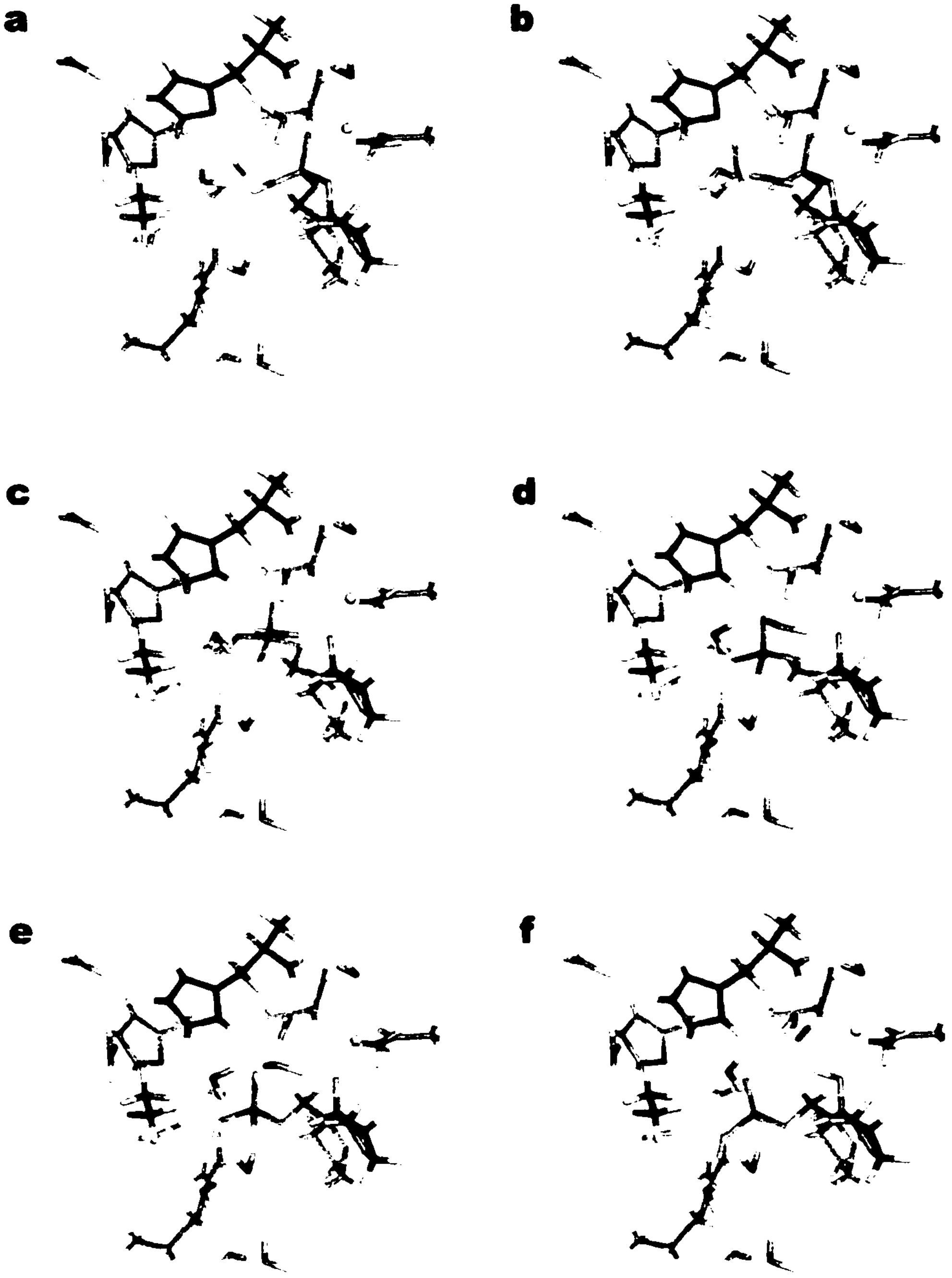


Figure 4.3: The structures of the six images that make up the elastic band (not including the substrate or product endpoints). The magnesium ion is in yellow and the nucleophilic water molecule is orange. Labels a-f correspond to the labels on the energy diagram in Figure 4.2: (b) stable intermediate; (d) transition state; (f) lower energy endpoint.

Sodium Substituted Path

The band calculation with the magnesium substituted with sodium reveals a much higher energy barrier to cleavage. This band is still being minimized so conclusions regarding this calculation would be premature.

Conclusions

Nudged elastic band calculations with clusers of 164 atoms representing the active site of I-*Ppo*I have been completed. The resulting energy profiles show a transition state barrier, a potential intermediate, and reveal the protonation state of the leaving group in the product complex. The WT path suggests that bond cleavage occurs before the transition state and that the energy barrier is due to the reorganization of the metal ion coordination sphere. The trends in calculated energy barriers agree with experiment with a lower barrier for magnesium compared to sodium. While the calculations seem to have captured qualitative components of the reaction mechanism, the quantitative energy values do not agree with experimental values. Based on the optimal rate of the steady-state reaction catalyzed by I-*Ppo*I (2 min^{-1}) (Wittmayer and Raines, 1996), the Arrhenius equation ($k = Ae^{-E/RT}$), and a pre-factor of 6×10^{12}, the activation energy must be lower than ~0.8 eV which means that our calculated activation energy is too high by at least a factor of four.

Future Directions

There are several obstacles that stand in the way of this technique becoming more accurate, more efficient, and of general use for macromolecular systems. The largest hurdle is the current computational speed for quantum mechanical calculations using DFT. Even at a cluster size of 164 atoms, minimizations took weeks and band calculations months. The small clusters that are required for the calculations to be completed in a reasonable time are also problematic. Critical interactions are likely left out and the definition of the important residues is skewed by current hypotheses, thus limiting the discovery of new information. This makes it difficult to test hypotheses about residues that are farther removed from the active center and severely limits the interpretation of any results.

Future studies will surely benefit from faster computers, but there are at least two other possibilities for improving the methodology and dealing with the problem of small cluster size. One possibility involves representing the interactions of atoms that are not included in the explicit calculation by an electrostatic field. The cluster of atoms could then be subject to forces according to the field and the total energy of the system would reflect the interactions of the cluster with its surroundings. The second possibility is to select a less computationally expensive potential to use in concert with the NEB technique. While the ultimate goal is to understand enzyme catalysis on a quantum mechanical level, fully *ab initio* calculations are currently too slow to be used with large macromolecular systems. Semi-empirical potentials like EVB, AM1, and PM3 that have been parametrized for protein systems could still yield interesting results and would run much faster with larger systems.

Another hurdle related to the length of time the band calculations take has to do with the initial guess for the path. The linear interpolation is taken as an initial path and represents the best guess when nothing is known about the intermediate states. However, this resulted in bond lengths and angles changing drastically from ideal values. These deviations from ideal geometry are not likely to contribute to the path and they cause the minimization procedure to take up the majority of its time simply optimizing bond lengths and angles for atoms that are not directly involved in the active site chemistry. For example, for a rotating group like the 5'-phosphate for I-*Ppo*I, the linear interpolation causes severe geometry and large forces as each atom proceeds linearly from its starting position to its final position. Initial band images have atoms passing very close to one another or even passing through each other! To remedy this, future calculations should initially introduce the same geometrical constraints that are used in X-ray crystallographic refinement. After the intermediate images on the band have been relaxed according to these restraints, the full NEB calculation could be run without restraints and much unnecessary computing time could be avoided.

Chapter 5

DNA Binding by I-*Ppo*I

Introduction

Many interactions between DNA-binding proteins and their targets involve a conformational change in one or both molecules. For non-enzymatic proteins such as transcription factors or histones, changes in protein and DNA structure facilitate high affinity binding, alter DNA accessibility, bring bound proteins into close proximity, or allow the assembly of multi-protein complexes at specific DNA target sites. Within these complexes, the perturbation of protein or nucleic acid may range from subtle to extreme. The DNA may remain essentially unperturbed while the protein undergoes a conformational change upon binding. Examples include bHLH transcription factors (Ferre-D'Amare et al., 1993; Ferre-D'Amare et al., 1994), arc repressor (Bowie and Sauer, 1989; Raumann et al., 1994), and λ repressor (Clarke et al., 1991; Beamer and Pabo, 1992). For other systems, protein binding results in significant distortions of the DNA. Proteins that bind primarily within the DNA major groove can induce DNA bend angles that range from 40° (λ Cro) (Brennan et al., 1990) to 90° (CAP) (Schultz et al., 1991). TATA binding protein (TBP) binds to the minor groove and induces an 80°-100° bend in its substrate by inserting phenylalanine residues between nucleotide bases (Kim et al., 1993; Kim et al., 1993). Integration host factor (IHF) disrupts base stacking via intercalating prolines and generates a 160° DNA bend (Rice et al., 1996). Chromatin DNA is wrapped around the nucleosomal core particle in a left-handed superhelix, with half of the binding interactions formed between protein main-chain and DNA backbone atoms (Luger et al., 1997).

Many enzymes that act on nucleic acid substrates also induce or undergo conformational changes. These perturbations contribute to catalysis by aligning reactive functional groups in the active site, by increasing binding energy in the transition state complex, or by inducing localized strain in the substrate. Cre recombinase (Guo et al.,

1997) catalyzes homologous recombination by binding to loxP sites and bending the DNA by approximately 100°. Methyltransferases such as *Hae*III (Reinisch et al., 1995) and *Hha*I (Klimasauskas et al., 1994) induce a conformational change in their substrates by flipping a cytosine nucleotide out of the DNA helix. Similarly, uracil-DNA glycosylase is a base excision enzyme that flips out and removes uracil residues (Slupphaug et al., 1996). Topoisomerases induce changes in DNA structure in order to resolve topological strain in the double helix. The type II family of topoisomerases pass an intact double helix through a double stranded break in another strand (Berger et al., 1996), while members of the type I family make a single-stranded break, allow the DNA to undergo a rotation along the helical axis, and then re-ligate the broken strand (Redinbo et al., 1998; Stewart et al., 1998).

Restriction endonucleases catalyze double stranded breaks in DNA and are believed to use lateral diffusion along the double helix to search out their target sites efficiently. Type II restriction endonucleases and their bound DNA exhibit a wide range of conformational changes when the correct cognate sequence is encountered. *Eco*RV binds non-cognate DNA sequences in an unperturbed B-form conformation, but induces a 50° bend in its specific restriction sequence (Winkler et al., 1993). The structure of the *Eco*RI-DNA complex reveals a kink of 25° in the DNA substrate (McClarin et al., 1986), whereas smooth bending is observed in DNA bound to *Bgl*II (23°) (Lukacs et al., 2000) and *Bgl*I (20°) (Newman et al., 1998). In contrast, cognate DNA sites are bound in a B-form conformation by both *Bam*HI (Newman et al., 1995) and *Pvu*II (Athanasiadis et al., 1994; Cheng et al., 1994). *Bgl*II and *Bam*HI share similar recognition sequences, yet bind their respective substrates in different conformations (Athanasiadis et al., 1994). A range of protein conformational changes have been observed upon restriction endonuclease binding. *Bam*HI and *Eco*RV show a wide range of changes upon binding including a rotation of subunits, the folding of disordered regions, and the unfolding of C-terminal α-helices (Winkler et al., 1993; Newman et al., 1995). *Pvu*II displays a conformational change that results in the formation of a histidine-histidine hydrogen bond by completely encircling the DNA (Cheng et al., 1994).

Homing endonucleases are encoded by open reading frames nested within mobile introns or inteins, and recognize lengthy DNA target sites (14 to 40 bp). Cleavage of these sites initiates 'homing': a high-frequency, site-specific gene conversion event that transfers a copy of the intervening sequence to a specific insertion point at or near the cleavage site (Dujon, 1989; Lambowitz and Belfort, 1993; Mueller et al., 1993; Belfort and Perlman, 1995; Belfort et al., 1995; Belfort and Roberts, 1997; Lambowitz et al., 1998; Jurica and Stoddard, 1999). Homing of Group I introns requires homology between the donor and recipient alleles. Their homing endonucleases have been grouped into four distinct families based on conserved sequence motifs: LAGLIDADG, His-Cys Box, GIY-YIG, and HNH. I-*Ppo*I was the first identified member of the His-Cys box family and is encoded by an open reading frame in the third intron of the nuclear, extrachromosomal 26S rRNA gene of the slime mold *Physarum polycephalum* (Muscarella and Vogt, 1989; Muscarella et al., 1990). The His-Cys box sequence motifs are responsible for forming two structural zinc binding sites per monomer and the enzyme is dependent on divalent cations (usually magnesium) for activity. I-*Ppo*I is a homodimer that cleaves a partially symmetric 15 base pair homing site, generating 4-base, 3' cohesive overhangs of sequence 5'-TTAA-3' (Lowery et al., 1992; Ellison and Vogt, 1993; Wittmayer and Raines, 1996).

Unlike restriction sites, homing sites may vary in sequence at many of their individual nucleotide positions while still being recognized and cleaved. I-*Ppo*I tolerates base pair substitutions at several positions within the homing site, but exhibits a strong preference for A-T base pairs in the central four positions of the homing site (Argast et al., 1998). The structures of I-*Ppo*I in complex with its DNA substrate and product have been determined at high resolution (Flick et al., 1998; Galburt et al., 1999). The homing site is severely bent in the protein-DNA complex, resulting in a sharp bend at the cleavage site. In the complex, leucines 116 and 116' are observed to make edge-on contacts to the adenine nucleotides directly at the primary bend sites. The conformational changes observed in this enzyme-DNA complex appear to be important for proper alignment of scissile phosphates with catalytic groups in the enzyme active site. These conformational changes may also promote binding and catalysis by maximizing surface complementarity, binding energy, and sequence-specific contacts between the enzyme and DNA. In this chapter I describe the results of structural and functional analyses of I-*Ppo*I, designed to determine the degree to which the bound complex represents an induced fit of enzyme and DNA substrate and to determine the role of Leu 116 in binding and catalysis.

Results

Protein Conformational Changes

I-*Ppo*I displays an unusual dimeric structure, with the final eighteen residues of each monomer (146 to 163) forming a domain-swapped C-terminal tail that wraps around the surface of the opposite subunit. The structure of individual I-*Ppo*I monomers, excluding these domain-swapped tails, are unchanged when comparing the unbound and DNA-bound proteins (r.m.s.d = 0.7 Å). The overall fold of the subunits are identical and all of the secondary structural elements are intact. The DNA-binding b-sheets are well formed in the absence of DNA, and can be independently superimposed from the same two structures with an r.m.s.d. of 0.8 Å. However, a rigid body rotation of the monomers relative to one another is observed when the bound and unbound protein structures are superimposed (Figure 5.1). The DNA binding sheets are approximately 5 Å closer together in the bound protein-DNA complex than in the unbound structure. This is a result of a hinge motion of 5° that changes the relative orientation of the two monomers. This motion involves a smooth, incremental series of protein backbone rotations between residues 144 and 148, near the dimer interface and directly preceding the domain-swapped tails of each subunit. When single monomers from each structure, excluding their C-terminal tails, are superimposed as shown in figure 5.1, the r.m.s.d between the unrestrained subunits is 4.4 Å, with deviations increasing from 1 Å close to the dimer interface to greater than 7 Å at the distal end of the DNA binding β-sheet.

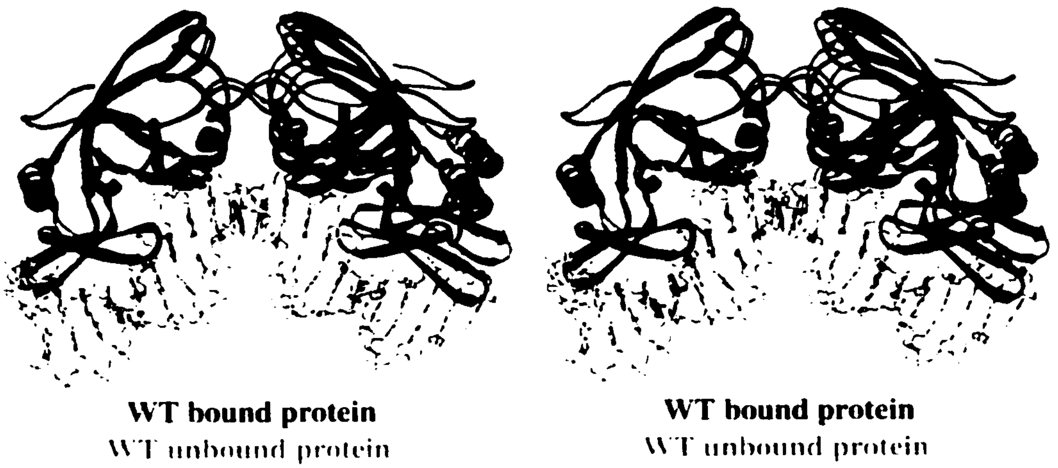


Figure 5.1: An alignment of the structure of the wild-type protein/DNA complex (1A73) with the structure of the wild-type protein alone (1EVX). The structures were aligned using only the DNA-binding β-sheet on the left. The bound protein is shown in blue and the unbound protein is shown in red. The bound DNA homing site is colored according to atom type.

The two structural zinc atoms are present in each monomer, but the catalytic magnesium atoms are not seen in either active site. This was expected, as two DNA oxygen atoms and only one uncharged protein ligand (Asn 119) directly contribute to the coordination of the active site metal in the complex. A bound magnesium-water cluster in a similar active site is observed in an apo structure of *Serratia* nuclease when the crystal is soaked in high concentrations of magnesium chloride (Miller et al., 1999). The active site residues (His 98, Asn 119, Cys 105, Arg 61) overlay very closely in the aligned monomer (r.m.s.d. = 0.9 Å); however, the positions of the two active sites relative to one another are significantly different as a result of the 5° hinge motion between monomers.

DNA Conformational Changes

DNA bending induced by binding of wild-type I-*Ppo*I to its homing site was measured using a circular permutation assay (data not shown). The bend angle calculated from subsequent analyses (see methods) is 81.7° ± 2.4°. No homing site DNA bending was detected in the absence of I-*Ppo*I, indicating that the bent conformation of the DNA is induced upon protein binding.

Effects of the L116A Mutation

The positioning of Leu 116 near Ade2 in the protein-DNA crystal structure suggested a role for this residue during DNA binding and/or cleavage. Substitution of an alanine at this position had a dramatic effect on both of these activities. Gel mobility shift assays (Figure 5.2) revealed that the binding constant (K_d) of the L116A mutant is increased at least 5,000-fold above that of wild-type I-*Ppo*I, It was not possible to assign an absolute Kd to L116A by this method, as protein concentrations greater than 50nM caused complexes to migrate aberrantly. The poor DNA binding activity of the mutant precluded use of the circular permutation assay to estimate DNA bending. The ability of L116A to cleave a supercoiled plasmid substrate is also compromised; the mutant protein is 72,000-fold less active than wild-type I-*Ppo*I at pH 10.0 (Figure 5.3).

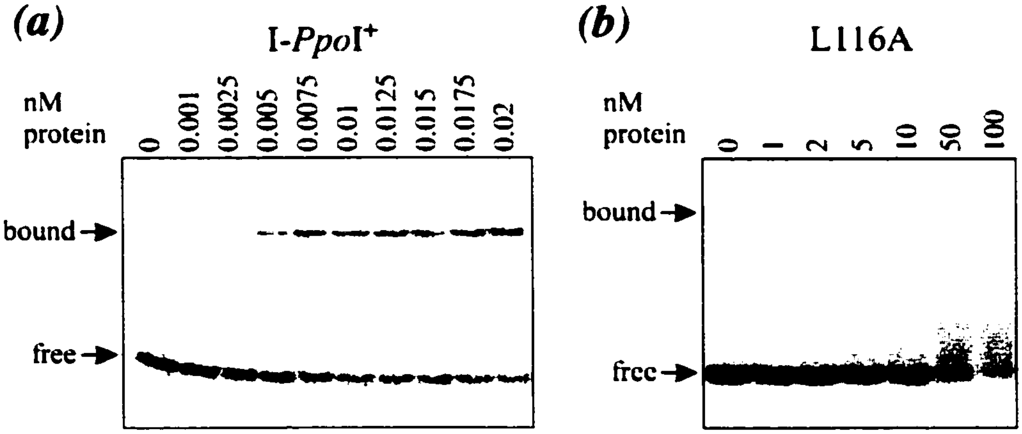


Figure 5.2: Gel mobility-shift assay of specific DNA-binding activity by (a) wild-type I-*Ppo*I and (b) L116A. Increasing levels of dimeric I-*Ppo*I (nM concentrations shown above lanes) were incubated with ^{32}P-labeled dsPpo38 oligonucleotide (2 pM in (a); 2 nM in (b)) in 25 mM Caps/Ches (pH 10.0), 50 mM NaCl, 2mM DTT, 10 mM $MgCl_2$, 0.1 mg/ml bovine serum albumin for 30 minutes at 25°C. Apparent K_d values of wild-type I-*Ppo*I and L116A are 10 pM and >50 nM, respectively. (M. Chadsey)

Figure 5.3: Agarose gels showing the relative cleavage activities of wild-type I-PpoI and L116A. Cleavage reactions (10 μl) contained the concentrations of protein indicated above the lanes (nM), 10 nM p42 plasmid substrate, 25 mM Caps/Ches (pH 10.0), 50 mM NaCl, 2 mM DTT, 10 mM MgCl2, 0.1 mg/ml bovine serum albumin. Reactions were incubated at 37C for 60 minutes before being quenched and electrophoresed. Arrows labeled r, l, and s indicate relaxed, linear, and supercoiled forms of p42, respectively. (M.Chadsey)

To study structural effects of Leu 116, the structure of an L116A/DNA complex was determined. The overall structure of the L116A/DNA complex is very similar to the wild-type complex, and individual contacts between the protein, nucleotides in the major groove, and the phosphate backbone are conserved. However, the overall protein conformation aligns better with the unbound wild-type structure (r.m.s.d. = 0.7 Å) than with the bound wild-type structure (r.m.s.d = 1.1 Å) (Figure 5.4). The difference in conformation between bound L116A and bound wild-type I-*Ppo*I is the same as the difference described above for the wild-type bound and unbound I-*Ppo*I structures. Because of the relative change in orientation between monomers, the DNA binding b-sheets of the subunits are further away from each other by approximately 5 Å. This less compact protein conformation is matched by a slightly less bent DNA duplex. The conformations of protein and DNA in the wild-type and L116A complexes are mutually exclusive.

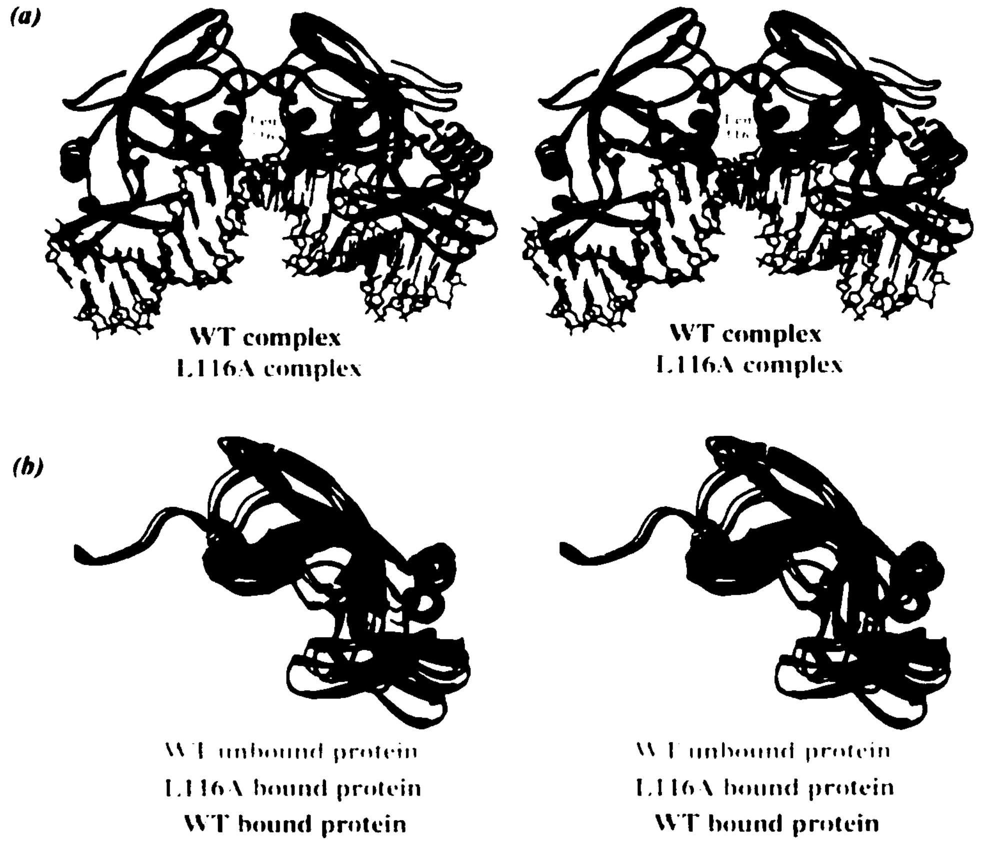


Figure 5.4: Alignments of the structures of the wild-type protein/DNA complex (1A73) and the L116A protein/DNA complex (1EVW). The alignments were performed in the same way as in Figure 5.1 by constraining only one b-sheet. (a) The wild-type complex is shown in blue with the mutated leucine residue in red van der Waals spheres and the L116A complex in green. (b) the unconstrained monomers of the unbound wild-type protein (red), the bound L116A protein from panel a (green), and the bound wild-type protein (blue) are rotated to show the extent of the conformational changes and the similarity obetween the bound L116A protein and the unbound wild-type protein.

For example, in the wild-type complex, there are 2.8 Å contacts between both Gln 63-Ade +6 and Arg 74-Gua +7. When the DNA structure from the *wild-type* complex is docked to the protein from the *mutant* L116A complex, these contacts extend to 6.1 Å and 6.2 Å respectively. However, these contact distances are still 2.8 Å in length in the mutant protein-DNA complex, indicating that enzyme-substrate complementarity is maintained across the interface of the L116A mutant despite the conformational changes.

In the wild-type complex, L116 makes an edge-on contact to Ade +2. The contact is such that these groups are not in close van der Waals contact, but their calculated solvent-accessible surfaces intersect (Figure 5.5). The closest approach between the two groups is 3.7 Å between Cδ2 of leucine and the N3 of the adenine. In the mutant complex, the average distance from the Cβ of alanine and the N3 of adenine is 6.5 Å. The L116A substitution creates a hole in the protein-DNA interface by removing three carbon atoms. This results in an increase in the solvent-accessible surface area and a decrease in the buried surface area of the interface at this position.

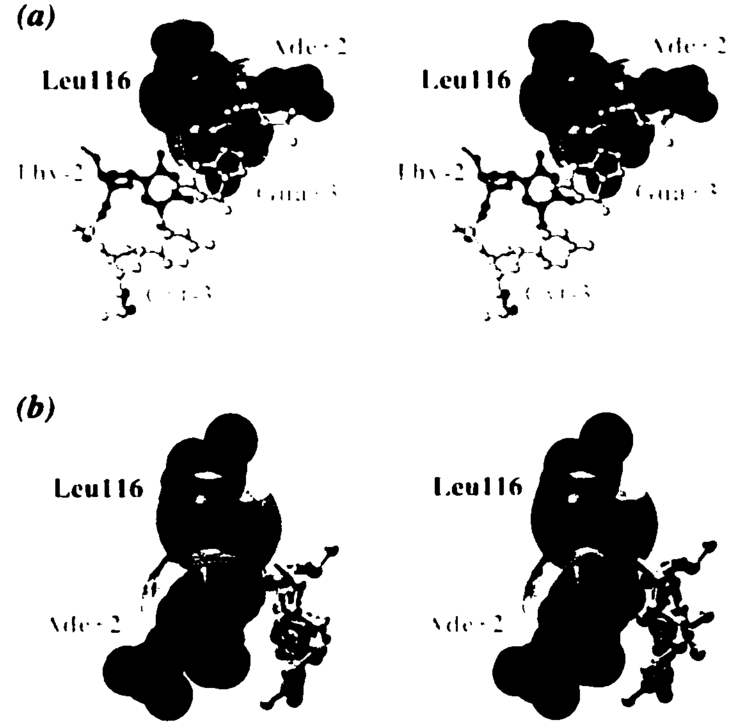
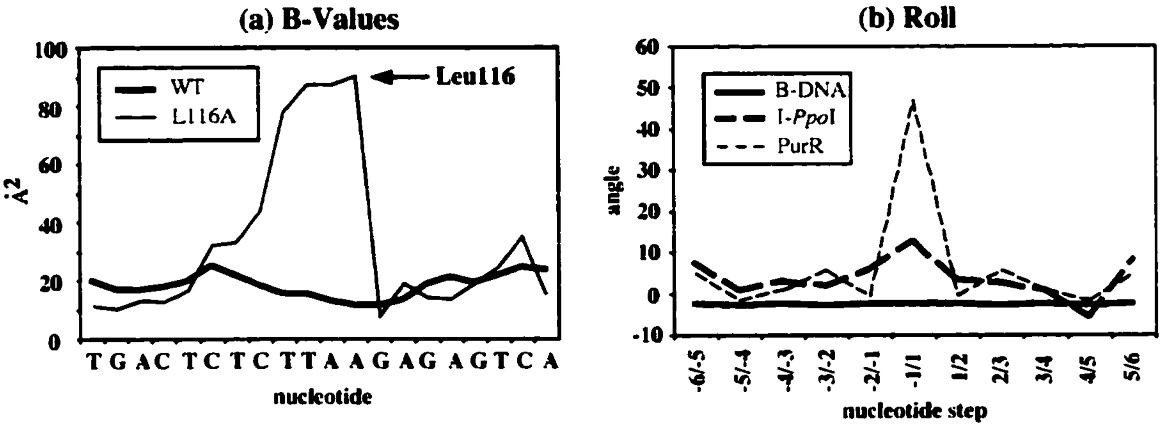


Figure 5.5: Protein-DNA contacts between L116 and A+2. Leu116 is in blue, A+2 and T-2 are in red, and the neighborind DNA nucleotides are in gray. The leucine residue and the adenine nucleotide are shown in van der Waals spheres with their corresponding solvent-accessible surface areas. The van der Waals spheres are not in contact, but the solvent-accessible surface areas intersect. (a) the large slip and slide between A+2 and G+3 is shown. (b) The structure has been rotated 90° and T-2 has been removed to show an edge-on view of the leucine-adenine interaction.

I-*Ppo*I cleaves two phosphodiester bonds, leaving 4-base, 3' overhangs. In the wild-type product complex, the four-base overhangs are well ordered (B_{avg}= 13.5 Å^2) compared to the rest of the DNA atoms (B_{avg}= 19.3 Å^2). In the L116A complex, the central four bases are extremely disordered. The poor visual quality of the map in this region was corroborated by the high group B-values for these atoms (B_{avg}= 86.9 Å^2) as compared to the rest of the DNA atoms (B_{avg}=21.5 Å^2). More specifically, the B-values increase dramatically at the adenine nucleotide that leucine 116 contacts in the wild-type structure. While the average B-value for the preceding guanine nucleotide is 7.8 Å^2, it jumps to 90.4 Å^2 for the adenine nucleotide lacking the leucine contact (Figure 5.6a).



-6 -5 -4 -3 -2 -1 1 2 3 4 5 6

5'-C T C T C T T A A G A G A G-3'

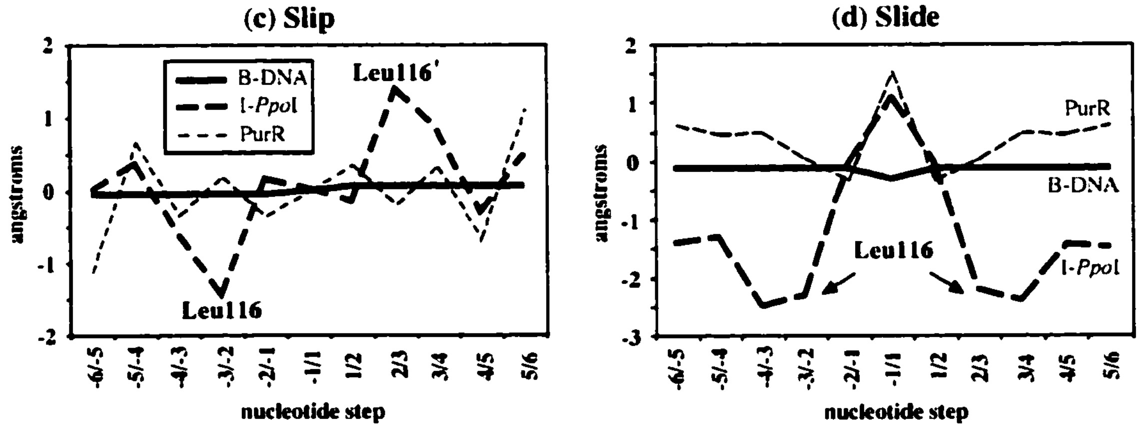


Figure 5.6: A compilation of DNA parameters calculated from the crystal structures using the program FREEHELIX (Dickerson, 1998). (a) A plot of DNA B-values for the wild-type and L116A complexes. (b), (c), and (d) Plots of DNA roll angles, slip distances, and slide distances for B-form DNA, wild-type I-*Ppo*I bound DNA, and PurR bound DNA, respectively. Roll refers to the total angle between subsequent base-pair normal vectors, slip refers to the displacement of base-pairs along their short axes, and slide is the analogous metric in the direction of the long axes. The homing site of I-*Ppo*I is showed in the middle of the figure with the numbering scheme used in the graphs.

Discussion

In the wild-type complex, a total surface area of 12,500 Å^2 is buried between I-*Ppo*I and DNA. As in many biomolecular complexes, it is likely that the burial of surface area drives the association through the free energy gain of releasing solvent molecules. For I-*Ppo*I, the formation and desolvation of the protein-DNA interface requires a severe distortion of B-form DNA to a geometry with lower twist angles and moderately high roll angles (Figure 5.6b). This distortion results in a compaction of the major groove and an expansion of the minor groove such that at the center of the homing site, the minor groove is 5 Å wider than the major groove. The nucleotide step at which L116 makes contact shows the largest deviation in base pair slip and slide, which results in the unstacking of an AT-CG step (Figure 5.6c,d). The 4-base pair stretch of DNA that lies between L116 and L116' in the complex exhibits the most distorted major and minor groove widths. As mentioned in the results, this region is well ordered in the WT-product complex and is disordered in the L116A-product crystal structure (Figure 5.6a). Based on these results, L116 appears to be important for maintaining a well ordered DNA complex at the cleavage site in which the DNA substrate is maximally deformed and for desolvating nucleotide bases that are partially unstacked in the DNA complex.

I-*Ppo*I's use of leucines to stabilize a deformed DNA conformation appears fundamentally different from other proteins that have been visualized structurally and that use hydrophobic residues in protein-DNA interfaces. We can contrast the leucine-DNA contact that I-*Ppo*I makes to the one made by the purine repressor (purR) bound to its operator sequence. The structure of the purR-purF operator complex reveals a bound DNA conformation that is somewhat similar to that of I-*Ppo*I bound DNA (Schumacher et al., 1994). An alignment of the two DNA structures shows that both have centrally compacted major grooves and expanded minor grooves, and that both helix axes are bent by similar amounts. However, a closer analysis yields insight into the different ways that proteins use hydrophobic residues to affect DNA conformation. PurR intercalates a leucine residue, L54, between the central CG-GC base step nucleotides to induce a large roll angle of almost 50° (Figure 5.6b) and a drastically reduced twist angle. In addition, the inserted leucine residue causes the individual base pairs to buckle such that they are aplanar by ~20°. This use of a protein residue as a wedge is similar to the TATA box binding protein's (TBP) use of phenylalanines to unstack the base pairs TA and CG (Kim et al., 1993; Kim et al., 1993). Bound TATA box DNA also displays large roll and buckle angles where the phenylalanine residues are positioned. In contrast, I-*Ppo*I uses leucines to make edge-on contacts with adenine nucleotides such that there is only a slight deviation in roll and twist angles from B-form. Instead, the majority of the distortion is caused by the sliding of the bases past one another.

The L116 residue of I-*Ppo*I is too far away and not properly positioned to act as a wedge, but it is close enough to locally desolvate the minor groove at the site where bases in the substrate are most dramatically unstacked in the protein complex. This local desolvation appears to stabilize the position of the adenine base in the bound complex, compensating for the energetic penalty of unstacking base pairs. The distorted DNA conformation has almost 10% more accessible surface area than B-form DNA of the same length and sequence; this increase is localized almost entirely to the central six base pairs of the homing site. To stabilize this DNA conformation and bind tightly, I-*Ppo*I must desolvate this additional surface area. Without L116, additional water molecules can enter the pocket between protein and DNA and solvate the edge of the adenine and any unstacked base pairs. The solvation of the exposed hydrophobic surface area would cause an increase in the free energy of the productive protein-DNA complex. This hydrophobic interaction can explain the experimentally observed reduction in binding affinity for L116A. This hypothesis is further supported by the observation that the central four base pairs are well ordered in the wild-type complex and completely disordered in the mutant complex. In summary, while both purR and TBP use hydrophobic residues as pry bars or wedges to lever apart stacked bases, I-*Ppo*I appears to use L116 to locally desolvate the minor groove thereby stabilizing a distorted DNA conformation and reducing the energetic cost of base-pair unstacking.

In general, I-*Ppo*I can recognize target sites with one or two base substitutions. However, randomized substrate screens reveal that the central sequence TTAA is strongly preferred by the enzyme (Argast et al., 1998). Most characterized restriction endonucleases specify their target site sequences by making all possible hydrogen bonds to the edges of the nucleotide bases that are accessible in the major and minor grooves. Interestingly, I-*Ppo*I makes no base-specific contacts to any of the four preferred central base pairs in the target sequence. The only interactions at these positions are between K120 and the O2 of Thy1 and the L116-Ade2 interaction discussed above. There are no steric clashes if other base pairs are modeled into these positions, and the hydrogen bond made by the lysine would almost certainly be conserved with a CG in place of the TA or with a switch of the base pair from TA to AT. Thus, the observed protein-DNA interactions are unable to explain the sequence conservation. The conservation of these base pairs might be due to the sequence-dependent conformational preferences of DNA or to the sequence-dependent flexibility of the DNA polymer. This dependence of binding affinity on non-contacted base pairs has been observed before for various repressor-operator complexes. The crystal structure of 434 repressor-operator complex reveals that no protein-DNA contacts are made to the inner 4 base pairs of the operator. Despite this, mutations in this region of the operator sequence can significantly alter repressor-operator affinities (Koudelka et al., 1987). More specifically, operators with an AT or TA base pair bind repressor with higher affinity than those with CG or GC. Further analysis revealed that higher affinity of the AT operators could be explained by a matching of the intrinsic twist of the DNA to the twist observed in the protein-DNA complex (Koudelka and Carlson, 1992). The affinity of the P22 operator varies similarly with the sequence of the central non-contacted bases. The central bases were observed to affect the structure (groove geometry) of both the unbound operator and the operator-repressor complex (Wu et al., 1992; Wu and Koudelka, 1993). The catabolite gene activator protein (CAP) and nucleosomes both induce severe DNA bending without base-specific protein contacts. For both proteins, sequence dependent analyses revealed that AT content is preferred where the minor groove faces the protein and GC content is preferred where the major groove faces the protein (Drew and Travers, 1985; Gartenberg and Crothers, 1988; Travers, 1989). Based on the lack of specific protein-DNA contacts to the central TTAA sequence in the I-*Ppo*I target site, it is likely that the preferred DNA sequence affects the physical characteristics of the DNA molecule such that it is able to more easily achieve the bent conformation required for high affinity binding and cleavage.

DNA bending is commonly observed in protein-DNA complexes. In some cases, the bend is important expressly for altering DNA conformation either for compaction (nucleosome core particle) (Luger et al., 1997) or for subsequent molecular recognition (CAP) (Schultz et al., 1991). In other cases, bent DNA is necessary to maximize the shape complementarity of surfaces between a globular protein and an initially linear DNA molecule. Lastly, DNA-bending enzymes may generate a DNA conformation that results in a productive catalytic alignment. I-*Ppo*I appears to bind and bend its DNA homing site for a combination of these reasons. The bent conformation increases the buried surface between protein and DNA, thus stabilizing the complex. At the same time, the bend results in the proper positioning of the substrate phosphodiester bonds relative to the two active sites and allows I-*Ppo*I to cleave across the minor groove. In B-form DNA, the scissile phosphates leading to 4 base, 3' overhangs are close together, making it difficult to surround them with two independent endonuclease active sites. The LAGLIDADG homing endonuclease family addresses this problem by tightly packing its active sites together, such that their catalytic aspartates and bound divalent cations are only 10 Å apart in the bound DNA complex. In contrast, I-*Ppo*I distorts the DNA substrate to widen the minor groove and facilitate the proper positioning of two sets of active site residues.

LAGLIDADG and His-Cys box family enzymes both cleave across the DNA minor groove (generating 4 base, 3' overhangs) and form sequence specific interfaces in the major grooves of the homing site. LAGLIDADG enzymes accomplish this by adopting a structure with tightly packed active sites. In contrast, I-*Ppo*I and its target site exhibit conformational motions that allow maximum complementarity across the protein-DNA interface (tight binding) and the alignment of scissile phosphates with more spatially separated active sites (efficient cleavage). It is interesting to note that a subfamily of His-Cys box endonucleases, typified by I-*Nja*I, have been shown to generate 5-base, 3' overhangs (Elde et al., 1999). In these cases, the homing site scissile phosphates would be further apart in unperturbed B-form DNA (15 Å instead of 10 Å), and their corresponding homing endonucleases do not contain residues that correspond to Leu 116 in I-*Ppo*I. This increased spacing of active sites and the corresponding lack of a leucine residue makes it likely that the DNA binding mode of these enzymes is different than the mode observed for I-*Ppo*I. In conclusion, even homing endonucleases within the same family appear to have evolved a variety of DNA binding modes and active site chemistries that accomplish the same biological function.

Methods

Oligonucleotides and Plasmids

DNA oligonucleotides used for the biochemical analyses (Table 6.1) were purchased from Operon Technologies (Alameda, CA). Single-stranded (ss) oligonucleotides were purified from denaturing (7 M urea) 15 % polyacrylamide gels; annealed double-stranded (ds) oligonucleotides were purified from non-denaturing 6 % polyacrylamide gels. Bands were excised, crushed, and soaked overnight in 0.3 M sodium acetate buffer (pH 5.2) containing 1 mM EDTA to elute DNA. Supernatant containing DNA was ethanol precipitated, resuspended in 1x STE (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA), and quantified by UV spectroscopy. To prepare substrate for the binding assays, a gel-purified 38 bp oligonucleotide, Ppo38^{-}, was radiolabeled on the 5'-OH group with [γ-^{32}P]ATP and T4 polynucleotide kinase (NEB), and then annealed with a 30-fold excess of the complementary oligonucleotide, Ppo38^{+}, by heating to 90°C in 10 mM Tris pH 8.0 and 200 mM NaCl followed by slow cooling to room temperature. Unlabeled ds nonspecific competitor DNA, PpoComp (chemically identical to Ppo38, but with a scrambled I-*Ppo*I homing site), was annealed prior to gel purification. pBENDPpo was constructed by cloning the DNA duplex oligomer PpoHSup/PpoHSlow into the *Xba*I site of the DNA circular permutation assay vector pBEND3 (gift from Dr. Sanker Adhya)(Zwieb and Adhya, 1994). The I-*Ppo*I cleavage assay target plasmid, p42, was a gift from Dr. Volker Vogt (Muscarella et al., 1990).

Protein Purification

The overexpression and purification of wild-type I-*Ppo*I and the crystallization of the wild-type enzyme in complex with DNA have been described (Flick et al., 1997). An I-*Ppo*I

L116A mutant was created using CLONTECH's Transformer Site-Directed Mutagenesis Kit. It was expressed and purified identically to wild-type.

Crystallography

Unbound wild-type I-*Ppo*I was crystallized in hanging drops (2 μl well + 2 μl protein at 6 mg/ml) with a well solution containing 200 mM ammonium sulfate, 18 % polyethylene glycol (PEG) 8K, 100 mM sodium cacodylate pH 6.5. The protein crystallized in the space group $P4_12_12$ with unit cell dimensions a = b = 52.8 Å, c = 278.8 Å. The crystal was flash frozen after soaking in a cryoprotectant (30 % glucose, 10 % glycerol). The asymmetric unit contains a single copy of a protein homodimer. The protein structure from the DNA complex was used as a search model for molecular replacement (EPMR) (Kissinger and Gehlhaar, 1997). During the initial rigid body refinement, each half of the homodimer was allowed to move independently, as were the DNA binding β-sheets. L166A I-*Ppo*I (4.5 mg/ml) was mixed with a DNA duplex oligomer [XtalOligo2 and its complement] as previously described (Flick et al., 1997) and the L116A/DNA complex was crystallized in hanging drops (2μl well + 2 μl of a 1:1 protein/DNA mixture) with a well solution containing 35 % MPD, 100 mM MES pH 6.5, 20 mM NaCl, 10 mM $MgCl_2$. The protein crystallized in space group C2 with unit cell dimensions a = 182.0 Å, b = 73.1 Å, c = 92.7 Å, β= 95.4°. The crystal was flash frozen directly out of the crystallization. A protein monomer combined with the extended C-terminal arm from its dimer partner was used as a search model for molecular replacement (AMORE(1979)). After finding four monomer solutions that corresponded to two dimers in the asymmetric unit, maps were calculated and a DNA model could be easily built into density for each enzyme dimer. Both data sets were collected on an in-house RAXIS IV imaging plate area detector. The DENZO/SCALEPACK program suite was used to process all data sets (Otwinowski and Minor, 1997). X-PLOR was used to perform the refinements of all structures (Brunger, 1992). To calculate an R_{free}, 6 % of reflections were set aside at the beginning of refinement for both structures(Brunger, 1993). Data and refinement statistics are shown for all data sets in Table 6.2. Both structures have been deposited into the PDB (Berman et al., 2000) (apo enzyme: 1EVX; L116A complex: 1EVW). The wild-type product complex structure has been solved and reported previously (1A73) (Flick et al., 1998).

Gel Mobility Shift Assay for DNA Binding

Gel mobility shift (GMS) assays were based on retardation of the electrophoretic mobility of a ^{32}P-labeled DNA molecule upon the binding of I-*Ppo*I(Fried and Crothers, 1981; Garner and Revzin, 1981). 0.02 fmols (I-*Ppo*I* assay) or 0.02 pmols (L116A assay) of a ^{32}P-labeled homing site oligonucleotide, ds Ppo38, were incubated for 30 min at 25°C with increasing amounts of I-*Ppo*I in 10 ml of binding buffer (25 mM CAPS/CHES pH 10.0; 2 mM EDTA; 50 mM NaCl; 10 % glycerol, and 50 µg/ml bovine serum albumin (BSA)) lacking Mg^{2+} to inhibit cleavage. Nonspecific competitor oligonucleotide, ds PpoComp, was present at a 50-fold molar excess. Samples were loaded on a 0.7 mm x 16 cm x 16 cm 10 % (75:1) polyacrylamide gel containing 2 % glycerol, and run at 175 V in 0.25x TBE for 3 hours at 4°C. Dried gels were imaged on a Storm Phosphorimager 840 (Molecular Dynamics, Sunnyvale, CA) and the intensity of the free and bound DNA bands quantified using ImageQuant Software (Molecular Dynamics). The apparent K_d the I-*Ppo*I target site complex was defined as the concentration of I-*Ppo*I at which 50% of the DNA was shifted into a complex having retarded electrophoretic mobility(Ausbel et al., 1989). The reported K_ds represent the average of three independent assays.

Circular Permutation Assay for DNA Bending

The degree to which I-*Ppo*I bends homing-site DNA upon binding was determined by circular permutation analysis, a variation of the gel mobility shift assay described above. This assay is based on the empirical observation that the mobility of a bent DNA fragment is related to the position of the bend relative to the ends of the molecule. Five different 140 bp fragments were generated from pBENDPpo by *Bam*HI, *Kpn*I, *Eco*RV, *Nhe*I, and *Bgl*II digestions for the circular permutation assay (Zwieb and Adhya, 1994). The center of the I-*Ppo*I target site is positioned 17, 32, 70, 108 and 114 bp from one end of the molecule, respectively. These DNA fragments were gel purified and labeled at their 5' ends by treatment with [γ-^{32}P]ATP and T4 polynucleotide kinase. DNA binding, gel electrophoresis, and visualization of the shifted complexes were carried out as described above, except that the concentration of DNA was increased to 0.2 nM, I-*Ppo*I was present at 0.5 nM, and nonspecific competitor was not included. The deviation from linearity, α, was calculated using the formula $\mu M/\mu E = \cos(\alpha/2)$, where μM is the migration distance of a fragment bent in the center, and μE that of a fragment bent at one end (Thompson and Landy, 1988). μM and μE were determined by measuring the distance of the protein-bound *Eco*RV and *Bam*HI fragments from their loading origin. The reported bend angle is an average of five independent assays. Bend angles of less than 10° were beyond the limit of detection of this method.

Catalytic Activity Assays

The relative catalytic activity of wild-type I-*Ppo*I and the L116A variant were measured as described (Mannino et al., 1999). Briefly, serial dilutions of enzyme were incubated with the substrate plasmid p42, which contains a single I-*Ppo*I target site, for 60 minutes at 37°C. The concentration of p42 in 10 µl reactions was 10 nM; the range of enzyme concentrations tested was 10 pM to 1 µM. Cleavage reactions contained 25 mM CAPS/CHES pH 10.0, 50 mM NaCl, 2 mM DTT, 10 mM $MgCl_2$ and 0.1 mg/ml BSA. The assay was performed at pH 10.0 where specific activity is optimal (Mannino et al., 1999). Reactions were quenched by the addition of 2x Stop buffer (2% SDS, 100 mM EDTA, 20% glycerol, and 0.2% bromophenol blue). Cleavage products were resolved from uncut DNA by agarose gel electrophoresis. The intensity of the ethidium bromide stained DNA bands was measured on a Fluorimager SI (Molecular Dynamics). The relative activities of L116A and wild-type I-*Ppo*I were determined by comparing the percentage of linear plasmid generated at each enzyme concentration using ImageQuant software (Molecular Dynamics).

Coordinates

Coordinates for the apo enzyme (1EVX) and the L116A complex (1EVW) have been deposited in the Protein Data Base along with the WT product complex (1A73).

Table 6.2: Crystallographic Data and Refinement Statistics

	WT apo	**L116A/product**	**WT/product**
Data			
Source	5.0.2; ALS	Anode	5.0.2; ALS
Space group	$P4_12_12$	C2	$P3_121$
Resolution (Å)	2.0	3.1	1.8
Wavelength (Å)	0.98	1.54	0.98
Unit cell (Å)	a=b=52.8	a=182.0	a=b=114.0
	c=278.8	b=73.1	c=89.0
		c=92.7	
		β=95.4°	
Unique reflections	25509	19707	71892
R-merge (last shell) (%)	5.2 (20.4)	11.0 (23.8)	4.9 (26.7)
Completeness (last shell) (%)	90.8 (74.8)	88.5 (63.4)	99.8 (98.1)
Refinement			
R-factor (%)	19.4	27.3	22.6
R-free (6%) (%)	22.9	32.0	24.7
Number of atoms	2827	6612	3751
Number of waters	352	0	403
R.m.s.d. bond length (Å)	0.006	0.012	0.011
R.m.s.d. bond angles (deg.)	1.30	1.45	1.70
R.m.s.d. impropers (deg.)	0.67	0.86	1.04
R.m.s.d. dihedrals (deg.)	26.9	28.2	27.6
Mean B-value, overall (Å^2)	17.4	22.3	19.8
Mean B-value, protein (Å^2)	16.2	20.0	15.5
Mean B-value, DNA (Å^2)	NA	19.9	17.3
Mean B-value, solvent (Å^2)	25.0	NA	20.7

Chapter 6

The Structure of T4 Polynucleotide Kinase

Introduction

T4 polynucleotide kinase (PNK) was initially discovered in protein extracts of *Escherichia coli* bacteria infected with T-even phage (Richardson, 1965; Novogrodsky and Hurwitz, 1966; Novogrodsky et al., 1966). In the 37 years since, the enzyme has become a standard reagent for the manipulation and study of polynucleotides (Berkner and Folk, 1980; Chaconas and van de Sande, 1980; Sambrook et al., 1989). The product of the T4 early gene *pse*T (Sirotkin et al., 1978), PNK catalyses the transfer of the γ-phosphate group from adenosine triphosphate (ATP) or other nucleoside triphosphates (Novogrodsky et al., 1966; Richardson, 1981) to the 5'-hydroxyl of polynucleotides of many different lengths, sequences, and types (Richardson, 1965). Acceptable substrates include double and single stranded DNA, RNA and individual 3'-phosphate nucleotide bases. PNK also possesses a 3'-phosphatase activity that is independent of ATP (Cameron and Uhlenbeck, 1977; Sirotkin et al., 1978) and has been reported to hydrolyze 2',3'-cyclic phosphodiesters (Amitsur et al., 1987). Each catalytic activity is dependent on magnesium as a cofactor. The *pse*T gene was cloned and sequenced (Midgley and Murray, 1985), which facilitated both the use of recombinant PNK as a reagent and the biochemical study of PNK itself.

The only known biological function of T4 PNK is related to the ability of T4 phage to overcome a suicide defense mechanism used by some strains of bacteria (Sirotkin et al., 1978; Amitsur et al., 1987). In response to T4 infection, a latent tRNA anticodon nuclease encoded by *prr* strains of *E. coli* is activated by the Stp DNA restriction inhibitory peptide (Levitz et al., 1990; Tyndall et al., 1994; Penner et al., 1995) and introduces a break in the bacterial lysine tRNA. This lesion is defined by 2',3'-cyclic phosphate and 5'-hydroxyl ends and effectively prevents protein synthesis, thus inhibiting phage replication. T4 PNK phosphorylates the 5'-hydroxyl and reprocesses the 3' end by opening the 2',3'-cyclic phosphate and then removing the 3'-phosphate. The reprocessed ends are substrates for T4 RNA ligase. Thus, the phage is able to circumvent the tRNA lesion and continue to propagate through the bacterial population.

T4 PNK is the prototypical member of a broad family of 5'-kinase/3'-phosphatase enzymes that mend broken strands in nucleic acids in conjunction with the appropriate RNA or DNA ligase. Human members of this family have recently been implicated in the repair of DNA strand breaks caused by oxidative damage (Jilani et al., 1999; Karimi-Busheri et al., 1999; Whitehouse et al., 2001). In mammalian cells, the conversion of 5'-hydroxyl/3'-phosphate DNA ends to 5'-phosphate/3'-hydroxyl is a necessary step for the repair of DNA nicks and gaps that may be introduced by ionizing radiation (Henner et al., 1983), certain alkylating agents, DNaseII (Lown and McLaughlin, 1979; Torriglia et al., 1998), and the enzymatic removal of dead-end complexes of topoisomerase I generated by camptothecin inhibition (Yang et al., 1996). Mammalian PNKs display the bifunctionality of T4 PNK, but differ in specificity. While T4 PNK has a decreased activity against both recessed 5' ends and nicks in double stranded DNA (Lillehaug et al., 1976), mammalian PNKs show no preference for 5' overhanging termini and are active against gaps and nicks, which further implicates them in DNA repair pathways (Karimi-Busheri et al., 1998). Some mammalian PNKs are DNA specific while others have a preference for RNA (Jilani et al., 1999). Sequence alignments of *H. sapiens*, *C. elegans*, and *S. pombe* PNK enzymes reveal that T4 and mammalian enzymes share two active site motifs: the ATP-binding P-loop motif, which defines a kinase active site, and a short motif (DxDxT) associated with a family of phosphatase domains characterized by L-2-haloacid dehalogenase (HAD) (Karimi-Busheri et al., 1999). These conserved motifs suggest that the chemical mechanisms of T4 PNK and mammalian PNKs are similar.

Kinetic studies of T4 PNK have shown that the kinase activity proceeds according to an ordered sequential mechanism (Lillehaug and Kleppe, 1975) and that phosphoryl transfer results in an inversion of configuration at the phosphorus atom (Jarvest and Lowe, 1981). Furthermore, the kinase reaction is reversible. For example, the ATP analog β,γ-imidoadenylyl 5'-triphosphate acts as an inhibitor to the forward reaction, but is able to replace ATP in the reverse reaction to generate β,γ-imidoadenylyl 5'-

tetraphosphate (Lillehaug, 1978). While the minimal phosphorylation substrates are 3' nucleotide monophosphates, T4 PNK is able to phosphorylate a variety of oligonucleotides with a slight preference for a 5' terminal guanosine base (van Houten et al., 1998). Despite the tetrameric form of the enzyme, it does not display kinetic cooperativity (Lillehaug and Kleppe, 1975). It has also been noted that N-protected deoxyoligonucleotides do not interfere with kinase activity (van de Sande and Bilsker, 1973) and that PNK will phosphorylate several non-nucleosidic moieties when they are 5'-attached to oligonucleotides (Fontanel et al., 1994). The phosphatase activity is known to require magnesium or cobalt, has a pH optimum of 6.0, and is more active against DNA than RNA (Cameron and Uhlenbeck, 1977).

The kinase and phosphatase activities of T4 PNK are separable and are loosely localized to the N- and C-termini of the protein, respectively (Soltis and Uhlenbeck, 1982). Further proteolytic experiments coupled with mutational analyses have refined this separation (Wang and Shuman, 2001; Wang et al., 2002). A truncated N-terminal PNK consisting of residues 1 to 181 forms a monomer with no phosphatase and low kinase activity, while a construct missing residues 1 to 47 forms a dimer with residual phosphatase but no kinase activity. Mutations near the amino terminus (K15A, S16A, D35A, R38A, and R126A) ablate kinase activity without affecting phosphatase activity or the quarternary structure of the enzyme. This constellation of residues is also present in the active sites of adenylate kinases and similar enzymes. In contrast, mutations nearer to the enzyme C-terminus (D165, D167A, R176A, R213A, D254A, and D278A) inactivate the phosphatase activity without affecting the kinase activity or the quarternary structure of the enzyme. The motif $D_{165}xD_{167}xT_{169}$ found in T4 PNK is shared by a superfamily of phosphotransferases (Wang et al., 2002).

The first crystal structure of a bifunctional kinase/phosphatase enzyme was that of the dimeric 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK/FBP) (Hasemann et al., 1996). This enzyme plays an indirect role in the regulation of glucose metabolism. The N-terminal kinase domain of PFK/FBP also shares sequence (and structural) homology with adenylate kinases, whereas the C-terminal phosphatase domain belongs to the phosphoglycerate mutase and acid phosphatase enzyme families. Both of these families have similar topologies to the haloacid dehalogenase (HAD) family, but have a catalytic histidine instead of an aspartic acid (Hasemann et al., 1996). The presence of two catalytic functions in one polypeptide chain usually indicates a functional linkage. For T4 PNK and the multifunctional eukaryotic polynucleotide kinases, the covalent linkage of catalytic activities surely exists due to the requirement for coupled kinase and phosphatase activities in the repair of RNA and DNA lesions.

Results

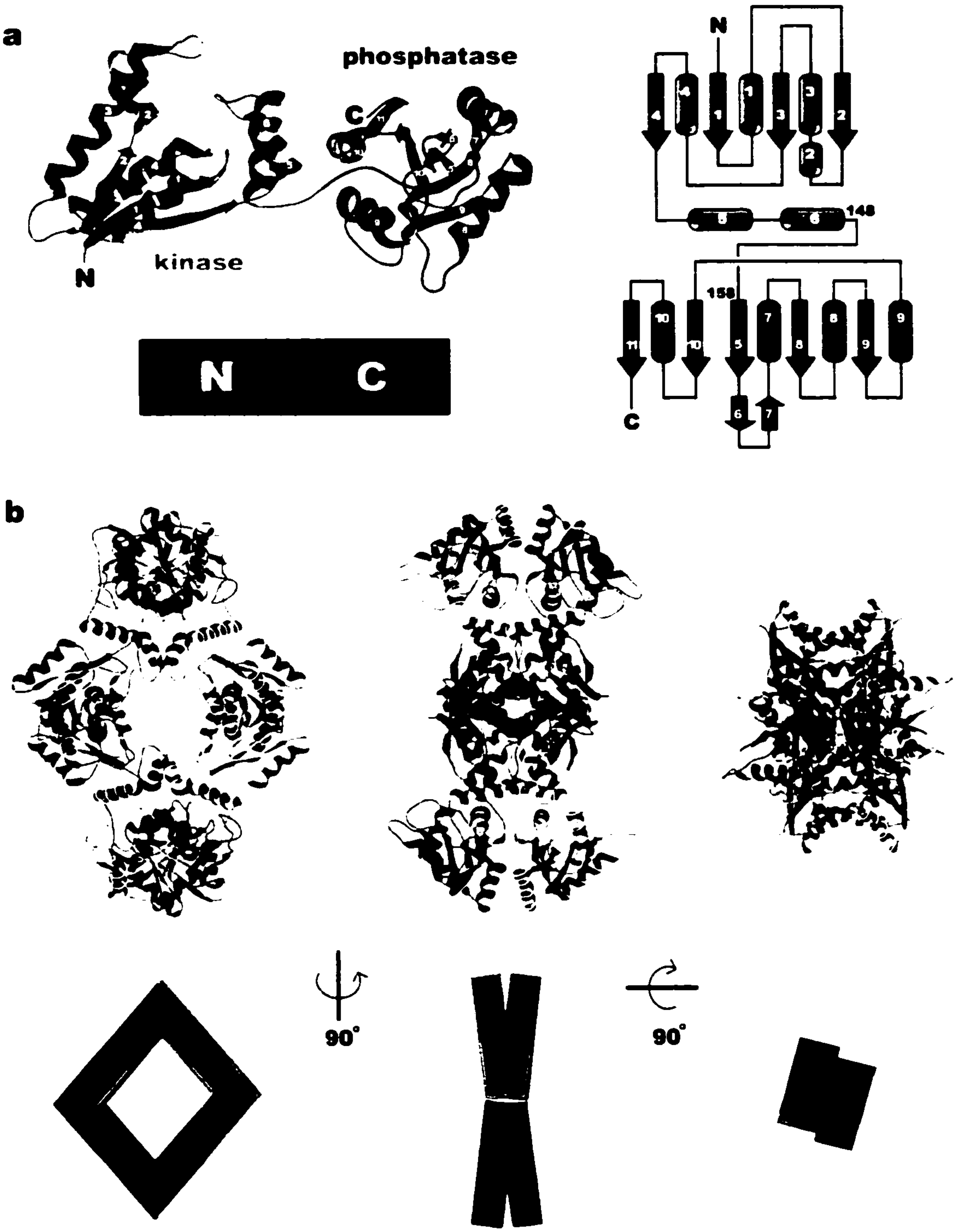
Enzyme fold and oligomerization

The structure of selenomethionyl substituted T4 PNK with bound nucleotide was determined by the single anomalous dispersion (SAD) method at 2.3 Å resolution. The structure was refined to an R_{work}/R_{free} of 0.236/0.261 with good geometry and an average protein B-factor of 36.5 $Å^2$ (Table 1 and side bar). The crystallographic asymmetric unit contains a single 34 kD subunit of T4 PNK and the I222 crystallographic symmetry operators generate two complete enzyme tetramers in the unit cell. Interpretable main chain electron density was observed for all but 3 residues (49-51) of the 301 amino acid residues in the enzyme monomer. Side chain density was observed for all but 9 lysines, 5 glutamic acids, 2 tyrosines, and 1 asparagine. These side chains are all solvent-exposed and have been truncated in the final model.

The PNK monomer is made up of two distinct N- and C-terminal α/β domains comprising residues 1 to 148 and 158 to 301 (Figure 6.1a). These domains correspond to the kinase and phosphatase activities, as described below. The tetramer is assembled through two separate interfaces (Figure 6.1b). One interface is formed exclusively with contacts between the N-terminal domains. This interface buries 964 $Å^2$ of accessible surface area and is formed mostly through the packing of helices 2 and 3 with their symmetry mates. The second interface is primarily formed by contacts between C-terminal domains and buries 1328 $Å^2$ of exposed surface area. This interface is anchored by strand 11, which interacts with its symmetry mate in an anti-parallel fashion, creating a pseudo-continuous 10-stranded β-sheet between two C-terminal domains. Helix 7 packs against its symmetry mate to complete the interface. This interface is further augmented by additional interactions between helix 5 and its symmetry mate from the N-terminal domain. The overall dimensions of the enzyme tetramer are approximately 120 × 80 × 50 Å and there is a solvent-exposed channel through the middle of the tetramer 15 to 20 Å in diameter (Figure 6.1b).

The distance between kinase and phosphatase active sites within an individual enzyme subunit is approximately 35 Å and is similar to the closest distance between active sites from separate subunits. The sites within an individual subunit point in opposite directions and are physically separated by helices 5, 6, 10, and strand 11 whereas the closest kinase and phosphatase active sites from separate subunits face each other across a surface valley.

Figure 6.1: The structure of T4 polynucleotide kinase. (a) A monomer of T4 polynucleotide kinase is shown as a ribbon diagram with the N-terminal kinase domain in red and the C-terminal phosphatase domain in blue. The secondary structural elements are numbered according to their occurrence in the chain. A topology diagram shows b-strands as arrows and a-helices as cylinders. (b) The tetrameric form of the enzyme is shown color-coded by monomer. Subsequent images are rotated by 90° as indicated and reveal the 222 symmetry of the tetramer. The block diagrams under each view of the tetramer show the positions of the N-terminal (red) and C-terminal (blue) domains.


a
phosphatase
kinase
N
C
N
C
N
C
148
158
1
2
3
4
5
6
7
8
9
10
11
b
90°
90°

The kinase domain

The N-terminal domain is composed of the first 148 residues and has a core consisting of a 4-stranded, parallel β-sheet with 4-1-3-2 topology (Figure 6.1a). The fold is completed by four α-helices that are found on both sides of the central sheet and 2 additional helices following the last strand. A structural similarity search using the DALI server (Holm and Sander, 1993) with the N-terminal domain of PNK identifies a large structural kinase family that displays homology to PNK. There are 272 structures with a Z-score ≥ 2.0, 25 structures with Z ≥ 5.0 and three structures with Z > 10.0. These three structures, in order of improving Z-score, are adenylate kinase (Z = 12.0, r.m.s.d. = 2.5 Å), chloramphenicol phosphotransferase (Z = 12.4, r.m.s.d. = 2.6 Å) and the N-terminal domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (Z = 14.2, r.m.s.d. = 2.5 Å). A superposition of PNK, of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK/FBP), and adenylate kinase (ADK) reveals interesting similarities and differences between these related structures (Figure 6.2a). The N-terminal domain of PNK is the smallest of the three domains; PFK/FBP and ADK contain 2 and 1 extra strand(s) in their β-sheets. ADK contains a functionally important zinc knuckle sub-domain inserted between helix 5 and 6 of PNK. In place of this sub-domain, PFK/FBP and PNK have short 12 and 7 amino acid loops respectively.

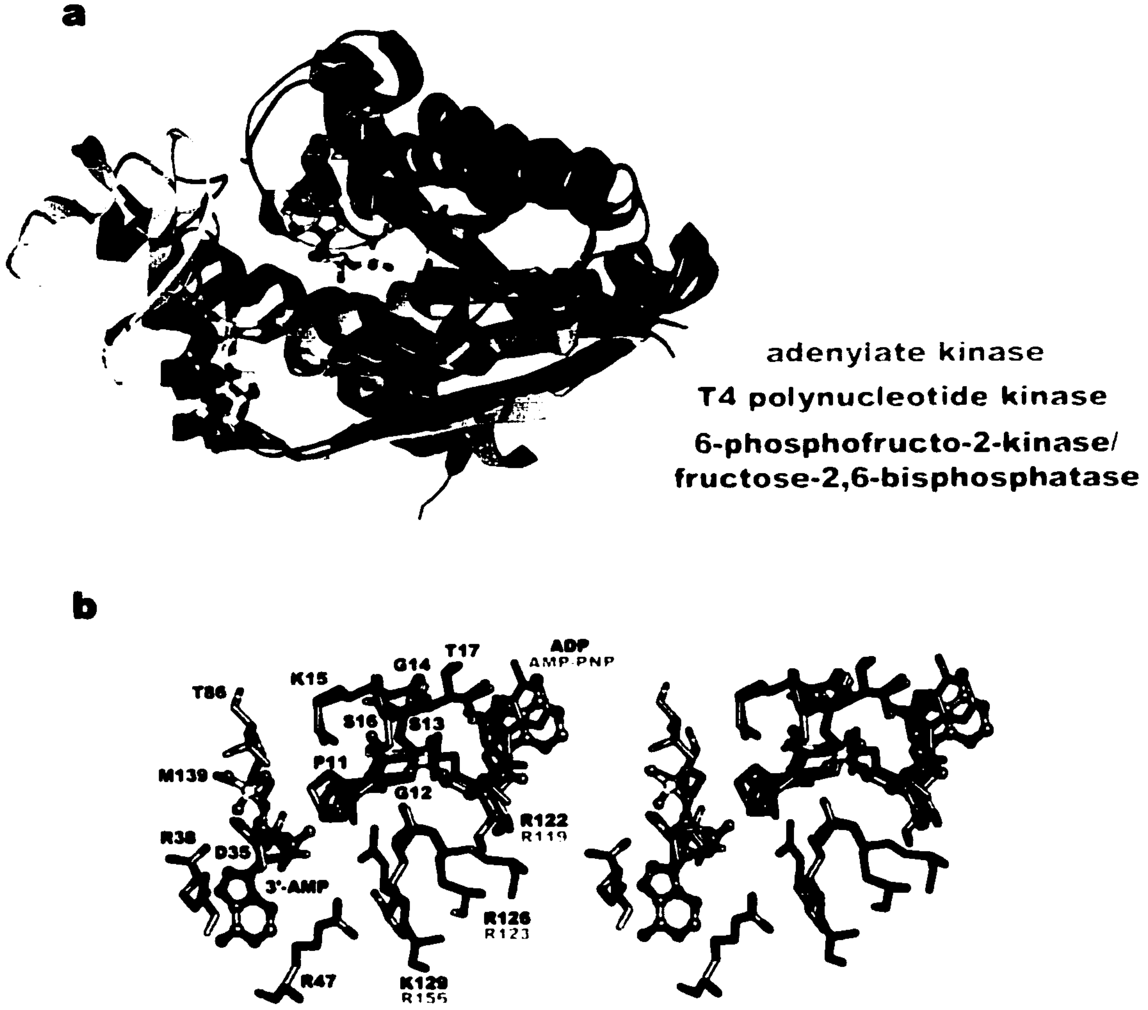


Figure 6.2: Superposition of the kinase domain and the kinase active site. (a) A superposition (DALI (Holm and Sander, 1993)) of T4 polynucleotide (red), adneylate kinase (1adk; gray), and the kinase domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (1bif; green) with both the observed ADP ligand and modeled 3'-AMP ligand of PNK. (b) The kinase active site of PNK is shown in a stick representation colored by atom type with residue labels in black. Both the observed ADP ligand and the modeled 3'-AMP ligand are displayed in ball and stick representations. On the ATP binding side, the corresponding residues from adenylate kinase along with its bound AMPPNP ligand (1ank) are shown colored in slate.

The kinase domain contains a bound ADP molecule in its active site, which represents the hydrolyzed product of ATP present in the crystallization buffer (Figure 6.2b). The kinase active site contains a shallow tunnel, formed as a result of contacts between two enzyme surface loops that create a lid over the phosphate tail of the bound ADP. The contacts between these loops include residues E46 and R47 in the helix 2-3 loop and residues T128 and K129 in the helix 5-6 loop. The surface of the active site tunnel consists almost entirely of charged or polar residues, including R126, K129, K15, S16, R138, D35, T86 and N33 (Figure 6.3, middle). Several hydrophobic residues (P11, V131, V135 and M139) also help to form the active site walls. The bound ADP molecule is located on the side of the tunnel that opens into the channel in the middle of the enzyme tetramer (Figure 6.3a). Its binding pocket is formed mostly through electrostatic interactions between its diphosphate tail and a series of basic residues (K15, S16 and R126) and backbone amide groups from residues 12 to 17 (Figure 6.2b). The adenine ring stacks against the R122 side chain. The phosphate tail of the ADP is more deeply buried in the active site and more ordered than its associated ribose and adenosyl groups. The nucleotide-binding pocket is well conserved in members of the ADK family and is formed by residues that are a part of the P-loop motif. The r.m.s.d. for backbone atoms in an alignment of the ADK (residues 10-15) and PNK (residues 12-17) ATP binding sites is 0.2 Å and the ADP molecule bound in the active site of PNK is in the same conformation as AMPPNP bound to ADK (Berry et al., 1994) (Figure 6.2b). This allowed us to easily model the position of the γ-phosphate into the PNK active site.

Based on the strong active-site homology between the kinase domain of PNK and ADK and the structure of ADK with bound adenylate and AMPPNP (Berry et al., 1994), an adenosine 3'-monophosphate (3'-AMP) substrate molecule can be modeled into the active site on the other side of the tunnel from the bound ADP. This binding site faces the exterior of the enzyme tetramer (Figure 6.3b) and opens into a broad surface valley capable of accommodating larger polynucleotide substrates such as tRNA.

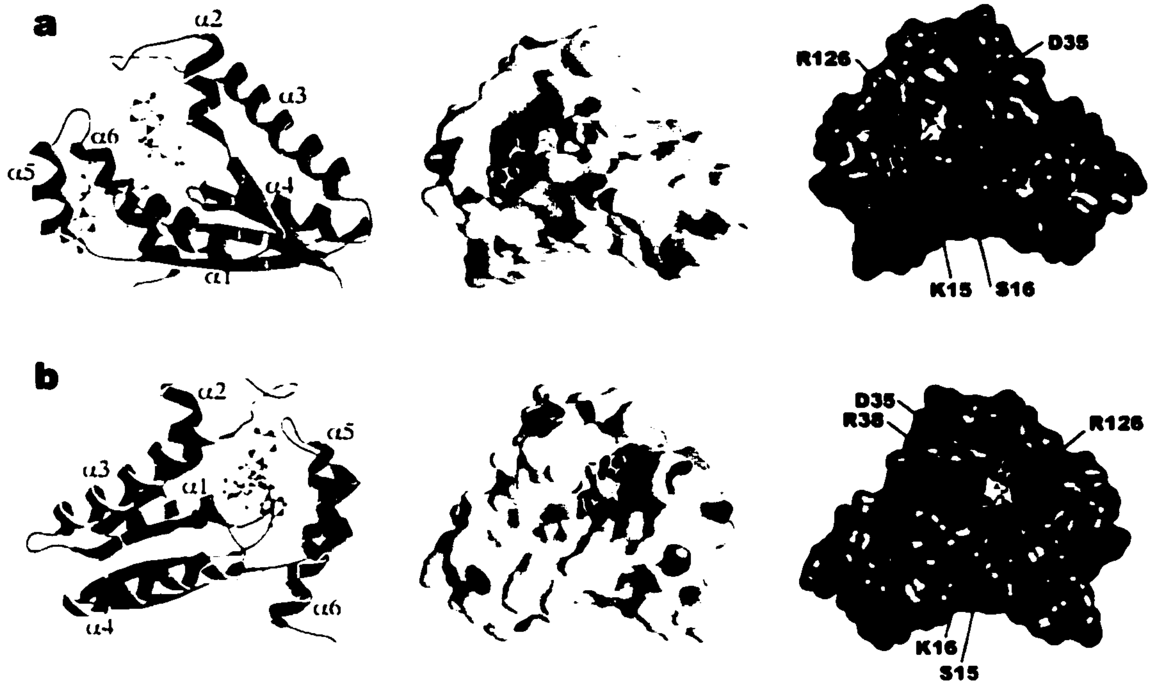


Figure 6.3: The kinase domain of PNK is shown in two views of three representations each. Panels (a) and (b) show two views that are related by a 180° rotation. The three representations are a ribbon diagram, a molecular surface color-coded by electrostatic potential (GRASP), and a molecular surface color-coded by mutational results. The ribbon diagram is labeled according to Figure 6.1 for reference. The GRASP surface shows a gradient of electrostatic potential ranging from –10 kT/e (red) to 15 kT/e (blue). In the third representation, the molecular surface is colored green and labeled where alanine mutations ablate kinase activity. All three illustrations contain the observed ADP ligand and the modeled 3'-AMP ligand in a ball and stick representation.

A 3'-AMP model was placed into the PNK kinase active site based on the position of bound adenylate in ADK. The conformation of the bound 3'-AMP model was energy minimized to alleviate any steric clash with PNK while holding the 5'-hydroxyl position fixed. In the resulting model of the ternary complex, the 3'-AMP phosphate makes contacts to the backbone amide of T86 (2.8 Å) and the side chains of R38 (3.0 Å) and K15 (4.4 Å). R34 is 5.2 Å away from the 3'-phosphate oxygen, but has room to rotate towards the putative phosphate binding position. In the proposed binding mode, the adenosine ring fits nicely in the binding pocket but there are no base-specific contacts.

The 3'-phosphate points out of the tunnel, and the 5'-hydroxyl points into the tunnel and is positioned correctly for an in-line phosphoryl transfer (3.3 Å from the modeled γ-phosphorous atom; Figure 6.2b). D35 is found on the 3'-AMP side of the tunnel and is observed in two conformations. In one conformation, it is engaged in a bidentate contact with R47 (3.4 and 4.0 Å) and in the other conformation it forms a salt bridge with R38 (3.6 Å). In either conformation, the D35 side chain is positioned close to the modeled position of the 3'-AMP 5'-hydroxyl (2.9 - 4.0 Å) and is a likely candidate for a magnesium-binding side chain.

The phosphatase domain

The C-terminal domain is composed of the last 143 residues and has a core consisting of a 5-stranded, parallel β-sheet with 11-10-5-8-9 topology (Figure 6.1a). The α/β fold is completed by four α-helixes inserted between consecutive β-strands. Two additional β-strands form a small 2-stranded anti-parallel sheet between strand 5 and helix 7. A structural similarity search using the DALI server indicates that the C-terminal domain of PNK belongs to the sizeable haloacid-dehalogenase (HAD) structural family with 457 known structures with a Z-score ≥ 2.0, 53 structures with Z ≥ 5.0 and 5 structures with Z > 8.0. These five domains are phosphonoacetaldehyde hydrolase (Z = 9.8, r.m.s.d. = 2.6 Å), ʟ-2-haloacid-dehalogenase (Z = 9.5, r.m.s.d. = 3.2 Å), epoxide hydrolase (Z = 8.9, r.m.s.d. = 3.3), phosphoserine phosphatase (Z = 8.8, r.m.s.d. = 3.1 Å), and hypothetical phosphatase protein hi1679 (Z = 8.5, r.m.s.d. = 2.7 Å). As with the N-terminal domain, the C-terminal domain of PNK is smaller than its closest structural neighbors. Both ʟ-2-haloacid-dehalogenase (HAD) and phosphoserine phosphatase (PSP) have additional helical sub-domains that reduce the solvent accessibility of the active site cleft (Figure 6.4a). The absence of such a sub-domain in PNK is likely responsible for the broader specificity of PNK compared to other members of this family.

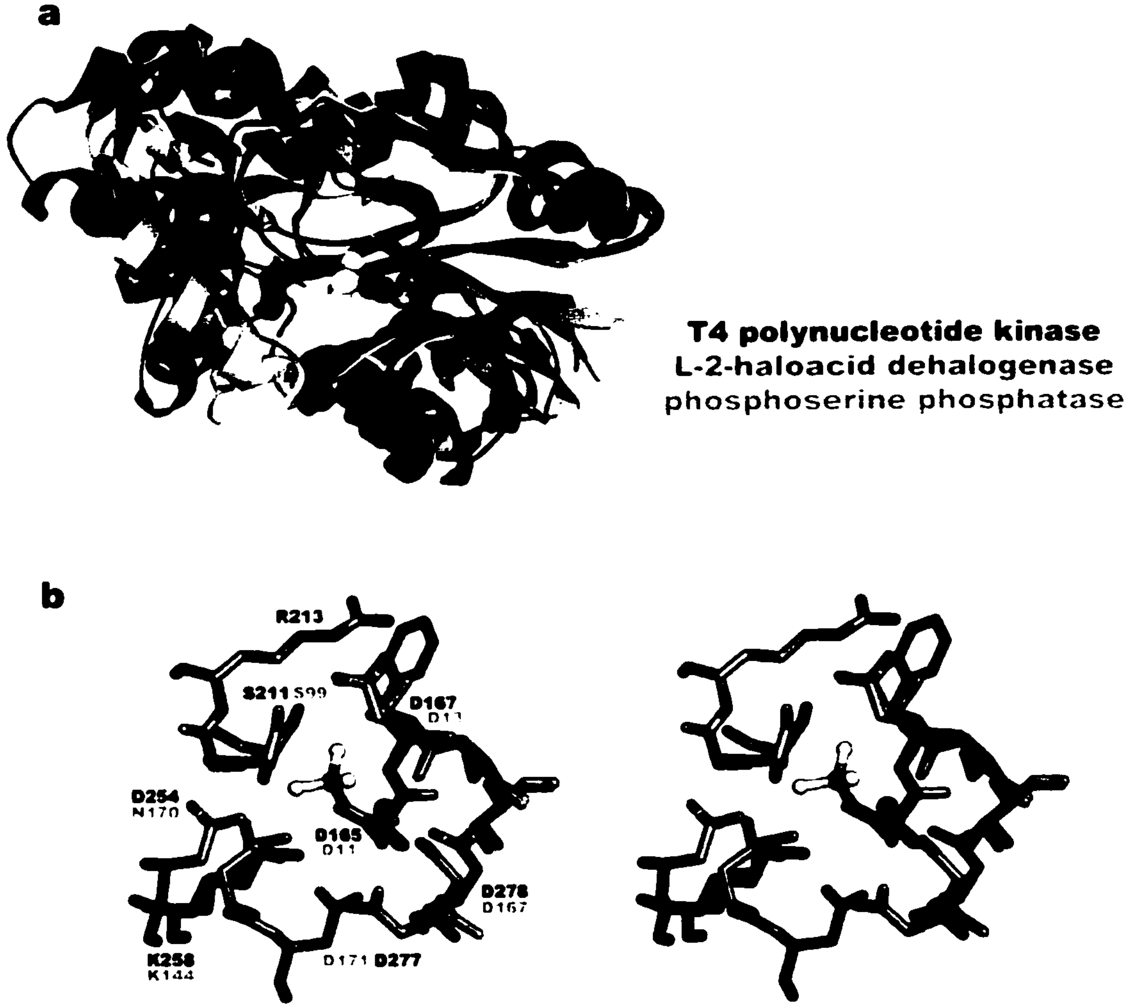


Figure 6.4: Superposition of the phosphatase doimain and the phosphatase active site. (a) A superposition of the phosphatase domain of T4 polynucleotide kinase (blue), phosphoserine phosphatase (1j97; gray), and L-2-haloacid dehalogenase (1qq5; green) with the beryllium fluoride aspartic acid ester from the phosphoserine phosphatase structure. (b) The phosphatase active site of PNK is shown in a stick representation colored by atom type with residue labels in black. The beryllium fluoride moiety is displayed in a ball and stick representations (orange = beryllium; yellow = fluoride; green = magnesium). Corresponding residues from phosphoserine phosphatase are displayed in slate.

The smaller size of both PNK domains has probably evolved both to broaden the substrate specificity of the enzyme and maximize the information content of the phage genome for effective propagation. The HAD family of enzymes uses a common active-site architecture defined by the D-X-D-X-T motif ($D_{165}VDGT_{169}$ in PNK). In all cases to date, these enzymes have been shown to proceed via a covalent phospho-aspartic acid intermediate and, with the exception of HAD, are dependent on magnesium for activity (Collet et al., 1999; Ridder and Dijkstra, 1999; Ridder et al., 1999). The phosphatase active site of PNK is centered on the primary aspartic acid in the conserved motif, D165, which is located on the C-terminal end of the central β-strand (strand 5). A representation of the electrostatic potential of the molecular surface reveals that the active site cleft is acidic. Furthermore, residues that have been mapped by mutagenesis to be required for phosphatase activity cluster in the active site of the C-terminal domain (Figure 6.5, right).

Figure 6.5: The phosphatase domain of PNK is shown in three representations: a ribbon diagram; a molecular surface color-coded by electrostatic potential (GRASP); and a molecular surface color-coded by mutational results. The ribbon diagram is labeled according to Figure 6.1 for reference. The GRASP surface shows a gradient of electrostatic potential ranging from –15 kT/e (red) to 15 kT/e (blue). The molecular surface is colored green and labeled where alanine mutations ablate kinase activity. The ribbon diagram contains the modeled phospho-enzyme intermediate in a spheres representation to define the center of the phosphatase active site.

The superposition of the active site of PNK with phosphoserine phosphatase inhibited by beryllium fluoride (Cho et al., 2001) (a mimic of a phospho-aspartate intermediate) (Figure 6.4b) reveals key functional residues and their likely roles in catalysis. The δO1 oxygen of D165 is the nucleophilic atom responsible for attacking the 3'-phosphorous in the S_N2 phosphoryl transfer reaction. ζN from K258 is 2.7 Å from D165 δO1. D165 δO2 appears to be a metal ligand (2.1 Å) along with D278 (2.7 Å), the backbone carbonyl from D167 (2.1 Å), and a substrate phosphate oxygen (2.0 Å). Two water molecules would likely complete the octahedral coordination of the magnesium ion. The modeled, non-bridging phosphate oxygens are contacted by the backbone amides of G212 (2.7 Å), V166 (3.0 Å), and D167 (2.8 Å), γOH of S211 (3.2 Å), and ζN of K258 (2.9 Å). The second aspartic acid in the conserved active site motif (D167) forms a salt bridge with R213 (2.4 and 2.6 Å). D277 is 3.6 Å from the modeled magnesium position and D254 is 4.5 Å from a modeled phosphate oxygen atom. These residues correspond spatially to D171 and N170, respectively, in phosphoserine phosphatase although they come from different positions within the sequence and overall folds.

Discussion

T4 PNK as a model for eukaryotic polynucleotide kinases

Recently, a human polynucleotide kinase has been described (Jilani et al., 1999; Karimi-Busheri et al., 1999; Mani et al., 2001; Whitehouse et al., 2001). The structure of T4 PNK provides a model for the structure function relationships of hPNK and of other eukaryotic polynucleotide kinases in this enzyme family. Sequence alignments reveal that the eukaryotic enzymes contain both the kinase and phosphatase active-site motifs found in T4 PNK (Karimi-Busheri et al., 1999). Unlike T4 PNK, these enzymes are active against either DNA or RNA (but not both) and only the DNA kinases posses the additional 3'-phosphatase activity. Mammalian PNKs are believed to play a role in the repair of DNA lesions with 3'-phosphate and 5'-hydroxyl termini. Human polynucleotide kinase (Jilani et al., 1999; Whitehouse et al., 2001) and a homolog from yeast (Meijer et al., 2002) have been implicated in the response to DNA damage by oxidation, γradiation, and camptothecin. Other than the active site regions, the enzymes lack significant sequence homology to T4 PNK. The kinase and phosphatase domains are reversed compared to T4 PNK with the P-loop motif found in the C-terminal domain and the phosphatase motif found in the N-terminal domain. Furthermore, the eukaryotic enzymes are nearly twice as large (60 kD) as T4 PNK and exist as monomers in solution (Mani et al., 2001). However, it is likely that these enzymes will have active-site structures similar to those in T4 PNK and it will be interesting to compare the structures of the eukaryotic DNA PNKs to that of T4 PNK in terms of substrate binding modes, domain organization, and overall structure.

Nicked tRNA ends can access kinase and phosphatase active sites simultaneously

The structure of PNK indicates that it is possible for nicked tRNA to bind in both the kinase and phosphatase active sites simultaneously (Figure 6.6). The tetrameric structure of PNK is crucial for this binding mode as the polynucleotide binding side of the kinase active site from one monomer faces the phosphatase active site of a neighboring monomer. The distance (~35 Å) between these sites appears appropriate for binding the nicked ends of a tRNA anticodon loop. It is formally possible that one PNK tetramer could bind up to four tRNAs simultaneously as symmetry related tRNA binding modes do not clash with each other. The anticodon stem-loop consists of 5 ribonucleotides that do not base pair with each other or make other non-local interactions. The only known *in vivo* substrate for PNK is lysine tRNA that has been nicked directly 5' of the anticodon nucleotides. Nicked tRNAs of this type should possess two flexible overhangs extending from the stable, base-paired stem structure. One overhang is 2 nucleotides long and has a 2',3'-cyclic phosphate terminus while the other is 3 nucleotides long, contains the anticodon sequence, and has a 5'-hydroxyl terminus. The ~35 Å distance between active sites can be spanned by these flexible overhangs suggesting that PNK performs both 3'- and 5'-terminal modifications within one tRNA binding event.

Figure 6.6: A hypothetical RNA/protein complex in which both PNK active sites are occupied simultaneously is shown with tRNA (gray) and two monomers from the tetramer of PNK (blue and red). The 5'-hydroxyl end of the nicked tRNA (yellow) is bound in the kinase active site of the red monomer while the 2',3'-cyclic phosphate end (orange) is bound in the phosphatase active site of the blue monomer.

The oligomeric state and catalytic activity of truncated constructs

Much of the published biochemical data on the domain organization of T4 PNK can be interpreted based on the enzyme structure. The oligomerization state and bifunctionality of PNK has been probed previously by proteolysis and truncation experiments. PNK comprised only of residues 1 to 181 is a monomer with no phosphatase activity and low kinase activity; PNK consisting only of residues 42 to 301 is a dimer with no kinase activity and residual phosphatase activity (Wang and Shuman, 2001). These results indicate that although no allostery has been observed and the active sites are well separated, the tetrameric form of the enzyme is critical for the full activity of both domains. The N-terminal/N-terminal domain interface includes helices 2 and 3, which also form one side of the kinase active-site tunnel. More specifically, a2 is the source of several residues that make contacts to the modeled 3'-AMP and have been implicated by mutagenesis experiments to be critical for kinase activity. The a2/a2 interface consists of two main contacts. Tyrosine 37 stacks against its symmetry mate and D36 makes contacts to both H44 (3.2 Å) and S40 (2.7 Å) of a symmetry-related monomer.

The removal of residues 1-41 [PNK (42-301)] removes a large amount of the N-terminal/N-terminal interface and prevents the tetramerization of the enzyme, leading to C-terminal linked dimers in solution. However, the phosphatase activity is also significantly reduced despite its C-terminal location ~30 Å away from the kinase active site. One likely possibility is that the remaining N-terminal domain sterically interferes with the phosphatase active site since it is connected to the C-terminal domain via what appears to be a flexible loop. This explanation is strengthened by the basic nature of the N-terminal domain and the acidic nature of the phosphatase active site. The removal of residues 182-301 [PNK (1-181)] eliminates nearly the entire C-terminal domain and thus we would predict an N-terminal linked dimer in solution. The observation of a monomer leads to the conclusion that the N-terminal domain interface is not as robust as the C-terminal interface. This hierarchy in the strengths of the domain interfaces correlates well with surface area burial, as the C-terminal interface buries 37% more surface area than the N-terminal interface. As mentioned above, the kinase active site is closely associated with the N-terminal interface via helix 2. The lack of such an interface with the monomeric PNK (1-181) might explain the decrease in kinase activity despite the presence of all the critical active site residues.

Substrate binding modes and substrate specificities of ADK family members

The substrate profiles of PNK and ADK, as well as additional kinase family members, can be explained by their structures. The bound ADP molecule observed in the active site of the kinase domain is bound in the same conformation as other P-loop bound nucleotides. The P-loop is a phosphate-binding motif with well conserved sequence and structure (Saraste et al., 1990). It has the consensus sequence (A/G)-X_4-G-K-(G/S/T) and is known to occur in adenylate kinases, ras proteins, elongation factors, ATP synthase β-subunits, myosin heavy chains, thymidine kinases, and phosphoglycerate kinases. The observed binding mode of ADP confirms that, as in other P-loop structures, ADP is recognized and bound mostly through interactions of the protein with the phosphate oxygens. This mode fits well with the observation that PNK is able to use any nucleotide triphosphate as the phosphoryl donor in the kinase reaction (Novogrodsky et al., 1966; Richardson, 1981). Similarly, the modeled binding mode of 3'-AMP shows a lack of base-specific contacts, which is reasonable since PNK can phosphorylate a polynucleotide with any 5'-terminal nucleotide (Richardson, 1981). Furthermore, the 3'-phosphate position allows the modeled mononucleotide to be easily extended to a variety of polynucleotides.

The ATP and polynucleotide binding sites are on opposite sides of the active site tunnel (Figure 6.3). In the context of the tetramer, the tunnel leads from the central channel to the exterior. The ATP binding sites open into the channel while the polynucleotide-binding site is on the exterior side of the tetramer. This allows large polynucleotide substrates to bind on the outside of the tetramer without encountering steric clashes from other PNK subunits or from other polynucleotides that might already be bound in the other active sites. Interestingly, the molecular surfaces of adenylate kinase do not exhibit a tunnel. In its open form, ADK contains a binding cleft for both ATP and adenylate. However, as the enzyme binds substrate, the latch domain closes down on the substrate, completely cutting it off from solvent (Berry et al., 1994). For ADK, the closing of the latch creates a specific binding pocket within the enzyme so that only the desired substrate molecule will bind productively. In contrast, the broad tunnel we observe for PNK allows the enzyme to have a very broad substrate specificity. The tunnel opens up to solvent allowing, polynucleotides of all types and sizes to bind and become phosphorylated. With other members of the family, such as PFK/FBP, remnants of a tunnel can be observed, but the tunnel openings are considerably narrower and more restrictive than in PNK. The tunnel/clamp structure presumably has diverged within the PNK/ADK structural family along with the polynucleotide-binding site. The selection of a particular tunnel or clamp backbone structure along with the specific residues that line the phosphoryl acceptor binding pocket appears to be an efficient way to generate a large variety of substrate specificities within the context of a single fold.

Kinase active site structure correlates with mutational analyses

Alanine scanning mutagenesis experiments on PNK have revealed a subset of residues that appear critical to the kinase activity (K15, S16, D35, R38, R126) (Wang et al., 2002). These five mutants all map to the tunnel region (Figure 6.3, right) and their specific roles in the catalytic mechanism of phosphoryl transfer can be described. K15 is one of the defining residues in the P-loop motif and is thought to be important for the structure of this motif because of its strict conservation (Saraste et al., 1990). It interacts with the β-phosphate of the bound ADP and with the modeled position of both the ATP γ-phosphate and the 3'-AMP 5'-hydroxyl. S16 is also a part of the P-loop structure as both the backbone amide of S16 and the side chain hydroxyl contact the β-phosphate of the bound ADP. Interestingly, this position in the motif is not as conserved as the preceding lysine. ADK has a glycine in this position and other P-loop proteins have a threonine (myosin heavy chains and ATP synthase β-subunits). Fitting with the apparent role of S16, kinase activity is maintained by PNK with the S16T mutation (Wang et al., 2002). R126 also contacts the diphosphate group of ADP, is conserved in ADK (R123), and is a lysine (K172) in PFK/FBP. Therefore, K15A, S16A, and R126A mutations are deficient in kinase activity because they disrupt the ATP binding site of PNK. D35 is observed in two conformations as described above and is located on the AMP side of the tunnel. In one conformation it forms a salt bridge with R38, which contacts the 3'-phosphate of 3'-AMP in our model, so one possible role for D35 is simply to help position R38 for phosphate binding. However, R38 is also positioned by a salt bridge with E47 and so a D35A mutation might not be expected to ablate kinase activity if this were its main function. D35 is unique in being the one acidic residue found near the active site of the kinase domain. Since the kinase activity is dependent on magnesium, a more likely function of D35 is metal binding. In its alternate conformation, it is well positioned for this role. Its delta oxygens here are 3.7 Å and 4.0 Å from the modeled position of the 5'-hydroxyl of 3'-AMP and pointing in the right direction to bind a metal ion that would bridge the gap between the 5'-hydroxyl and the γ-phosphate of ATP.

Phosphatase active site structure correlates with mutational analyses

Alanine scanning mutagenesis experiments on PNK have also been performed on the C-terminal phosphatase domain and have yielded a set of six residues that are critical for activity (D165, D167, R176, R213, D254, and D278) and a set of two residues that partially reduce activity (R246 and R279) (Wang et al., 2002). D165 and D167 make up the core of the conserved active site motif D-X-D-X-T. Based on the structural alignment of PNK with the beryllium fluoride-bound phosphoserine phosphatase structure, D165 functions both as the nucleophile responsible for attacking the phosphorous atom in a S_N2 type reaction and as a magnesium ligand. Main-chain atoms of D167 are involved in both phosphate (amide) and metal (carbonyl) binding, but the D167A mutant should still be able to function in that capacity. The side chain of D167 forms a salt bridge with R213 and the two residues are widely separated in sequence as D167 is at the C-terminal end of strand 5 and R213 is at the C-terminal end of strand 10. Therefore, the D167-R213 salt bridge appears to be important for stabilizing the structure of the active site. Strand 10 is a critical constituent of the active-site structure, since both S211 and G212 interact with the bound phosphate. Furthermore, since both D167A and R213A mutants are defective in phosphatase activity, the salt bridge between them evidently plays a critical role in the formation of the active site. This also explains why the D167N and D167E enzymes are not active (Wang et al., 2002). Asparagine is unable to form a strong electrostatic interaction with arginine and while a glutamate residue can form a salt bridge with arginine, its extra length would perturb the positions of S211 and G212 relative to the rest of the active site. Similarly, the modified salt bridge that would exist between lysine, in an R213K mutant, and aspartic acid must not position the rest of strand 10 appropriately for phosphate binding because this mutant is also inactive. D254 is positioned away from the phosphate-binding site (4.5 Å) and forms a salt bridge with R279 (3.0 Å). Both of these residues may be important for polynucleotide binding as they are positioned in the entrance to the active site cleft and replacement of R279 with alanine partially reduces activity. The partially reduced activity of R246A involves another arginine residue that contacts D254 on the opposite side from R279 (3.0 and 3.2 Å) and is again likely to be important for substrate binding. D278 is located in close proximity to the modeled magnesium ion and would therefore be important for metal binding. The fact that D278E is partially active is consistent with this role. Out of the six identified residues, R176 is the furthest from the phosphate-binding center (11.0 Å) and does not contribute in any clear way to the active site.

The 2',3'-cyclic phosphodiesterase activity must also reside in the phosphatase domain active site, but there is a dearth of data on this activity. Future experiments involving co-crystal structures with the various substrates of both domains along with bound magnesium ions should clarify the relationship of this activity to the 3'-phosphatase activity.

Summary

The structure of T4 PNK is the first crystal structure of a bifunctional RNA repair enzyme and serves as a model for similar eukaryotic DNA repair enzymes like hPNK. The structure verifies that the kinase and phosphatase activities reside in structurally distinct domains that are members of the adenylate kinase family and the L-2-haloacid dehalogenase family, respectively. Both active sites have been located in the context of the enzyme tetramer and described via a combination of the observation of bound substrates and homology modeling. Truncation and mutational results have been mapped onto the structure and explained. Lastly, a tRNA binding mode has been proposed in which both 3' and 5' modifications occur in the context of a single binding event.

Methods

Subcloning, expression, and purification

Subcloning was required to move the T4 PNK gene into an inducible vector for expression of selenomethionine-derivitized protein. The gene for T4 PNK was amplified from p.PNK plasmid (New England Biolabs (NEB)) using the polymerase chain reaction (PCR) with a forward primer containing an *Nde*I cleavage site at the start codon (5'-GGTAGGTCATATGAAAAAGATTATTTTGACTATTG-3') and a reverse primer containing a *Bam*HI site just 3' of the stop codon (5'-CGAAGCCCTCTAAAAATTCCTAGGAT-3'). pAII17 (a pET-based vector , W. Jack, NEB) and the PCR generated fragments were digested with *Bam*HI and *Nde*I and then ligated to generate the pET-PNK expression plasmid.

Selenomethionine containing T4 PNK was expressed in minimal media from the *E. coli* strain BL21(DE3) adapted for growth with methionine pathway inhibition (Doublie, 1997). Cells were grown in minimal media at 37 C to an OD_{600} of 0.8 and the following amino acids were added to inhibit the methionine biosynthetic pathway: 100 mg/L lysine, threonine, phenylalanine; 75 mg/L selenomethionine; 50mg/L leucine, isoleucine, valine. Following a 15-minute incubation at 37 C, 0.5 mM isopropyl-thio-b-D-galactosidase (IPTG) was added to induce expression and the cultures were left at 37 C for 10-14 hours.

The cell pellets (25 g) were resuspended in 100 ml of Phenyl column buffer A (0.2 M NaCl, 20 mM Tris-HCl, 0.1 mM EDTA, 10 mM beta-mercaptoethanol, pH 7.5 at room temperature). The resuspended cells were lysed by sonication with a Sonicator Cell Disruptor (Heat Systems-Ultrasonics, Inc., Plainview, NY) with 2-minute intervals at 70 % power until the soluble protein reached a plateau. The resulting lysate was centrifuged for 30 min. at 12,000 RPM in a Beckman JA17 rotor. The crude cell supernatant was applied to an 80 ml Phenyl Sepharose column, which was washed with 3 column volumes (CV) of Phenyl column buffer A. A gradient was then applied from 0-50% ethylene glycol (Fisher) in the same buffer and the eluent was collected in 18 ml fractions. The peak T4 PNK-containing fractions were determined by running samples on a 10-20% SDS-PAGE gel stained with Coomassie Blue and were pooled (175 ml final volume). The Phenyl Sepharose pool (175 ml) was diluted to a final volume of 1.0 L with Heparin Column Buffer A (0.1M NaCl, 20 mM Tris-HCl, 0.1 mM EDTA, 10 mM beta-mercaptoethanol, 10 % glycerol) and was applied to a 7.8 ml AF-Heparin 5PW column. The column was washed with 3 CV of the same buffer and an NaCl gradient from 0.1 M NaCl to 1.0 M NaCl in 10 CV of the same buffer was applied to elute the enzyme. The peak T4 PNK-containing fractions were determined by running samples on a 10-20% SDS-PAGE gel stained with Coomassie Blue and were pooled (12 ml). The resulting pool was concentrated by dialysis into 0.1 M NaCl, 10 mM Tris-HCl, 0.1 mM EDTA, 2 mM DTT, pH 8.0 at 4 C.) The concentrated AF-Heparin 5PW pool (approximately 4 ml) was applied to a 318 ml HiPrep Sephacryl S-200 size exclusion column and was eluted with 0.3M NaCl, 20 mM Tris-HCl, 10 mM beta-mercaptoethanol, 0.1 mM EDTA, 10% glycerol, pH 8.0 at 4 C, collected in 4 ml fractions. The peak T4 PNK-containing fractions were determined by running samples on a 10-20% SDS-PAGE gel stained with Coomassie Blue and were pooled (28 ml). The HiPrep Sephacryl S-200 pool was dialyzed for 16 hours against 2.0 L of Source Q15 buffer A (0.1 M NaCl, 20 mM Tris-HCl, 10% glycerol, 0.1 mM EDTA, 10 mM beta-Mercaptoethanol, pH 8.5 at 4 C) and was applied to a 7.8 ml Source Q15 column. The column was washed with 3 CV of Source Q15 buffer A and was eluted with an NaCl gradient from 0.1 to 1.0 M in the same buffer. The peak T4 PNK-containing fractions were determined by running samples on a 10-20% SDS-PAGE gel stained with Coomassie Blue and were pooled (34 ml). The final pool was dialyzed against storage buffer (50 mM NaCl, 10 mM Tris-HCl, 0.1 mM EDTA, 1 mM DTT, 50 % glycerol, 0.1 mM ATP, pH 7.4 at room temperature).

Crystallography

Selenomethionine T4 PNK was dialyzed from storage conditions into 10 mM Tris pH 7.4, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, 0.1 mM ATP. It was crystallized in hanging drops (2 µl of protein solution at 1 mg/ml with 2 µl of well solution). The well solutions ranged from 5-10% PEG 4000 and 0-5 mM MES pH 6.5. The protein crystals grew within a week and were between 50-200 µm on a side. They were initially transferred to a cryo-solution of well solution plus 30% (w/v) sucrose in 2 steps of increasing sucrose, and flash frozen in liquid nitrogen. With this treatment the crystals diffracted in a primitive orthorhombic space group (P222) with split spot profiles and diffracted to 2.8 Å with unit cell dimensions a = 77.6 Å, b = 97.3 Å, c = 121.5 Å. These data did not lead to a solution of the phase problem. Later, crystals were transferred from the crystallization drop into a cryo-solution of well solution plus 30% (v/v) dimethyl sulfoxide (DMSO) in four steps of increasing DMSO concentration and flash frozen. With the new cryo-condition, the crystals diffracted to 2.3 Å in a body-centered orthorhombic space group (I222) with unit cell dimensions a = 78.6 Å, b = 93.7 Å, c = 124.2 Å. A three-wavelength multiple anomalous dispersion (MAD) (Hendrickson, 1991) data set was collected on beam-line 5.0.2 at the ALS (Advanced Light Source, Lawrence Berkeley Laboratory, Berkeley) using a four panel ADSC CCD area detector, but the crystal showed significant decay during data collection. Data were processed and scaled using HKL2000 (Otwinowski and Minor, 1997). Subsequent data analysis was performed using the CNS package (Brunger et al., 1998). Two selenium sites were located by manual inspection of the anomalous Patterson map and 9 more were found through anomalous difference Fourier methods. Because of the crystal decay, phases were obtained with the set of 11 selenium sites using single anomalous dispersion (SAD) (Brodersen et al., 2000). An interpretable electron density map was obtained after density modification with solvent flipping with a solvent content of 50% (CNS) and an initial model was built using XtalView (McRee, 1999). The model was refined with CNS using the mlf target (maximum likelihood, F's) with 7% of the data excluded for the calculation of the cross-validating free R (Kleywegt and Brunger, 1996). 90% of all the built residues are in the most favorable regions of Ramachandran space and 10% are in the allowed regions (Laskowski et al., 1993). Statistics from phasing and refinement are shown in Table 6.1 and typical examples of the experimental electron density map generated with XtalView and Raster 3D (Merritt and Bacon, 1997) are shown in Figure 6.7. The structure has been deposited into the PDB with the accession code 1LTQ.

Table 6.1: Crystallographic Data and Refinement Statistics

Data collection	
Space Group	I222
Unit Cell (Å)	a=78.6 b=93.7 c=124.2
Wavelength (Å)	0.97873
Resolution (highest shell)	50-2.3 (2.38-2.30)
Total reflections	191,084
Unique reflections	38,288
Completeness (last shell) (%)	99.1 (97.7)
R_{merge} (last shell) (%)	7.3 (28.8)
Average $I/\sigma I$ (last shell)	33.4 (5.7)
Phasing	
Phasing Power	2.56
Figure of Merit (before/after DM)	0.40/0.96
Refinement	
R_{crsyt} (%)	23.6
R_{free} (7.5%) (%)	26.1
Total number of atoms	2532
Number of water molecules	144
Number of ligand/cryo atoms	27/4
r.m.s.d. bond length (Å)	0.007
r.m.s.d. bond angles (deg.)	1.4
Mean B value, mainchain (Å^2)	35.9
Mean B value, sidechain (Å^2)	37.6
Mean B value, solvent (Å^2)	47.3

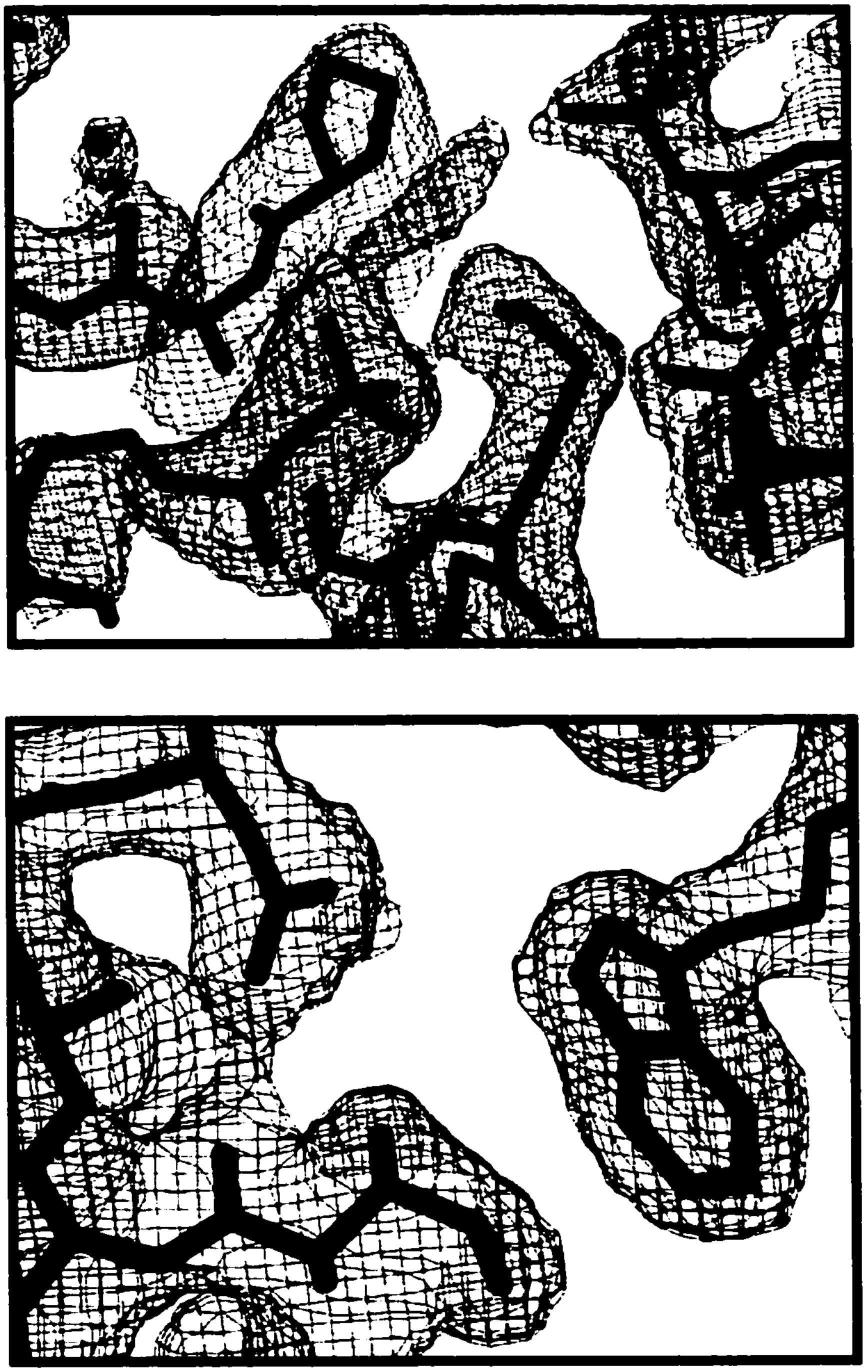


Figure 6.7: Examples of the experimental electron density map overlayed with the final atomic model of PNK.

Bibliography

Aggarwal, A. K. and D. A. Wah (1998). "Novel Site-Specific DNA Endonucleases. [Review] [40 Refs]." Current Opinion in Structural Biology **8**(1): 19-25.

Amitsur, M., R. Levitz, et al. (1987). "Bacteriophage T4 Anticodon Nuclease, Polynucleotide Kinase and Rna Ligase Reprocess the Host Lysine Trna." Embo J **6**(8): 2499-503.

Argast, G. M., K. M. Stephens, et al. (1998). "I-Ppoi and I-Crei Homing Site Sequence Degeneracy Determined by Random Mutagenesis and Sequential in Vitro Enrichment." Journal of Molecular Biology **280**(3): 345-53.

Athanasiadis, A., M. Vlassi, et al. (1994). "Crystal Structure of Pvuii Endonuclease Reveals Extensive Structural Homologies to Ecorv." Nat Struct Biol **1**(7): 469-75.

Ausbel, F. M., P. J. Thomas, et al. (1989). Current Protocols in Molecular Biology. New York, Wiley.

Baldwin, G. S., R. B. Sessions, et al. (1999). "DNA Cleavage by the Ecorv Restriction Endonuclease: Roles of Divalent Metal Ions in Specificity and Catalysis." J Mol Biol **288**(1): 87-103.

Beamer, L. J. and C. O. Pabo (1992). "Refined 1.8 a Crystal Structure of the Lambda Repressor-Operator Complex." J Mol Biol **227**(1): 177-96.

Belfort, M. and P. S. Perlman (1995). "Mechanisms of Intron Mobility." J. Biol. Chem. **270**(51): 30237-40.

Belfort, M., M. E. Reaban, et al. (1995). "Prokaryotic Introns and Inteins: A Panoply of Form and Function." J Bacteriol **177**(14): 3897-903.

Belfort, M. and R. J. Roberts (1997). "Homing Endonucleases - Keeping the House in Order." Nucleic Acids Res. **25**(17): 3379-3388.

Berger, J. M., S. J. Gamblin, et al. (1996). "Structure and Mechanism of DNA Topoisomerase Ii." Nature **379**(6562): 225-32.

Berkner, K. L. and W. R. Folk (1980). "Polynucleotide Kinase Exchange as an Assay for Class Ii Restriction Endonucleases." Methods Enzymol **65**(1): 28-36.

Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-42.

Berry, M. B., B. Meador, et al. (1994). "The Closed Conformation of a Highly Flexible Protein: The Structure of E. Coli Adenylate Kinase with Bound Amp and Amppnp." Proteins **19**(3): 183-98.

Bowie, J. U. and R. T. Sauer (1989). "Equilibrium Dissociation and Unfolding of the Arc Repressor Dimer." Biochemistry **28**(18): 7139-43.

Bozic, D., S. Grazulis, et al. (1996). "Crystal Structure of Citrobacter Freundii Restriction Endonuclease Cfr10i at 2.15 a Resolution." J Mol Biol **255**(1): 176-86.

Brennan, R. G., S. L. Roderick, et al. (1990). "Protein-DNA Conformational Changes in the Crystal Structure of a Lambda Cro-Operator Complex." Proc Natl Acad Sci U S A **87**(20): 8165-9.

Brodersen, D. E., E. de La Fortelle, et al. (2000). "Applications of Single-Wavelength Anomalous Dispersion at High and Atomic Resolution." Acta Crystallogr D Biol Crystallogr **56 (Pt 4)**: 431-41.

Brooks, B. R., R. E. Bruccoleri, et al. (1983). "Charmm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." Journal of Computational Chemistry **4**: 187-217.

Brudler, R., R. Rammelsberg, et al. (2001). "Structure of the I1 Early Intermediate of Photoactive Yellow Protein by Ftir Spectroscopy." Nat Struct Biol **8**(3): 265-70.

Brunger, A. (1993). "Assessment of Phase Accuracy by Cross Validation: The Free R Value. Methods and Applications." Acta Cryst **D49**: 24 - 36.

Brunger, A. T. (1992). X-Plor: A System for X-Ray Crystallography and Nmr. New Haven, Connecticut, Yale University Press.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & Nmr System: A New Software Suite for Macromolecular Structure Determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Cameron, V. and O. C. Uhlenbeck (1977). "3'-Phosphatase Activity in T4 Polynucleotide Kinase." Biochemistry **16**(23): 5120-6.

Chaconas, G. and J. H. van de Sande (1980). "5'-32p Labeling of Rna and DNA Restriction Fragments." Methods Enzymol **65**(1): 75-85.

Cheng, X., K. Balendiran, et al. (1994). "Structure of Pvuii Endonuclease with Cognate DNA." Embo J **13**(17): 3927-35.

Chevalier, B. S., R. J. Monnat, Jr., et al. (2001). "The Homing Endonuclease I-Crei Uses Three Metals, One of Which Is Shared between the Two Active Sites. [See Comments]." Nature Structural Biology **8**(4): 312-6.

Chevalier, B. S. and B. L. Stoddard (2001). "Homing Endonucleases: Structural and Functional Insight into the Catalysts of Intron/Intein Mobility. [Review] [144 Refs]." Nucleic Acids Research **29**(18): 3757-74.

Cho, H., W. Wang, et al. (2001). "Bef(3)(-) Acts as a Phosphate Analog in Proteins Phosphorylated on Aspartate: Structure of a Bef(3)(-) Complex with Phosphoserine Phosphatase." Proc Natl Acad Sci U S A **98**(15): 8525-30.

Chu, K., J. Vojtchovsky, et al. (2000). "Structure of a Ligand-Binding Intermediate in Wild-Type Carbonmonoxy Myoglobin." Nature **403**(6772): 921-3.

Clarke, N. D., L. J. Beamer, et al. (1991). "The DNA Binding Arm of Lambda Repressor: Critical Contacts from a Flexible Region." Science **254**(5029): 267-70.

Collet, J. F., V. Stroobant, et al. (1999). "Mechanistic Studies of Phosphoserine Phosphatase, an Enzyme Related to P-Type Atpases." J Biol Chem **274**(48): 33985-90.

Cornell, W. D., P. Cieplak, et al. (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." J Am Chem Soc **117**(19): 5179.

Cotton, F. S. and G. Wilkinson (1980). Advanced Inorganic Chemistry, John Wiley and Sons, New York.

Dean, A. B., M. J. Stanger, et al. (2002). "Inaugural Article: Zinc Finger as Distance Determinant in the Flexible Linker of Intron Endonuclease I-Tevi." Proc Natl Acad Sci U S A **99**(13): 8554-61.

Deibert, M., S. Grazulis, et al. (1999). "Crystal Structure of Muni Restriction Endonuclease in Complex with Cognate DNA at 1.7 a Resolution." Embo J **18**(21): 5805-16.

Deibert, M., S. Grazulis, et al. (2000). "Structure of the Tetrameric Restriction Endonuclease Ngomiv in Complex with Cleaved DNA." Nat Struct Biol **7**(9): 792-9.

Derbyshire, V., J. C. Kowalski, et al. (1997). "Two-Domain Structure of the Td Intron-Encoded Endonuclease I-Tevi Correlates with the Two-Domain Configuration of the Homing Site." Journal of Molecular Biology **265**(5): 494-506.

Dinner, A. R., G. M. Blackburn, et al. (2001). "Uracil-DNA Glycosylase Acts by Substrate Autocatalysis." Nature **413**(6857): 752-5.

Drew, H. R. and A. A. Travers (1985). "DNA Bending and Its Relation to Nucleosome Positioning." J Mol Biol **186**(4): 773-90.

Drouin, M., P. Lucas, et al. (2000). "Biochemical Characterization of I-Cmoei Reveals That This H-N-H Homing Endonuclease Shares Functional Similarities with H-N-H Colicins." Nucleic Acids Research **28**(22): 4566-72.

Duan, X., F. S. Gimble, et al. (1997). "Crystal Structure of Pi-Scei, a Homing Endonuclease with Protein Splicing Activity." Cell **89**(4): 555-64.

Dujon, B. (1989). "Group I Introns as Mobile Genetic Elements: Facts and Mechanistic Speculations--a Review." Gene **82**(1): 91-114.

Eddy, S. R. and L. Gold (1991). "The Phage T4 Nrdb Intron: A Deletion Mutant of a Version Found in the Wild." Genes & Development **5**(6): 1032-41.

Edgell, D. R. and D. A. Shub (2001). "Related Homing Endonucleases I-Bmoi and I-Tevi Use Different Strategies to Cleave Homologous Recognition Sites." Proceedings of the National Academy of Sciences of the United States of America **98**(14): 7898-903.

Elde, M., P. Haugen, et al. (1999). "I-Njai, a Nuclear Intron-Encoded Homing Endonuclease from Naegleria, Generates a Pentanucleotide 3' Cleavage-Overhang within a 19 Base-Pair Partially Symmetric DNA Recognition Site." European Journal of Biochemistry **259**(1-2): 281-8.

Elde, M., N. P. Willassen, et al. (2000). "Functional Characterization of Isoschizomeric His-Cys Box Homing Endonucleases from Naegleria." European Journal of Biochemistry **267**(24): 7257-66.

Ellison, E. L. and V. M. Vogt (1993). "Interaction of the Intron-Encoded Mobility Endonuclease I-Ppoi with Its Target Site." Molecular & Cellular Biology **13**(12): 7531-9.

Ferre-D'Amare, A. R., P. Pognonec, et al. (1994). "Structure and Function of the B/Hlh/Z Domain of Usf." Embo J **13**(1): 180-9.

Ferre-D'Amare, A. R., G. C. Prendergast, et al. (1993). "Recognition by Max of Its Cognate DNA through a Dimeric B/Hlh/Z Domain." Nature **363**(6424): 38-45.

Fersht, A. (1999). Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. New York, W. H. Freeman and Company.

Flick, K. E., M. S. Jurica, et al. (1998). "DNA Binding and Cleavage by the Nuclear Intron-Encoded Homing Endonuclease I-Ppoi." Nature **394**(6688): 96-101.

Flick, K. E., D. McHugh, et al. (1997). "Crystallization and Preliminary X-Ray Studies of I-Ppoi: A Nuclear, Intron-Encoded Homing Endonuclease from Physarum Polycephalum." Protein Sci **6**(12): 2677-80.

Fontanel, M. L., H. Bazin, et al. (1994). "Sterical Recognition by T4 Polynucleotide Kinase of Non-Nucleosidic Moieties 5'-Attached to Oligonucleotides." Nucleic Acids Res **22**(11): 2022-7.

Fried, M. and D. M. Crothers (1981). "Equilibria and Kinetics of Lac Repressor-Operator Interactions by Polyacrylamide Gel Electrophoresis." Nucleic Acids Res **9**(23): 6505-25.

Friedhoff, P., I. Franke, et al. (1999). "Cleavage Experiments with Deoxythymidine 3',5'-Bis-(P-Nitrophenyl Phosphate) Suggest That the Homing Endonuclease I-Ppoi Follows the Same Mechanism of Phosphodiester Bond Hydrolysis as the Non-Specific Serratia Nuclease." FEBS Letters **443**(2): 209-14.

Friedhoff, P., I. Franke, et al. (1999). "A Similar Active Site for Non-Specific and Specific Endonucleases." Nat Struct Biol **6**(2): 112-3.

Friedhoff, P., B. Kolmes, et al. (1996). "Analysis of the Mechanism of the Serratia Nuclease Using Site-Directed Mutagenesis." Nucleic Acids Research **24**(14): 2632-9.

Fulop, V., R. P. Phizackerley, et al. (1994). "Laue Diffraction Study on the Structure of Cytochrome C Peroxidase Compound I." Structure **2**(3): 201-8.

Galburt, E. A., B. Chevalier, et al. (1999). "A Novel Endonuclease Mechanism Directly Visualized for I-Ppoi." Nature Structural Biology **6**(12): 1096-9.

Garinot-Schneider, C., A. J. Pommer, et al. (1996). "Identification of Putative Active-Site Residues in the Dnase Domain of Colicin E9 by Random Mutagenesis." J Mol Biol **260**(5): 731-42.

Garner, M. M. and A. Revzin (1981). "A Gel Electrophoresis Method for Quantifying the Binding of Proteins to Specific DNA Regions: Application to Components of the Escherichia Coli Lactose Operon Regulatory System." Nucleic Acids Res **9**(13): 3047-60.

Gartenberg, M. R. and D. M. Crothers (1988). "DNA Sequence Determinants of Cap-Induced Bending and Protein Binding Affinity." Nature **333**(6176): 824-9.

Glennon, T. M., J. Villa, et al. (2000). "How Does Gap Catalyze the Gtpase Reaction of Ras? A Computer Simulation Study." Biochemistry **39**(32): 9641-51.

Goodrich-Blair, H. and D. A. Shub (1996). "Beyond Homing: Competition between Intron Endonucleases Confers a Selective Advantage on Flanking Genetic Markers." Cell **84**(2): 211-21.

Gouet, P., H. M. Jouve, et al. (1996). "Ferryl Intermediates of Catalase Captured by Time-Resolved Weissenberg Crystallography and Uv-Vis Spectroscopy." Nat Struct Biol **3**(11): 951-6.

Grazulis, S., M. Deibert, et al. (2002). "Crystal Structure of the Bse634i Restriction Endonuclease: Comparison of Two Enzymes Recognizing the Same DNA Sequence." Nucleic Acids Res **30**(4): 876-85.

Guallar, V., B. F. Gherman, et al. (2002). "Quantum Chemical Studies of Methane Monooxygenase: Comparision with P450." Curr Opin Chem Biol **6**(2): 236-42.

Guo, F., D. N. Gopaul, et al. (1997). "Structure of Cre Recombinase Complexed with DNA in a Site-Specific Recombination Synapse." Nature **389**(6646): 40-6.

Hajdu, J. (2000). "Single-Molecule X-Ray Diffraction." Curr Opin Struct Biol **10**(5): 569-73.

Hannan, J. P., S. B. Whittaker, et al. (1999). "Nmr Study of Ni2+ Binding to the H-N-H Endonuclease Domain of Colicin E9." Protein Science **8**(8): 1711-3.

Hasemann, C. A., E. S. Istvan, et al. (1996). "The Crystal Structure of the Bifunctional Enzyme 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase Reveals Distinct Domain Homologies." Structure **4**(9): 1017-29.

Heath, P. J., K. M. Stephens, et al. (1997). "The Structure of I-Crel, a Group I Intron-Encoded Homing Endonuclease. [See Comments]." Nature Structural Biology **4**(6): 468-76.

Hendrickson, W. A. (1991). "Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation." Science **254**(5028): 51-8.

Henkelman, G., G. Johannesson, et al. (2000). Progress on Theoretical Chemistry and Physics. S. D. Schwartz, Kluwer Academic: 269.

Henner, W. D., L. O. Rodriguez, et al. (1983). "Gamma Ray Induced Deoxyribonucleic Acid Strand Breaks. 3' Glycolate Termini." J Biol Chem **258**(2): 711-3.

Holm, L. and C. Sander (1993). "Protein Structure Comparison by Alignment of Distance Matrices." J Mol Biol **233**(1): 123-38.

Horton, J. R. and X. Cheng (2000). "Pvuii Endonuclease Contains Two Calcium Ions in Active Sites." J Mol Biol **300**(5): 1049-56.

Horton, N. C., K. J. Newberry, et al. (1998). "Metal Ion-Mediated Substrate-Assisted Catalysis in Type Ii Restriction Endonucleases." Proc Natl Acad Sci U S A **95**(23): 13489-94.

Horton, N. C. and J. J. Perona (2000). "Crystallographic Snapshots Along a Protein-Induced DNA-Bending Pathway." Proc Natl Acad Sci U S A **97**(11): 5729-34.

Huai, Q., J. D. Colandene, et al. (2001). "Structure of Naei-DNA Complex Reveals Dual-Mode DNA Recognition and Complete Dimer Rearrangement." Nat Struct Biol **8**(8): 665-9.

Ichiyanagi, K., Y. Ishino, et al. (2000). "Crystal Structure of an Archaeal Intein-Encoded Homing Endonuclease Pi-Pfui." Journal of Molecular Biology **300**(4): 889-901.

Jarvest, R. L. and G. Lowe (1981). "The Stereochemical Course of Phosphoryl Transfer Catalysed by Polynucleotide Kinase (Bacteriophage-T4-Infected Escherichia Coli B)." Biochem J **199**(1): 273-6.

Jeltsch, A., J. Alves, et al. (1993). "Substrate-Assisted Catalysis in the Cleavage of DNA by the Ecori and Ecorv Restriction Enzymes." Proc Natl Acad Sci U S A **90**(18): 8499-503.

Jilani, A., D. Ramotar, et al. (1999). "Molecular Cloning of the Human Gene, Pnkp, Encoding a Polynucleotide Kinase 3'-Phosphatase and Evidence for Its Role in Repair of DNA Strand Breaks Caused by Oxidative Damage." J Biol Chem **274**(34): 24176-86.

Jonsson, H., G. Mills, et al. (1998). Classical and Quantum Dynamics in Condensed Phase Simulations. B. J. Berne, G. Ciccotti and D. F. Coker. Singapore, World Scientific: 385.

Jurica, M. S. and B. L. Stoddard (1999). "Homing Endonucleases: Structure, Function and Evolution." Cell Mol Life Sci **55**(10): 1304-26.

Karimi-Busheri, F., G. Daly, et al. (1999). "Molecular Characterization of a Human DNA Kinase." J Biol Chem **274**(34): 24187-94.

Karimi-Busheri, F., J. Lee, et al. (1998). "Repair of DNA Strand Gaps and Nicks Containing 3'-Phosphate and 5'-Hydroxyl Termini by Purified Mammalian Enzymes." Nucleic Acids Res **26**(19): 4395-400.

Kim, J. L., D. B. Nikolov, et al. (1993). "Co-Crystal Structure of Tbp Recognizing the Minor Groove of a Tata Element." Nature **365**(6446): 520-7.

Kim, Y., J. H. Geiger, et al. (1993). "Crystal Structure of a Yeast Tbp/Tata-Box Complex." Nature **365**(6446): 512-20.

Kissinger, C. R. and D. K. Gehlhaar (1997). Epmr: A Program for Crystallographic Molecular Replacement by Evolutionary Search. La Jolla, CA, Agouron Pharmaceuticals.

Kleanthous, C., U. C. Kuhlmann, et al. (1999). "Structural and Mechanistic Basis of Immunity toward Endonuclease Colicins. [See Comments]." Nature Structural Biology **6**(3): 243-52.

Kleywegt, G. J. and A. T. Brunger (1996). "Checking Your Imagination: Applications of the Free R Value." Structure **4**(8): 897-904.

Klimasauskas, S., S. Kumar, et al. (1994). "Hhai Methyltransferase Flips Its Target Base out of the DNA Helix." Cell **76**(2): 357-69.

Ko, T. P., C. C. Liao, et al. (1999). "The Crystal Structure of the Dnase Domain of Colicin E7 in Complex with Its Inhibitor Im7 Protein." Structure **7**(1): 91-102.

Kollman, P. A., B. Kuhn, et al. (2001). "Elucidating the Nature of Enzyme Catalysis Utilizing a New Twist on an Old Methodology: Quantum Mechanical-Free Energy Calculations on Chemical Reactions in Enzymes and in Aqueous Solution." Acc Chem Res **34**(1): 72-9.

Kostrewa, D. and F. K. Winkler (1995). "Mg2+ Binding to the Active Site of Ecorv Endonuclease: A Crystallographic Study of Complexes with Substrate and Product DNA at 2 a Resolution." Biochemistry **34**(2): 683-96.

Koudelka, G. B. and P. Carlson (1992). "DNA Twisting and the Effects of Non-Contacted Bases on Affinity of 434 Operator for 434 Repressor." Nature **355**(6355): 89-91.

Koudelka, G. B., S. C. Harrison, et al. (1987). "Effect of Non-Contacted Bases on the Affinity of 434 Operator for 434 Repressor and Cro." Nature **326**(6116): 886-8.

Kowalski, J. C., M. Belfort, et al. (1999). "Configuration of the Catalytic Giy-Yig Domain of Intron Endonuclease I-Tevi: Coincidence of Computational and Molecular Findings." Nucleic Acids Research **27**(10): 2115-25.

Kuhlmann, U. C., G. R. Moore, et al. (1999). "Structural Parsimony in Endonuclease Active Sites: Should the Number of Homing Endonuclease Families Be Redefined?" FEBS Letters **463**(1-2): 1-2.

Lambowitz, A. M. and M. Belfort (1993). "Introns as Mobile Genetic Elements." Annu Rev Biochem **62**: 587-622.

Lambowitz, A. M., M. G. Caprara, et al. (1998). Group I and Group Ii Ribozymes as Rnps: Clues to the Past and Guides to the Future. Rna World Ii. Cold Spring Harbor, New York, Cold Spring Harbor Press.

Laskowski, R. J., M. W. Macarthur, et al. (1993). "Procheck: A Program to Check the Stereochemical Quality of Protein Structures." Journal of Applied Crystallography **26**(283-290).

Lau, E. Y. and T. C. Bruice (1999). "Active Site Dynamics of the Hhai Methyltransferase: Insights from Computer Simulation." J Mol Biol **293**(1): 9-18.

Levitt, M., M. Hirshberg, et al. (1995). "Potential Energy Function and Parameters for Simulations of the Molecular Dynamics of Proteins and Nucleic Acids in Solution." Computer Physics Communications **91**: 215-231.

Levitz, R., D. Chapman, et al. (1990). "The Optional E. Coli Prr Locus Encodes a Latent Form of Phage T4-Induced Anticodon Nuclease." Embo J **9**(5): 1383-9.

Lillehaug, J. R. (1978). "Inhibition of T4 Polynucleotide Kinase by the Atp Analog, Beta, Gamma-Imidoadenylyl 5'-Triphosphate." Biochim Biophys Acta **525**(2): 357-63.

Lillehaug, J. R. and K. Kleppe (1975). "Kinetics and Specificity of T4 Polynucleotide Kinase." Biochemistry **14**(6): 1221-5.

Lillehaug, J. R., R. K. Kleppe, et al. (1976). "Phosphorylation of Double-Stranded Dnas by T4 Polynucleotide Kinase." Biochemistry **15**(9): 1858-65.

Lowery, R., L. Hung, et al. (1992). "Properties of I-*Ppo*i: A Rare-Cutting Intron-Encoded Endonuclease." Promega Notes **38**: 8-12.

Lown, J. W. and L. W. McLaughlin (1979). "Nitrosourea-Induced DNA Single-Strand Breaks." Biochem Pharmacol **28**(10): 1631-8.

Luger, K., A. W. Mader, et al. (1997). "Crystal Structure of the Nucleosome Core Particle at 2.8 a Resolution." Nature **389**(6648): 251-60.

Lukacs, C. M., R. Kucera, et al. (2000). "Understanding the Immutability of Restriction Enzymes: Crystal Structure of Bglii and Its DNA Substrate at 1.5 a Resolution." Nat Struct Biol **7**(2): 134-40.

Mani, R. S., F. Karimi-Busheri, et al. (2001). "Physical Properties of Human Polynucleotide Kinase: Hydrodynamic and Spectroscopic Studies." Biochemistry **40**(43): 12967-73.

Mannino, S. J., C. L. Jenkins, et al. (1999). "Chemical Mechanism of DNA Cleavage by the Homing Endonuclease I-Ppoi." Biochemistry **38**(49): 16178-86.

Martin, A. M., M. D. Sam, et al. (1999). "Structural and Energetic Origins of Indirect Readout in Site-Specific DNA Cleavage by a Restriction Endonuclease." Nat Struct Biol **6**(3): 269-77.

McClarin, J. A., C. A. Frederick, et al. (1986). "Structure of the DNA-Eco Ri Endonuclease Recognition Complex at 3 a Resolution." Science **234**(4783): 1526-41.

McRee, D. E. (1999). "Xtalview/Xfit--a Versatile Program for Manipulating Atomic Coordinates and Electron Density." J Struct Biol **125**(2-3): 156-65.

Meijer, M., F. Karimi-Busheri, et al. (2002). "Pnk1, a DNA Kinase/Phosphatase Required for Normal Response to DNA Damage by Gamma-Radiation or Camptothecin in Schizosaccharomyces Pombe." J Biol Chem **277**(6): 4050-5.

Meiss, G., O. Gimadutdinow, et al. (2000). "Mechanism of DNA Cleavage by the DNA/Rna-Non-Specific Anabaena Sp. Pcc 7120 Endonuclease Nuca and Its Inhibition by Nuia." Journal of Molecular Biology **297**(2): 521-34.

Merritt, E. A. and D. J. Bacon (1997). "Raster3d: Photorealistic Molecular Graphics." Methods Enzymol **277**: 493-505.

Midgley, C. A. and N. E. Murray (1985). "T4 Polynucleotide Kinase; Cloning of the Gene (Pset) and Amplification of Its Product." Embo J **4**(10): 2695-703.

Miller, M. D., J. Cai, et al. (1999). "The Active Site of Serratia Endonuclease Contains a Conserved Magnesium-Water Cluster." Journal of Molecular Biology **288**(5): 975-87.

Miller, M. D., J. Tanner, et al. (1994). "2.1 a Structure of Serratia Endonuclease Suggests a Mechanism for Binding to Double-Stranded DNA." Nat Struct Biol **1**(7): 461-8.

Mueller, J. E., M. Brysk, et al. (1993). Homing Endonucleases. Nucleases. S. M. Linn, R. S. Lloyd and R. J. Roberts. Cold Spring Harbor, New York, Cold Spring Harbor Press. **2**: 111-143.

Murray, J. B. and W. G. Scott (2000). "Does a Single Metal Ion Bridge the a-9 and Scissile Phosphate Groups in the Catalytically Active Hammerhead Ribozyme Structure?" J Mol Biol **296**(1): 33-41.

Muscarella, D. E., E. L. Ellison, et al. (1990). "Characterization of I-Ppo, an Intron-Encoded Endonuclease That Mediates Homing of a Group I Intron in the Ribosomal DNA of Physarum Polycephalum." Mol Cell Biol **10**(7): 3386-96.

Muscarella, D. E. and V. M. Vogt (1989). "A Mobile Group I Intron in the Nuclear Rdna of Physarum Polycephalum." Cell **56**(3): 443-54.

Neutze, R., R. Wouts, et al. (2000). "Potential for Biomolecular Imaging with Femtosecond X-Ray Pulses." Nature **406**(6797): 752-7.

Newman, M., K. Lunnen, et al. (1998). "Crystal Structure of Restriction Endonuclease Bgli Bound to Its Interrupted DNA Recognition Sequence." Embo J **17**(18): 5466-76.

Newman, M., T. Strzelecka, et al. (1994). "Structure of Restriction Endonuclease Bamhi Phased at 1.95 a Resolution by Mad Analysis." Structure **2**(5): 439-52.

Newman, M., T. Strzelecka, et al. (1995). "Structure of Bam Hi Endonuclease Bound to DNA: Partial Folding and Unfolding on DNA Binding." Science **269**(5224): 656-63.

Novogrodsky, A. and J. Hurwitz (1966). "The Enzymatic Phosphorylation of Ribonucleic Acid and Deoxyribonucleic Acid. I. Phosphorylation at 5'-Hydroxyl Termini." J Biol Chem **241**(12): 2923-32.

Novogrodsky, A., M. Tal, et al. (1966). "The Enzymatic Phosphorylation of Ribonucleic Acid and Deoxyribonucleic Acid. Ii. Further Properties of the 5'-Hydroxyl Polynucleotide Kinase." J Biol Chem **241**(12): 2933-43.

Otwinowski, Z. and W. Minor (1997). "Processing of X-Ray Diffraction Data Collected in Oscillation Mode." Methods Enzymol **276**: 307-326.

Penner, M., I. Morad, et al. (1995). "Phage T4-Coded Stp: Double-Edged Effector of Coupled DNA and Trna-Restriction Systems." J Mol Biol **249**(5): 857-68.

Perona, J. J. and A. M. Martin (1997). "Conformational Transitions and Structural Deformability of Ecorv Endonuclease Revealed by Crystallographic Analysis." J Mol Biol **273**(1): 207-25.

Pietrokovski, S. (1998). "Modular Organization of Inteins and C-Terminal Autocatalytic Domains." Protein Sci **7**(1): 64-71.

Pingoud, A. and A. Jeltsch (1997). "Recognition and Cleavage of DNA by Type-Ii Restriction Endonucleases." Eur J Biochem **246**(1): 1-22.

Pingoud, A. and A. Jeltsch (2001). "Structure and Function of Type Ii Restriction Endonucleases." Nucleic Acids Res **29**(18): 3705-27.

Pommer, A. J., C. S., et al. (2001). "Mechanism and Cleavage Specificity of the H-N-H Endonuclease Colicin E9." Journal of Molecular Biology **314**(4): 735-49.

Raaijmakers, H., I. Toro, et al. (2001). "Conformational Flexibility in T4 Endonuclease Vii Revealed by Crystallography: Implications for Substrate Binding and Cleavage." Journal of Molecular Biology **308**(2): 311-23.

Raumann, B. E., M. A. Rould, et al. (1994). "DNA Recognition by Beta-Sheets in the Arc Repressor-Operator Crystal Structure." Nature **367**(6465): 754-7.

Redinbo, M. R., L. Stewart, et al. (1998). "Crystal Structures of Human Topoisomerase I in Covalent and Noncovalent Complexes with DNA." Science **279**(5356): 1504-13.

Reinisch, K. M., L. Chen, et al. (1995). "The Crystal Structure of Haeiii Methyltransferase Convalently Complexed to DNA: An Extrahelical Cytosine and Rearranged Base Pairing." Cell **82**(1): 143-53.

Ren, Z., B. Perman, et al. (2001). "A Molecular Movie at 1.8 a Resolution Displays the Photocycle of Photoactive Yellow Protein, a Eubacterial Blue-Light Receptor, from Nanoseconds to Seconds." Biochemistry **40**(46): 13788-801.

Rice, P. A., S. Yang, et al. (1996). "Crystal Structure of an Ihf-DNA Complex: A Protein-Induced DNA U-Turn." Cell **87**(7): 1295-306.

Richardson, C. C. (1965). "Phosphorylation of Nucleic Acid by an Enzyme from T4 Bacteriophage-Infected Escherichia Coli." Proc Natl Acad Sci U S A **54**(1): 158-65.

Richardson, C. C. (1981). Baceriophage T4 Polynucleotide Kinase. The Enzymes. P. D. Boyer. San Diego, Academic Press, Inc. **14:** 299-314.

Ridder, I. S. and B. W. Dijkstra (1999). "Identification of the Mg2+-Binding Site in the P-Type Atpase and Phosphatase Members of the Had (Haloacid Dehalogenase) Superfamily by Structural Similarity to the Response Regulator Protein Chey." Biochem J **339 (Pt 2)**: 223-6.

Ridder, I. S., H. J. Rozeboom, et al. (1999). "Crystal Structures of Intermediates in the Dehalogenation of Haloalkanoates by L-2-Haloacid Dehalogenase." J Biol Chem **274**(43): 30672-8.

Rupert, P. B. and A. R. Ferre-D'Amare (2001). "Crystal Structure of a Hairpin Ribozyme-Inhibitor Complex with Implications for Catalysis. [See Comments]." Nature **410**(6830): 780-6.

Sambrook, J., E. F. Fritsch, et al. (1989). Enzymes Used in Molecular Cloning. Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory Press. **1**.

Saraste, M., P. R. Sibbald, et al. (1990). "The P-Loop--a Common Motif in Atp- and Gtp-Binding Proteins." Trends Biochem Sci **15**(11): 430-4.

Schlichting, I., J. Berendzen, et al. (2000). "The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution." Science **287**(5458): 1615-22.

Schlichting, I. and K. Chu (2000). "Trapping Intermediates in the Crystal: Ligand Binding to Myoglobin." Curr Opin Struct Biol **10**(6): 744-52.

Schultz, S. C., G. C. Shields, et al. (1991). "Crystal Structure of a Cap-DNA Complex: The DNA Is Bent by 90 Degrees." Science **253**(5023): 1001-7.

Schumacher, M. A., K. Y. Choi, et al. (1994). "Crystal Structure of Laci Member, Purr, Bound to DNA: Minor Groove Binding by Alpha Helices." Science **266**(5186): 763-70.

Schweins, T., M. Geyer, et al. (1995). "Substrate-Assisted Catalysis as a Mechanism for Gtp Hydrolysis of P21ras and Other Gtp-Binding Proteins. [See Comments]." Nature Structural Biology **2**(1): 36-44.

Schweins, T., R. Langen, et al. (1994). "Why Have Mutagenesis Studies Not Located the General Base in Ras P21." Nature Structural Biology **1**(7): 476-84.

Scott, W. G. and J. B. Murray (2000). "Conventional and Time-Resolved Ribozyme X-Ray Crystallography." Methods Enzymol **317**: 180-98.

Seligman, L. M., K. M. Stephens, et al. (1997). "Genetic Analysis of the Chlamydomonas Reinhardtii I-Crei Mobile Intron Homing System in Escherichia Coli." Genetics **147**(4): 1653-64.

Silva, G. H., J. Z. Dalgaard, et al. (1999). "Crystal Structure of the Thermostable Archaeal Intron-Encoded Endonuclease I-Dmoi." Journal of Molecular Biology **286**(4): 1123-36.

Sirotkin, K., W. Cooley, et al. (1978). "A Role in True-Late Gene Expression for the T4 Bacteriophage 5' Polynucleotide Kinase 3' Phosphatase." J Mol Biol **123**(2): 221-33.

Slupphaug, G., C. D. Mol, et al. (1996). "A Nucleotide-Flipping Mechanism from the Structure of Human Uracil-DNA Glycosylase Bound to DNA." Nature **384**(6604): 87-92.

Soltis, D. A. and O. C. Uhlenbeck (1982). "Independent Locations of Kinase and 3'-Phosphatase Activities on T4 Polynucleotide Kinase." J Biol Chem **257**(19): 11340-5.

Stewart, L., M. R. Redinbo, et al. (1998). "A Model for the Mechanism of Human Topoisomerase I." Science **279**(5356): 1534-41.

Subramaniam, S., M. Lindahl, et al. (1999). "Protein Conformational Changes in the Bacteriorhodopsin Photocycle." J Mol Biol **287**(1): 145-61.

Teng, T. Y., V. Srajer, et al. (1997). "Initial Trajectory of Carbon Monoxide after Photodissociation from Myoglobin at Cryogenic Temperatures." Biochemistry **36**(40): 12087-100.

Thompson, J. F. and A. Landy (1988). "Empirical Estimation of Protein-Induced DNA Bending Angles: Applications to Lambda Site-Specific Recombination Complexes." Nucleic Acids Res **16**(20): 9687-705.

Torriglia, A., P. Perani, et al. (1998). "L-Dnase Ii, a Molecule That Links Proteases and Endonucleases in Apoptosis, Derives from the Ubiquitous Serpin Leukocyte Elastase Inhibitor." Mol Cell Biol **18**(6): 3612-9.

Travers, A. A. (1989). DNA Conformation and Protein Binding. Annual Review of Biochemistry. C. C. Richardson. Palo Alto, Annual Reviews, Inc. **58:** 427-452.

Turmel, M., C. Otis, et al. (1997). "Evolutionarily Conserved and Functionally Important Residues in the I-Ceui Homing Endonuclease." Nucleic Acids Research **25**(13): 2610-9.

Tyndall, C., J. Meister, et al. (1994). "The Escherichia Coli Prr Region Encodes a Functional Type Ic DNA Restriction System Closely Integrated with an Anticodon Nuclease Gene." J Mol Biol **237**(3): 266-74.

van de Sande, J. H. and M. Bilsker (1973). "Phosphorylation of N-Protected Deoxyoligonucleotides by T4 Polynucleotide Kinase." Biochemistry **12**(25): 5056-62.

van der Woerd, M. J., J. J. Pelletier, et al. (2001). "Restriction Enzyme Bsobi-DNA Complex: A Tunnel for Recognition of Degenerate DNA Sequences and Potential Histidine Catalysis." Structure (Camb) **9**(2): 133-44.

van Houten, V., F. Denkers, et al. (1998). "Labeling Efficiency of Oligonucleotides by T4 Polynucleotide Kinase Depends on 5'-Nucleotide." Anal Biochem **265**(2): 386-9.

Van Roey, P., L. Meehan, et al. (in press).

Van Roey, P., C. A. Waddling, et al. (2001). "Intertwined Structure of the DNA-Binding Domain of Intron Endonuclease I-Tevi with Its Substrate." EMBO Journal **20**(14): 3631-7.

Viadiu, H. and A. K. Aggarwal (1998). "The Role of Metals in Catalysis by the Restriction Endonuclease Bamhi." Nat Struct Biol **5**(10): 910-6.

Vineyard, G. H. (1957). J. Phys. Chem. Solids **3**: 121.

Wah, D. A., J. A. Hirsch, et al. (1997). "Structure of the Multimodular Endonuclease Foki Bound to DNA." Nature **388**(6637): 97-100.

Walsh, C. (2001). "Enabling the Chemistry of Life." Nature **409**(6817): 226-31.

Wang, L. K., C. D. Lima, et al. (2002). "Structure and Mechanism of T4 Polynucleotide Kinase: An Rna Repair Enzyme." Embo J **21**(14): 3873-3880.

Wang, L. K. and S. Shuman (2001). "Domain Structure and Mutational Analysis of T4 Polynucleotide Kinase." J Biol Chem **276**(29): 26868-74.

Warshel, A. (1991). Computer Modelling of Chemical Reactions in Enzymes and Solutions, Wiley, New York.

Westheimer, F. H. (1987). "Why Nature Chose Phosphates." Science **235**(4793): 1173-8.

Whitehouse, C. J., R. M. Taylor, et al. (2001). "Xrcc1 Stimulates Human Polynucleotide Kinase Activity at Damaged DNA Termini and Accelerates DNA Single-Strand Break Repair." Cell **104**(1): 107-17.

Winkler, F. K., D. W. Banner, et al. (1993). "The Crystal Structure of Ecorv Endonuclease and of Its Complexes with Cognate and Non-Cognate DNA Fragments." Embo J **12**(5): 1781-95.

Wittmayer, P. K., J. L. McKenzie, et al. (1998). "Degenerate DNA Recognition by I-Ppoi Endonuclease." Gene **206**(1): 11-21.

Wittmayer, P. K. and R. T. Raines (1996). "Substrate Binding and Turnover by the Highly Specific I-Ppoi Endonuclease." Biochemistry **35**(3): 1076-83.

Wu, L. and G. B. Koudelka (1993). "Sequence-Dependent Differences in DNA Structure Influence the Affinity of P22 Operator for P22 Repressor." J Biol Chem **268**(25): 18975-81.

Wu, L., A. Vertino, et al. (1992). "Non-Contacted Bases Affect the Affinity of Synthetic P22 Operators for P22 Repressor." J Biol Chem **267**(13): 9134-9.

Xu, S. Y. and I. Schildkraut (1991). "Isolation of Bamhi Variants with Reduced Cleavage Activities." J Biol Chem **266**(7): 4425-9.

Yang, S. W., A. B. Burgin, Jr., et al. (1996). "A Eukaryotic Enzyme That Can Disjoin Dead-End Covalent Complexes between DNA and Type I Topoisomerases." Proc Natl Acad Sci U S A **93**(21): 11534-9.

Zwieb, C. and S. Adhya (1994). "Improved Plasmid Vectors for the Analysis of Protein-Induced DNA Bending." Methods Mol Biol **30**: 281-94.

APPENDIX A

Crystallography of H98A I-PpoI/DNA substrate complex

```
rocky2.fhcrc.org
files: /net/barry/eric/ppo/ppoh98a/
       /net/barry/eric/ppo/ppoh98a/xplor/

5/29/98

DATA: ppoh98a.sca

RESOLUTION: 2.0
MOSAICITY: 0.344
COMPLETENESS: 98.7 (82.4)
R-SYM: 5.1 (23.9)
<I/sigma>: 14.2

SPACE GROUP: p3(1)21  #152
UNIT CELL: 113.866 113.866 89.040 90.000 90.000 120.000

   Shell    Summary of observation redundancies by shells:
Lower Upper No. of reflections with given # of observations
limit limit      0    1    2    3      4    5-6  total
 2.03  2.00     396  303  480  414    442   215   1854
  All hkl       603 1166 5779 13284 17017 7516  44762

   Shell            I/Sigma in resolution shells:
Lower Upper No. of reflections with I / Sigma less than
limit limit  0     1    2    3     5     10     20   >20  total
2.03  2.00 117   361  649  864  1201  1593  1816   38    1854
All hkl   1253  3241 5805 8187 12332 19613 28268 16494 44762
```

```
   Shell
Lower Upper  Average    Average     Norm.   Linear Square
limit limit     I     error    stat. Chi**2  R-fac  R-fac
2.03  2.00    295.7   63.9    63.3  1.058   0.239  0.220
All hkl      1606.6   66.2    54.6  1.075   0.051  0.040
```

----

ppoh98a.sca > ppoh98a.hkl > ppoh98a.mtz > ppoh98a.fobs
number of reflections output: 43498

----

ran setup_fr.inp -> ppoh98a.frf
working set: 40879 reflections
test set : 2613 reflections (6%)

6/1/98

generate.inp on starting model (410bref-h20.pdb from MJ / broken into ind. chain pdbs)
   > ppoh98a_gen.pdb & ppoh98a_gen.psf

Starting model pdbs can be found in "st_model" directory

rigid.inp returned R-values of > %50 indicating that one of the reciprocal lattice axis had been indexed backwards compared to previous data. Will re-index data with the

```
correction matrix [-1  0  0] to remedy this problem.
                  [ 0 -1  0]
                  [ 0  0 -1]
```

In addition, since re-indexing anyway, I am indexing out to 1.9A.

Decided to scale out to 1.93A since completeness dive bombed to ~%50 in 1.93-1.90 shell.

so lets just put the other data in a folder "oops" and treat this as a NEW DATA SET...

****

DATA: ppoh98a.sca

RESOLUTION: 1.93
MOSAICITY: 0.347
COMPLETENESS: 98.3 (82.2)
R-SYM: 5.3 (32.9)
<I/sigma>: 13.5

SPACE GROUP: p3121 #152
UNIT CELL: 113.860 113.860 89.024 90.000 90.000 120.000

```
   Shell     Summary of observation redundancies by shells:
Lower Upper  No. of reflections with given # of observations
limit limit   0    1    2    3     4     5-6   total
1.96  1.93   444  881  566  269   165   164    2045
 All hkl     857 2754 7283 14058 17497 7947   49539



   Shell          I/Sigma in resolution shells:
Lower Upper   No. of reflections with I / Sigma less than
```

```
limit limit  0    1    2     3     5     10     20    >20   total
1.96   1.93 248  620  1008  1312  1655  1956  2040    5     2045
All hkl   1747 4521 7956 11002 16026 24095 33020 16519 49539

   Shell
Lower Upper    Average     Average       Norm.    Linear Square
limit limit       I      error    stat.  Chi**2   R-fac  R-fac
1.96    1.93   195.2     73.9     73.5   1.090    0.329  0.302
All hkl       1473.2     66.8     56.2   1.075    0.053  0.040

------------------------------------------------------------

ppoh98a.sca > ppoh98a.Is > ppoh98a.mtz > ppoh98a.fobs
number of reflections output: 47774

------------------------------------------------------------

ran setup_fr.inp -> ppoh98a.frf
working set: 44905 reflections
test set : 2869 reflections (6%)

------------------------------------------------------------

Using generated structures from first try, plug into rigid.inp

rigid.inp: ppoh98a_gen.pdb > rigid_h98a.pdb

Start:
      R = 0.366
      Rfree = 0.371
Finish:
      R = 0.328
      Rfree = 0.332
```

6/2/98

check.inp > wa=260641

positional.inp (100 steps, wa=260641) > positional.out > pos_h98a.pdb
Start:
R = 33.8
Rfree = 34.2
Finish:
R = 32.2
Rfree = 34.4

Used pos_h98a.pdb to do solvent flattening
setup_bulksol.inp > created ppoh98a.sol (solvent corrected reflections)

```
XMASK: average mask radius around selected atoms  1.6960 A
XMASK: probe radius=  1.4000 shrink radius=  1.1000
XMASK: volume inside mask=   36.8000% (MASK<=0)
XMASK: volume outside mask=   63.2000% (MASK=1)
```

R-factor = 27.2

check.inp > check2.out > wa=273795

6/3/98

positional.inp (100 steps, wa=273795) > positional2.out > pos2_h98a.pdb

R = 26.7
Rfree = 29.2

calculated maps (fofc.map & 3fo2fc.map) and saw strong negative difference
density at mutated residues (H98/H298)
converted residues into alanines: new PDB > mod1_h98a.pdb

need to generate new structure file after modification of residue:
generate.inp > generate2.out > ppoh98a_gen2.psf & ppoh98a_gen2.pdb

positional.inp (100 steps, wa = 273795) > positional3.out > pos3_h98a.pdb
R = 26.7
Rfree = 29.2

6/4/98

three strong difference peaks (fofc-2.map) can be seen at the vertices of an equilateral triangle with the model Mn at the center.
Moved the Mn to my best guess density (strongest peak by ~1 sigma) in both positions.

also lots of similarly strong peaks found elsewhere... probably strongly bound waters...

positional.inp (100 steps, wa=273795) -> positional4.out > pos4_h98a.pdb

R = 26.8
Rfree = 28.9

calculated difference map (fofc-3.map) and saw strong negative density at new Mn location and positive density where the metal no longer was. Perhaps metal is in fact
Mg which would account for the lack of supporting density.

Regenerated structure files with Mg instead of Mn and placed the Mg in its original location where it appears that it would be well coordinated.

new difference maps (fofc-4.map) show no negative peaks indicating that the Mg is likely correct.
used x-solvate to add 362 waters. exported new pdbs:
mod4_h98a.pdb & waters.pdb

generate.inp > generate4.out > ppoh98a_gen4.pdb & ppoh98a_gen4.psf

check.inp > check3.out to determine new weight... wa=269319

positional.inp (100 cycles, wa=269319) > positional5.out > pos5_h98a.pdb
R = 21.5
Rfree = 23.8

mapnfo-mfc.inp >
fo-fc-5.map
2fo-fc-3.map

positive difference density in one monomer, but not the other (fofc-5.map) found on top of the phosphorous, the magnesium, and a neighboring water.

brefine.inp > brefine.out > bref_h98a.pdb
R = 20.7
Rfree = 23.7

histidine98 > activating the water or protonating the leaving group???
what's the pKa of the leaving oxygen? can water do it?

strain of imperfect octahedral coordination a drive towards cleavage and thus
a more perfect coordination?

6/8/98

took out metal and waters in active site in order to do an unbiased difference map
looking for positive peaks for metal and waters as well as nice difference peak for histidine residue.
First do rigid and 30 steps of positional, then calculate maps...

noligands.pdb > starting model
ppoh98a_nolig.pdb,psf > generate output
rigid_nolig.pdb > see rigid_nolig.out
check_nolig.out > wa=270585
positional_nolig.pdb > (30 rounds)
fofc-nolig.map

calculated 2fofc-6.map and fofc-6.map using bref_h98a.pdb look at phosphate position, can it be moved so as to better coordinate the metal???

6/30/98

moved one of the metals slighty to better position in density.

ran positional.inp to see if Rs go down any further with wa doubled to 500000 > pos6_h98a.pdb

R = 20.4

Rfree = 23.5

7/17/98

ran procheck on pos6_h98a.pdb & stored output in procheck directory: /net/barry/egalburt/ppo/ppoh98a/procheck/

APPENDIX B

Crystallography of I-PpoI/DNA Na^+ substrate complex

rocky2.fhcrc.org
files: /net/barry/egalburt/ppo/nometSYNC/
data: nomet_inv.sca
resolution: 2.1 A

```
************
* 4/22/99 *
************
```

+attempted to throw starting model (start.pdb) into rigid refinement

+neither version of the scaled data (hkl <-> -h-kl) yielded R's less than 55%

+noticed that unit cell parameters were all ~10% smaller than what is ordinarily seen for Ppo/DNA xtals

+tried to use epmr to find a molecular replacement solution reasoning that maybe things have moved around a bit

+other possibility is that distance was refined incorrectly and that led to erroneous unit cell lengths

```
************
* 4/23/99 *
************
```

+changed unit cell parameters in xplor to what I know unit cell usually is [113.86 113.86 89.04]

+using nomet_inv.fobs, R's refined to <40% using rigid refinement

+maybe shoot xtal again if unit cell parameters cause substantial problems

```
>rigid_all.pdb   Rf=39.9    R=39.2
>check.log             WA=253487
>pos1.pdb              Rf=39.9 R=37.2
>bulksol1.log          nomet_inv.sol
>pos2.pdb              Rf=32.9 R=31.2
>bgroup.pdb            Rf=32.3 R=30.1
>bref1.pdb       Rf=31.6    R=29.0 [fofc1.map]
>prpst.pdb
>slwcl1.pdb            Rf=30.7 R=27.6 [2fofc1.mbk/fofc2.mbk]
```

```
***********
* 4/25/99 *
***********
```

+NOTE+ calculated true resolution based upon number of reflections to be 2.07A. deviation from recorded resolution is due to erroneous xtal/detector distance which was discovered by observation of the unit cell parameters (see 4/23/99)

+NOTE+ setup_fr.log reports correct resolution with given HKLs (2.09A)

```
>genwat1.log [searchwaters1.pdb + slwcl1.pdb = addwat1.pdb(psf)]
      +added 166 waters

>pos3.pdb        Rf=27.4 R=25.2
>bref2.pdb       Rf=27.3 R=24.6 [2fofc2.mbk/fofc3.mbk]
```

```
***********
```

```
* 4/26/99 *
***********

>genwat2.log [searchwaters2.pdb + bref2.pdb = addwat2.pdb(psf)]
      +added 158 waters (total waters = 324)

>pos4.pdb             Rf=26.6 R=24.5
>bulksol2.log         nomet_inv.sol2
>bref3.pdb       Rf=25.5 R=22.9 [fofc4.mbk]

>gennuc.log [bref3-DNA.pdb + addnuc.pdb = gennuc.pdb(psf)]
      +added in bracketing nucleotides after modeling into
       omit density
      +set b=20.00

>pos5.pdb             Rf=23.4 R=20.7
>bref4.pdb       Rf=23.3 R=20.6 [2fofc5.mbk/fofc5.mbk]

>genwat3.log [bref4.pdb + searchwaters3.pdb] = addwat3.pdb(psf)]
      +added 72 waters (total waters = 396)

>pos6.pdb        Rf=22.7 R=20.3
>bulksol3.log    nomet_inv.sol3
>bref5.pdb Rf=21.6 R=19.1 [2fofc6.mbk/fofc6.mbk]
```

APPENDIX C

Crystallography of Apo I-PpoI

```
rocky2.fhcrc.org
/net/barry/egalburt/ppo/apo/
/net/barry/egalburt/ppo/apo/epmr/
/net/barry/egalburt/ppo/apo/xplor/merged/setb/

ppo_apo.sca
ppo_apo_p41212.sca
merged together two data sets > scalemerge2.log
      >>>> apomerge2.sca

resolution: 2.0
space group: P421212
unit cell: 52.8 52.8 278.8 90 90 90
completeness: 90.8 (74.8)
R-merge: 5.2 (20.4)
<I/σI>: 12.2 (7.0)

-----------------------------------------------------------

7/24/98

molecular replacement search model is a trimmed version
of 410bref-h20.pdb > apo_search.pdb
DNA binding beta sheets have been removed.
neither P422 or P4(3)2(1)2 resulted in high correlation
coefficients.

-----------------------------------------------------------
```

7/25/98

** redo MR search without -m2 because search model is already a dimer!

trying both P4(1)2(1)2 & P4(3)2(1)2

****

also trying to get AMoRe to run

ran SORT
ran PDBSET to renumber pdb
ran TABLING
     -> had to exclude metals because of discrepancies in atomsf.lib (ATOM ID Z)
     -> apo_search_nm.brk

****

no cc's over ~13% in either 4(3)2(1)2 or 4(1)2(1)2 from epmr

7/26/98

running epmr in space groups P4(3)22 and P4(1)22

****

continuation of AMoRe run...
ran ROTATION and got a series of rotation solutions (in 422 I believe)

running TRANSLATION on those ROT solutions in P43212 overnight.

7/27/98

epmr runs in both P4(3)22 and P4(1)22 result in no cc's over 12%
trying runs with full ppo structure (including DNA binding domains)

got the following solutions:

```
++++++++++
+ P43212
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Solution: rot = (77.6906 87.8668 182.041), trans = (52.346
14.1567 92.8458), cc = 0.215327, R = 0.593666
Solution: rot = (265.93 91.534 93.2739), trans = (12.0313 23.2622
84.2165), cc = 0.165465, R = 0.615453
Solution: rot = (88.2367 88.5907 358.386), trans = (44.1141
14.1274 255.942), cc = 0.166022, R = 0.622713
Solution: rot = (78.6724 89.0108 93.1464), trans = (30.4035
0.0662612 48.2549), cc = 0.144674, R = 0.639209
Solution: rot = (258.318 91.5025 358.362), trans = (28.518
32.7572 96.2916), cc = 0.148477, R = 0.637244
Solution: rot = (267.958 91.4889 90.0561), trans = (45.1324
30.8533 113.046), cc = 0.140589, R = 0.635192
Solution: rot = (268.222 91.9696 182.198), trans = (26.9994
12.2559 130.822), cc = 0.158371, R = 0.615907
Solution: rot = (267.499 91.4944 180.44), trans = (49.4214
40.3502 245.134), cc = 0.141914, R = 0.617451
Solution: rot = (77.689 87.6332 182.863), trans = (26.0261
40.6369 104.259), cc = 0.132807, R = 0.629166
```

```
Solution: rot = (258.798 90.5918 357.699), trans = (0.47866
31.0112 78.0351), cc = 0.14282, R = 0.631705

++++++++++
+ P41212
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Solution: rot = (266.992 91.8973 182.819), trans = (0.600588
38.6991 23.4103), cc = 0.376943, R = 0.536302
Solution: rot = (258.238 92.0652 177.165), trans = (0.621102
14.163 46.5734), cc = 0.371835, R = 0.535764
Solution: rot = (78.2974 88.0435 93.1773), trans = (12.2577
25.8193 93.1064), cc = 0.372637, R = 0.535716
Solution: rot = (86.5147 88.0273 176.607), trans = (25.712
12.2787 127.746), cc = 0.231315, R = 0.589084
Solution: rot = (269.067 119.207 245.305), trans = (23.3001
6.40883 226.671), cc = 0.109105, R = 0.653591
Solution: rot = (87.0268 88.1632 267.158), trans = (38.7082
0.604171 201.641), cc = 0.21945, R = 0.605679
Solution: rot = (79.1882 88.3514 274.197), trans = (40.6691
27.0613 182.536), cc = 0.204659, R = 0.611733
Solution: rot = (258.162 92.0992 88.2279), trans = (38.6876
0.604649 186.869), cc = 0.20071, R = 0.599413
Solution: rot = (258.791 92.0153 267.133), trans = (40.623
25.7994 183.768), cc = 0.219353, R = 0.600641
Solution: rot = (258.698 91.9531 176.535), trans = (27.0301
40.6106 10.9894), cc = 0.217443, R = 0.602178
```

tried full model with P4(1)22 and P4(3)22 with cc's less than 23% and amongst noise.

will rerun full_P4(1)2(1)2 in order to write out multiple structures with high cc's to compare...
files: rerun37.*

also rerun37_2.*

8/2/98

reindexed and scaled data in P41212 : ppo_apo_p41212.sca > p41212.fobs (w/ sclpckxplor*)
took the pdbs with the highest cc's (~37%) from epmr and copied them to best(123).pdb in /xplor
look and see if they refine well (or at all...)

```
SOLUTIONS+++++++++++++++++++++++++++++++++++++++++++++++++

(epmr_full_P41212.log > best1.pdb)
Solution: rot = (266.992 91.8973 182.819), trans = (0.600588
38.6991 23.4103), cc = 0.376943, R = 0.536302
Solution: rot = (258.238 92.0652 177.165), trans = (0.621102
14.163 46.5734), cc = 0.371835, R = 0.535764
Solution: rot = (78.2974 88.0435 93.1773), trans = (12.2577
25.8193 93.1064), cc = 0.372637, R = 0.535716

(rerun.log > best2.pdb)
Solution: rot = (87.0511 88.0333 177.27), trans = (25.8352
12.2602 186.059), cc = 0.376608, R = 0.535892

(fourthrun.log -> best3.pdb)
Solution: rot = (267.004 91.909 92.8273), trans = (40.6298
27.0403 93.232), cc = 0.376969, R = 0.536046

+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

8/3/98

looked at three solutions to see if they are all actually the same solution.

best2 is a translation of 1/2 the unit cell in the long dimension of best1. this arises from the ambiguity
introduced by a screw axis as to where the unit cell is. If the whole molecule is to be in the unit cell, this
leaves 4 possible placed to place a solution within the unit cell defined from best1.

best3 is just a symmetry copy of best1.

this certainly lends credence to the solution from epmr.

```
===========================================================
NOTES REGARDING REFINEMENT OF MERGED APO PPO DATA
/net/barry/egalburt/ppo/apo/xplor/merged/setb/
===========================================================

*reset model B's to 20.00

*STRICT NCS
***********************************************************

RIGID
     rigmon > rigmon2 > rigmon3 > R=47.2 Rf=45.5

POSITIONAL
     posmon > posmon2 > posmon3 > R=42.5 Rf=44.2

BUILD
     model4.brk
```

```
POSITIONAL
      posmon4 > R=43.3 Rf=43.9

SOLVENT CORRECT
      apomerge.sol

POSITIONAL
      posmon4sol > posmon5sol > posmon6sol > R=34.9 Rf=36.8

OMIT MAP/MODEL
      omitmodel.pdb

BGROUP
      bgroupmon > R=32.7 Rf=25.3

*RESTRAIN NCS
***********************************************************

POSITIONAL
      posrest2 > posrest3 > R=32.4 Rf=34.5

BREFINE
      bref > R=31.1 Rf=34.2

BUILD
      modeldim.pdb

POSITIONAL
      posmod4 > R=32.4 Rf=34.7

*FREE NCS
***********************************************************
```

```
POSITIONAL
     posmod4free60 > R=29.7 Rf=33.3

OMIT MAP to look for A89-92 and B57-61
     omit2.map

BUILD
     remod2.pdb

POSITIONAL
     remod2pos > R=29.8 Rf=32.9

ADDED WATERS (40/40)
     poswat1 > R=28.4 Rf=31.9

BREFINE
     brefwat1 > R=27.7 Rf=31.2

BUILD
     model6.pdb

POSITIONAL
     poswat1mod > R=28.0 Rf=31.1

ADD WATERS (114/154)
     addwat2.pdb

POSITIONAL
     poswat2 > R=25.6 Rf=28.6

BREFINE
     brefwat2 > R=25.1 Rf=28.2
```

```
BUILD
     model7m.pdb

POSITIONAL
     posmod7m > posmod7ma > R=25.0 Rf=27.9

ADDED WATERS (102/253)
     addwat3.pdb

POSITIONAL
     poswat3 > R=24.4 Rf=27.0

BREFINE
     brefwat3 > R-23.7 Rf=26.5

SOLVENT CORRECTION
     apomerge.sol2

POSITIONAL
     poswat3bulk2 > R=23.0 Rf=26.1

ADD SULFATES (1&2), WATERS, BUILD - mutated arg61 to ala61
     (1030.pdb)

POSITIONAL
     (posso4_15) > R=22.3 Rf=25.7

ADDED WATERS (50/303)
     addwat4.pdb

POSITIONAL
     poswat4 > R=22.1 Rf=25.4
```

```
BUILD
     model9.xpl

POSITIONAL
     posmod9 > R=21.9 Rf=25.0

BREFINE
     brefmod9 > R=21.2 Rf=24.8

BUILD - mutate ala 61 back to arg61
     loopmod.pdb

POSITIONAL
     posmodloop > R=21.4 Rf=24.7

BREFINE
     brefloop > R=21.2 Rf=24.5

PROCHECK
     brefloop_0*.ps

BUILD/POSITIONAL
     posmod119 > posmod119b > R=21.3 Rf=24.5

REMOVE WATERS
     watdel.xpl

POSITIONAL
     poswatdel > R=21.6 Rf=24.6

ADD WATERS (336)
     1124.xpl
```

```
POSITIONAL
     pos1124 > R=21.4 Rf=24.1

BUILD
     mod1124.xpl

POSITIONAL
     posmod1124 > R=21.8 Rf=24.1

BREFINE
     bref1124 > R=21.3 Rf=23.8

ADDED TWO MORE SULFATES (3&4)
     sulf1124.xpl

POSITIONAL
     possulf1124 > R=21.1 Rf=23.6

BREFINE
     brefsulf1124 > R=21.0 Rf=23.5

REMOVED ONE SULFATE BECAUSE OF NEGATIVE DENSITY...BEST MODEL...
     possulfdel.xpl > R=20.7 Rf=23.6

BUILD/POSITIONAL
     pos1125 > R=20.7 Rf=23.5

ADDED 4th SULFATE/POSITIONAL
     possulf1125 > R=20.4 Rf=23.4

BUILD
     posref1.xpl
```

```
XXXX    POSITIONAL
        posrefl > R=19.7 Rf=23.8  XXXX

TORSIONAL
        torrefl > R=19.6 Rf=23.2

XXXX    BUILD/POSITIONAL
        pos1127 > R=20.7 Rf=23.5  XXXX

BUILD/POSITIONAL
        pos1128a > R=19.5 Rf=23.0

ONE LAST WATER ANALYSIS (REMOVAL OF BAD WATERS) + POSITIONAL
        pos1129 > R=19.9 Rf=22.9


===========================================================
BEST MODEL: pos1129.xpl   R=19.9 Rf=22.9
===========================================================

PROCHECK
        pos1129*.ps
        > BGLN 70 & BPRO 59 need some attention *********
        > PRO 49 is CIS
        > main chain bond angles >10.0 from ideal : T60 & V136
        > looks good
```

APPENDIX D
Crystallography of L116A I-PpoI/DNA complex
(M. Jurica and D. Erho)

```
8-24-98
rocky2.fhcrc.org
~mjurica/L116A/
Data: Note: this is data collected with a 1 degree oscillation.
There were less than 10 full reflections on each frame.  But we
got it to scale so we're going to try it out anyways.
>>>> ppol116a.sca

Total number of reflections:   120188
Resolution Limit: 3.1
Mosaicity: 0.9 on average

Space group C2
Unit Cell: 182.0 73.1 92.7 90 95.4 90
Unit cell volume= 1227825.9
Vm = 1227825.9/2 x4(18300 + 18300 + 13860)=3.0416
(solvent content = -1.488214(vm2) + 28.29413(vm) - 16.33381)=>54%
=>expect 2 dimers/ A.U.

Ave I/sigI  All films   7.4

   Shell  Summary of observation redundancies by shells:
Lower Upper No. of reflections with given # of observations
limit limit   0      1      2      3      4      total
All hkl      2559   4256 10649   4483   319    19707

   Shell
Lower Upper  Average  Average    Norm.  Linear Square
```

```
limit limit    I    error   stat. Chi**2  R-fac  R-fac
All hkl     920.3  129.2   97.3  1.387   0.114  0.110

   Shell            I/Sigma in resolution shells:
Lower Upper   No. of reflections with I / Sigma less than
limit limit   0    1    2    3    5    10    20    >20  total
 All hkl    3.3 12.2 22.7 30.4 42.8 66.6 88.3    0.2   88.5
```

Try MOLECULAR REPLACEMENT in AMORE for starters: 2 different models:

1- Search model: apo_swmon.pdb=>this is monomer1, residues 2-146 and the C-domain
of monomer2, residues 347 to end. Plus 2 zinc ions.

```
CELL 93 93 93
SPHERE 3.0 25
```

Ran from 10 to 4 A
Ran from 10 to 5 A
Ran from 10 to 6 A

These are our top 4 solutions

```
 SOLUTIONF     1  100.47   35.72   85.61
                    0.2708 -0.0003  0.3080 55.5 41.7   1
 SOLUTIONF     1  185.39   90.65  214.09
                    0.4177  0.0496  0.6278 55.5 41.7   2
 SOLUTIONF     1  194.88  100.15  165.56
                    0.6490  0.3571  0.9043 55.5 41.7   3
 SOLUTIONF     1  299.05  162.23  130.09
                    0.9972  0.0882  0.3628 55.5 41.7   4
```

8-25-98
So these solutions pack to form two dimers with room for DNA. So it looks like our solution!!

Take this into XPLOR
~mjurica/L116A/xplor

Create a free R data set: ppol116a.frf
=>6% of data reserved for free R
working set= 17852 reflections
test set= 1109 reflections

Generate: 824ppo_gen.pdb

Rigid body refinement: 825rigid2.pdb
first a.u., then dimers, then globular monomers:

Start:
Rfree= 0.446
R= 0.443
Finish:
Rfree= 0.444
R= 0.439

Calculate maps and have a looksee.
Can already see some density for the DNA!!! YES!

Go into RAVE to try averaging: See memo in ~mjurica/L116A/rave

8-26-98
Rave map looks very good. Will start to build our DNA model. Start with Karen's oligo7 bent DNA model. Make changes to get to oligo6 sequence, but with out including the mistake at the 3rd position.
DNA has the following sequence:

X   Z   5'-TTGGCTACCTTAAGAGAGTCA-3'
W   Y   3'-AACCGATGGAATTCTCTCAGT-5'

The DNA is pseudocontinuous in the maps, so we'll have to be carefull about the terminii. Also, density for DNA is weak in the middle of the oligo. Kind of interesting?

So fit the DNA to density for the AB dimer and will rotate so that it fits into the CD dimer. It seems that A and D are more closely coorelated, so we'll match them up for the DNA orientation.

8-27-98
Fit the protein of dimer1 to averaged map. Then rotated dim1+dna1 onto second dimer. Put that into generate to start a little refinement in xplor.

NOTE: 4 protein monomers with segid: APPO, BPPO, CPPO, DPPO
      2 DNA duplexes with strands Z, Y, X, W

      2 dimers are then A+B , with DNA Z+Y
                 and   C+D, with DNA X+W

Generate: 827ppo_gen.pdb, psf
          827dna_gen.pdb, psf

Rigid:first a.u., then dimers=> 827rigid.pdb
Start:
  Rfree=   0.432
  R=   0.416
Finish:
  Rfree=   0.429

R= 0.415

CHECK: wa= 312785

Prpst: 827prpst.pdb
wa=200000
Start:
Rfree= 0.427
R= 0.425
Finish:
Rfree= 0.418
R= 0.402

1)Positional refinement with 4fncs: 827pos3.pdb
wa=200000 cycles=150 NCS=100MC, 50SC
Start:
Rfree= 0.418
R= 0.402
Finish:
Rfree= 0.399
R= 0.367

OK- we'll try a bulk solvent correction because we broke 40 and we have a complete model. => ppol116a.sol

Start:
Rfree= 0.3986
R= 0.3674

```
 XMASK: probe radius=  1.4000 shrink radius=  1.1000
```

```
 XMASK: volume inside mask=   41.0545% (MASK<=0)
 XMASK: volume outside mask=   58.9455% (MASK=1)
ppoFinish:
  Rfree=   0.3775
  R=       0.3451
```

2)Positional refinement with 4fncs: 827pos6.pdb
wa=200000 cycles=150 NCS=100MC, 50SC
Start:
  Rfree=   0.378
  R=   0.345
Finish:
  Rfree=   0.365
  R=   0.327

1)B group refinement with 4nfcs: 827bgp3.pdb
Refines each residues as a group of mc and sc atoms, and each nucleotide as
a group. The metals are all refined as one group.
wa=200000 steps=15 NCS=100MC, 50SC
Start:
  Rfree=   0.365
  R=   0.327
Finish:
  Rfree=   0.338
  R=   0.300

9-2-98

After examining maps, it looks like #1-the dna is cleaved, #2-a Mg ion is bound in the active site, and #3 (most disconcerting) this is a 20 bp oligo, not a 21 bp oligo. The correct scissile phosphates are lining up at all 4 active sites but there is a clashing bp with the symmetry mate at position Z1/Y21 and X1/W21. So I'm going to try the following model and go back to refinement:

Remove Z1/Y21 and X1/W21 bp's.

Create cleaved dna model.

Add mg to the model.

Generate:

92dna_gen.pdb,psf

92mg_gen.pdb,psf

(prot remains the same)

2)Positional refinement with 4fncs: 92pos.pdb

wa=200000 cycles=150 NCS=100MC, 50SC

Start:

Rfree=	0.335
R=	0.299

Finish:

Rfree=	0.328
R=	0.288

1)B group refinement with 4nfcs: 92bgrp.pdb

Refines each residues as a group of mc and sc atoms, and each nucleotide as a group. The metals are all refined as one group.

wa=200000 steps=15 NCS=100MC, 50SC

Start:

Rfree=	0.328
R=	0.288

Finish:

Rfree=	0.327
R=	0.285

Refined chain APPO by hand in quanta. Saved as dave93_a.pdb

APPENDIX E

Crystallography of T4 Polynucleotide Kinase

```
sussman.fhcrc.org
data files: /home/eric/pnk_data/
ALS 5.0.2, HKL2000
      alspnk_*.img.bz2
      alspnk_*.x
      peak.sca remote.sca
      peak.out remote.out

space group: I222
unit cell: 78.585 93.684 124.165 90 90 90
volume: 914122 A^3
Vm: 3.4 A^3/kDa

sad phasing: /home/eric/pnk_sad
CNS, XtalView
      peak.sca > to_cns > peak.hkl
      peak.hkl > scale.inp > sad_scale.hkl
      sad_scale.hkl > analyse23.inp > analyse.*
      sad_scale.hkl > patterson.inp > anom_patterson.map
      found 6 selenium positions manually > 6sites.sdb
      found 5 more by difference fourier >
            6site_anom_diff_positive.peaks > 11sites.sdb
      11sites.sdb, sad_scale.hkl > sad_phase23.inp >
            11sites_phase.hkl
      sad_scale.hkl, 11sites_phase.hkl >
            density_modify23.inp (50% solvent) >
                  density_modify_11sites_50.hkl,
                  density_modify_11sites_50.map
```

```
chain tracing: /home/eric/pnk_sad/trace/
XtalView
     c-alpha trace: ca_pnk_3_22.pdb
     poly alananine trace: polyala_trace.pdb
     series of residue by residue building:
          build_10res.pdb (build_3_22.pdb)
          build_20res.pdb
          build_40res.pdb
          build_60res.pdb
          build_80res.pdb
          build_100res.pdb
          build_120res.pdb
          build_150res.pdb (build_3_23.pdb)
          build_170res.pdb
          build_200res.pdb
          build240res.pdb
          build270res.pdb
          build301res.pdb (build_complete.pdb)

refinement: /home/eric/pnk_sad/trace/refine/
CNS

3/25/02

-made cross validation data set (7.5%)
     -make_cv.inp
     -density_modify_11sites_50.hkl > dm_50.cv
-generated pdb and mtf files
     -generate_easy.inp
     -build_complete.pdb > generate.pdb(mtf)
-checked model
     -model_stats.inp
     -model_stats_start.*
```

```
      -R-values:
       initial r = 0.3865 free_r = 0.3950
       after B-factor and/or bulk solvent correction
               r = 0.3440 free_r = 0.3530
      -many high energy geometries
-minimized model
      -minimize.inp
      -generate.pdb > minimize.pdb
      -REMARK starting r = 0.3441 free_r = 0.3532
       REMARK final    r = 0.3002 free_r = 0.3385
-rechecked model
      -model_stats.inp
      -model_stats_min.*
      -geometry looks much better
-simulated annealing (1 run)
      -anneal.inp
      -minimize.pdb > anneal.pdb
      -REMARK starting r = 0.2943 free_r = 0.3325
       REMARK final    r = 0.2930 free_r = 0.3289

-changed MET residues to MSE residues; tell CNS anomalous
scattering f'/f''; mlf refinement
      -generate_easy.inp
      -mse_anneal.pdb > mse_generate.pdb(mtf)
-minimized with the 12 anomalous scatterers to check refinement
(20 steps)
      -minimize.inp
      -mse_anneal.pdb > minimize_mse.pdb
      -REMARK starting r = 0.2973 free_r = 0.3296
       REMARK final    r = 0.2957 free_r = 0.3300
      -go back to mse_anneal.pdb
-group Bfactor refinement
      -bgroup.inp
```

```
     -mse_generate.pdb > bgroup.pdb
     -REMARK starting r = 0.2962 free_r = 0.3293
      REMARK final    r = 0.2723 free_r = 0.3011

3/27/02
-manual rebuild residue by residue using fofc.fsf
     -A25K
     -added A48 (ASP) and K51 to loop1 region
     -G74A (D)
     -A102E
     -A121K
     -A124S
     -A180D, A183K salt bridge?
     -A218E
     -A230K
     -bgroup.pdb > rebuild.pdb

-regenerate
     -rebuild.pdb > rebuild_generate.pdb
-minimize
     -rebuild_generate.pdb > minimize_rebuild.pdb
     -REMARK starting r = 0.2976 free_r = 0.3239
      REMARK final    r = 0.2920 free_r = 0.3217
-bindividual
     -minimize_rebuild.pdb > bindividual.pdb
     -REMARK starting r = 0.2919 free_r = 0.3203
      REMARK final    r = 0.2785 free_r = 0.3043
-modelled diphosphate (dip.pdb,dip.top,dip.param)
     -generate > generate_dip.pdb(mtf)
-minimize
     -generate_dip.pdb > minimize_dip.pdb
     -REMARK starting r = 0.2756 free_r = 0.3022
      REMARK final    r = 0.2640 free_r = 0.2951
```

```
-bindividual
      -minimize_dip.pdb > bind_dip.pdb
      -REMARK starting r = 0.2633 free_r = 0.2942
       REMARK final    r = 0.2588 free_r = 0.2869

3/28/02
-picked waters (41)
      -water_pick_1.pdb(mtf)
      -minimize_wat1.pdb
      -bind_wat1.pdb
      -REMARK starting r = 0.2484 free_r = 0.2832
       REMARK final    r = 0.2477 free_r = 0.2807
-manual rebuild and look at maps
      -rebuild2.pdb
-picked more waters (122 total)
      -bind_wat2.pdb
      -REMARK starting r = 0.2454 free_r = 0.2683
       REMARK final    r = 0.2445 free_r = 0.2688
-picked more waters (194 total)
      -bind_wat3.pdb
      -REMARK starting r = 0.2395 free_r = 0.2648
       REMARK final    r = 0.2375 free_r = 0.2641


XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
4/2/02
-removed some waters, moved some side chains, added some waters
manually
      -bind_man.pdb (removed because of OXT/TER problems)
      -REMARK starting r = 0.2411 free_r = 0.2645
       REMARK final    r = 0.2411 free_r = 0.2636
-going over each residue; noting changes
      -THR 7
```

-GLU 21
-PRO 27
-ASP 35
-ILE 41 removed erroneous OXT
-HIS 44 flipped base
-removed waters from gap between 48-51
-TYR 73
-GLU 91
-GLU 118
-LYS 121
-LYS 125
-126/127 peptide bond flip
-THR 128
-ASP 134
-176,177,178 mainchain
-ASP 187
-VAL 210
-LYS 218 two conformers?
-LYS 230
-GLU 233

something is wack with MSE residue peptide bonds

Xfit doesn't recognize MSE so it writes out PDBs with TER card...not good

***added MSE to dict.pdb that Xfit uses to write out pdbs...

still doesn't work...

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

MAKE SURE TO REMOVE TER CARDS THAT XFIT PUTS IN BEFORE RUNNING CNS

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

@$#%@#$ go back to bind_wat3.pdb R = 24.1 Rf = 26.4
and use maps made with this pdb > (2)fofc_wat3.fsf

-going over entire model; note changes
*removing water molecules from putative active sites
for unbiased maps
-ILE 4
-ILE 5
-THR 7
-ILE 8
-GLU 21
-PRO 27
-ASP 35
-HIS 44
-LYS 51/ALA 52
-ASP 66
-ILE 71
-TYR 73
-GLU 91
-GLU 108
-GLU 118
-LYS 121
-ARG 128/GLY 127
-THR 128
-ASP 134
-SER 138
-MSE 139
-PRO 156
-ILE 163
-ASP 167/GLY 168
-LYS 172
-PRO 178/ALA 179

-ASP 187

-ILE 189

-VAL 210

-GLU 219

-ASP 255

-ILE 266

-ILE 276

-ASP 278

-removing TER cards from check1.pdb and HETATM from MSE residues

-minimize and bindividual

-bind_check1.pdb

-REMARK starting r = 0.2467 free_r = 0.2668
REMARK final r = 0.2404 free_r = 0.2598

-renumbered waters so that I may "go to" them with xfit starting with residue #500

-rechecked maps, removed waters that were not in 2fofc density,or that seemed to be in too large density (still +fofc density on them > ADP?, PO4?)

-T7,E21,D35,H44,Y73,E91,K121,T128,D134,S128,M139,D167,K172, D187,V210,E247,D278

-also noted residues missing side chains

-D48,Y52,K54,K55,D76,K101,Y179,E182,K264,E219

-generated check2.pdb after removing TER cards, etc....

-minimize and bindividual

-bind_check2.pdb

-REMARK starting r = 0.2424 free_r = 0.2676
REMARK final r = 0.2385 free_r = 0.2631

-tried modelling in a phosphate in large positive peak near H109, could it be DMSO?

-bind_photest.pdb

-REMARK starting r = 0.2348 free_r = 0.2603
REMARK final r = 0.2346 free_r = 0.2584

4/7/02

-phosphate model shows strong negative density in difference map...try dmso

-dmso seems to fit better even though free_r goes up a bit...

    -bind_dmso2.pdb
    -REMARK final   r = 0.2360 free_r = 0.2596

-tried to model in the rest of the adp molecule...

    -bfactors refine to around 70, but there is a huge negative difference peak

    I don't think I can say where the rest of the molecule is...

    -bind_adp.pdb
    -REMARK final   r = 0.2375 free_r = 0.2607

-go back to diphophate model for now...

    -bind_dmso2.pdb
    -REMARK final   r = 0.2360 free_r = 0.2596

-looked over chain again, added in ALA(ASN)174, GLY175 to model

    -still need to check D35, Y73, E91, D134, M139, E247
    -reasonable stats and density
    -bind_check3.pdb
    -REMARK final   r = 0.2367 free_r = 0.2599

-maybe found that loop (174-5), modified a bit check4.pdb and checked residues:

    -ASP35 has 2 conformers
    -TYR73 is wack...near a symmetry axis...
    -MSE139 might be 2 conformers too
    -GLU247 needs to be checked once more...truncate to ALA?
    -also attempted to add in ALA(GLU)49, ALA(TYR)50 to the model

4/8/02

-remodeled ALA(GLU)49, ALA(TYR)50

-modelled ASP35 as two conformers

```
-removed ala(tyr)50 from model
-added some more waters to a total of 151
-minimize, bindividual,
     -bind_ww.pdb
     -REMARK final    r = 0.2325 free_r = 0.2577

4/12/02
-looked at ramachandran plots and remodelled problem areas...
-no residues in generously allowed or disallowed regions -=>
bind_r4.pdb
-REMARK final    r = 0.2340 free_r = 0.2584
-bind_r4.ps -=> procheck output

-inserted all side chains and residue names and truncated them
with generate
-fixed wandering AC2 35 mainchain
-rounds of min and bind with checking stats
-final.pdb
-final.ps final.sum 90% most favored/10% allowed regions
-model_stats_final.list R = 23.6 Rf = 26.2

4/17/02
-added the rest of the ADP molecule and refined
-still negative density on the base...set base b's to 99 and run
bindividual, bs refined down ~90
-acceptable amounts of negative density now, r's improved a bit
too.
-bind_testb.pdb
-REMARK final    r = 0.2356 free_r = 0.2606
-LUZZATI error = 0.31/0.36 angstroms

6/7/02
-changed GLU48 to ASP48 in pdb submission (revise_submit.pdb)
```

```
-changed GLU76 to ASP76 as well
>>>>>submit_model.pdb
```

Eric A. Galburt

Fred Hutchinson Cancer Research Center, University of Washington
1100 Fairview Avenue North, A3-023, Seattle, WA 98109
egalburt@u.washington.edu
(206) 667-4066

Education:

1997-2002	Doctor of Philosophy, Biochemistry Bio-Molecular Structure and Design Program University of Washington
1993-1997	Bachelor of Science, Physics Brown University

Publications:

Galburt, E.A. and Stoddard, B.L. (2002). The catalytic mechanisms of restriction and homing endonucleases. Biochemistry. *in review.*

Galburt, E.A., Pelletier, J., Wilson, G., Stoddard, B.L. (2002). Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. (2002). Structure. 10(9), *in press.*

Galburt, E.A. and Stoddard, B.L. (2001). Time-resolved macromolecular crystallography. Physics Today. 54(7), 33-39.

Galburt, E.A., Chadsey, M.S., Jurica, M.S., Chevalier, B.S., Erho, D., Tang, W., Monnat, R.J. Jr., and Stoddard, B.L. (2000). Conformational changes and cleavage by the homing endonuclease I-*Ppo*I: a critical role for a leucine residue in the active site. Journal of Molecular Biology. 300(4), 877-887.

Galburt, E.A. and Stoddard, B.L. (2000). Restriction endonucleases: one of these things is not like the others. Nature Structural Biology. 7(2), 89-91.

Galburt, E.A., Chevalier, B.S., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J. Jr., and Stoddard, B.L. (1999). Nature Structural Biology. 6(12), 1096-1099.