Title
The neutron structure of perdeuterated staphylococcal nuclease

Author
Gamble, Theresa R.

Publication Date
1995

Peer reviewed|Thesis/dissertation

The Neutron Structure of Perdeuterated Staphylococcal Nuclease

by

Theresa R. Gamble

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

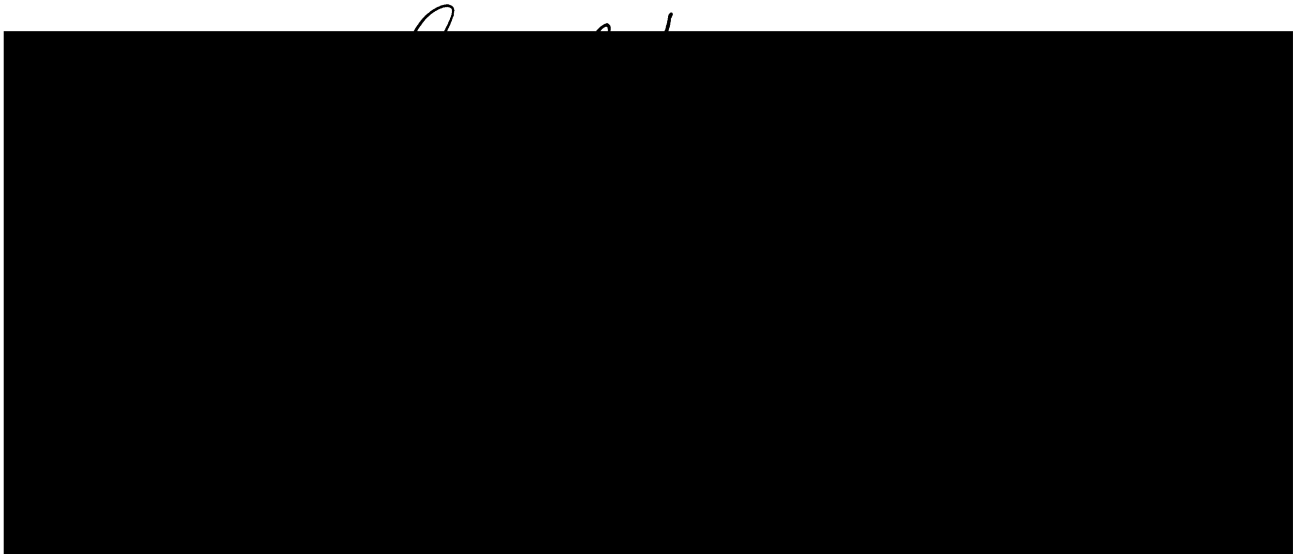
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Date **University Librarian**

Degree Conferred: . . . . . .

This thesis is dedicated to my parents

Eleanore and Andrew Gamble

and

to my friend and mentor

Abraham (Bart) de Vos

Preface

I wish to acknowledge my graduate advisor, Tony Kossiakoff, for his support throughout this long, and often arduous, project. I am particularly grateful for the opportunity he created for me in his laboratory. He has been visionary, cheerleader, idea generator, ambassador, and task master all of which were necessary to get the job done. In addition, I have learned much about leadership and scientific integrity from him. Fortunately for Tony, he only has to endure his "first graduate student" once.

Next, I wish to acknowledge and thank my "practical" advisor, Bart de Vos. It has been an honor and a privilege to work with such an amazing mind and stellar teacher. I am grateful for his patience, humor, diplomacy, and rationality. By example, he set a high standard, and always expected me to do the same. To the extent that Tony "let me in," Bart has "kept me in" and "pushed me out." From him, I have learned the true meaning of the words ""Godverdomme" and "hubris."

I thank my committee, Ken Dill, Jim Wells, and Bob Stroud, who could see the project's potential before I could. Rich Korszun, Dieter Schneider, and Jim Pflugrath made it possible to get all of the neutron data despite many obstacles.

Many people at Genentech have made my life as a graduate student much easier and a whole lot more fun. Mike Randal, Hans Christinger, Will Somers, Axel Scheidig, Karen Butcher, Laura Pellitier, Yves Muller, Celia Schiffer, Linda Chee, Mark Ultsch, Arlene Stuart, Brad Snedecor, Chuck Duarte, Bob McDowell, Ton Hynes, Tania Gonzalez, Len Presta, Andrew Braisted, Jana Ulmer, Mark Dennis, Melissa Starovasnik, Wayne Fairbrother, Nick Skelton, Shane Atwell, and Dave Wood. A special thanks goes to Charlie Eigenbrot for his dramatic flair and his suggestion to "teach" crystallography as a way to learn it. Also to Kevin Judice, for a positive collaboration experience, his cynical world view, and tips on how to function in a male-dominated society. And last, to Angie Namenuk, who taught me the joys and responsibilities of working with a "student."

The Graduate Group in Biophysics has afforded me many opportunities to grow into a broad-minded and thinking scientist. It is hard to straddle different fields and do it well, and the group in Biophysics is setting a good example. I especially thank Tak Kuntz, who encouraged me to approach Tony in the first place, and Julie Ransom, who has the thankless task of administration.

The Graduate Students' Association (GSA) and Women in Life Sciences (WILS) have allowed me to use my leadership skills, practice my politics, and feed my spirit. Thanks to Candy Clemens, Eric Koenig, and Ann O'Brien for helping me to hold it all together. Co-founding WILS has taught me much about networking and speaking with authority. My interactions with the women in WILS has given me a model for collaborative work that I intend to use throughout my career.

I thank the men in my life who have helped me through these past years: my brother David, Mark Anderson, Karl Clauser, and Greg Couch. I am particularly grateful for my Father's support - there's no feminist like a man with daughters.

And most importantly, I thank the women who have lent me their strength in ways they'll never fully know. Renée Williard, Tina Settineri, Donna Hendrix, Cindy Corwin, Karen Johnson, Fran Ravel, Thena Trygstad, Alma Sisco-Smith, Sherrie Hans, Linda Frederick, Angie Namenuk, Tracy Ware, Corrie Detweiler, Tia Renée Babin, Patty Lemley, my Mother, my sisters Sue, Kate, and Laura, Sheila Allen, Maria Longuemare, Heather Elmer, Margeret Lynch, Brenda Stine, Camille Claudel, and Simone de Beauvoir.

Epitaphs

"Neutrons are my life."

Theresa Gamble, second year graduate student

"Just gut it out."

Tony Kossiakoff

"What doesn't kill you, makes you stronger."

Theresa Gamble, seventh year graduate student, paraphrasing Friedrich Nietzsche

The Neutron Structure of Perdeuterated *Staphylcoccal* Nuclease

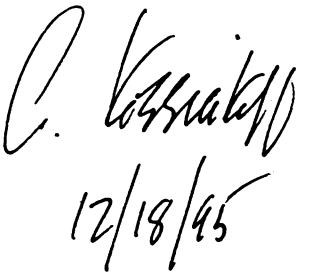
by

Theresa R. Gamble

Abstract

The structure of perdeuterated *Staphylococcal* nuclease (Snase) was determined by neutron crystallography. The protein was expressed in *E. coli* grown in a medium containing deuterated amino acids and D_2O and then purified. The mean perdeuteration level at non-labile hydrogen sites was found to be 96% by electrospray ionization mass spectrometry. The perdeuterated enzyme was crystallized and its X-ray structure determined. Comparison of the X-ray structures of native and perdeuterated Snase revealed no significant structural differences at 1.9Å resolution. Both native and perdeuterated crystals were soaked in deuterated mother liquor prior to neutron data collection to reduce background noise and to determine the hydrogen/deuterium (H/D) exchange pattern for crystallized Snase. Neutron diffracton data were collected at the H3A beamline at the High Flux Beam Reactor at Brookhaven National Laboratory from both native and perdeuterated crystals. Even in the absence of a diffracting crystal, high, variable background levels were found in the data, thus, no conclusion could be drawn as to the effect of protein perdeuteration on the quality of neutron data. A data set to 2.4Å resolution was collected from a 2.2mm^3 perdeuterated crystal. The data were processed using a modified version of the program MADNES. The final data set contains 5765 unique reflections corresponding to an overall completeness of 83% (64% complete between 2.5Å and 2.4Å resolution). XPLOR was used for crystallographic refinement

yielding a final R-value of 22.1%. The final model contains one protein molecule (residues 7-141), one calcium ion, one inhibitor molecule (pdTp), and 13 water molecules. There are 128 exchangeable backbone amide sites; 28 are occupied by hydrogen atoms, 64 by deuterium atoms, and 36 could not be determined. The combination of mid resolution data and phase bias affected structure analysis; neither hydroxyl rotor nor water orientation could be defined. Comparison of the H/D exchange pattern of Snase in the crystal to that in solution led to the conclusion that the effect of crystal contacts on H/D exchange, and thus flexibility, are both local and global.



12/18/95

List of Abbreviations

BNL	Brookhaven National Laboratory
BPTI	bovine pancreatic trypsin inhibitor
DNA	deoxyribonucleic acid
F_c	calculated structure factor amplitude
F_o	observed structure factor amplitude
H/D	hydrogen/deuterium
HFBR	High Flux Beam Reactor
MADNES	Munich Area Detector Non-Enraf System
MPD	methyl pentanediol
NMR	nuclear magnetic resonance
OB	oligonucleotide/oligosaccharide binding
pdb	protein data bank
pdTp	thymidine-3',5'-bisphosphate
psf	protein structure file
RMSD	root mean square deviation
RNA	ribonucleic acid
Snase	*Staphylococcal* nuclease

Table of Contents

Appendices

List of Tables

List of Figures

CHAPTER 1. Introduction

1.1 Objectives of Dissertation

This dissertation summarizes the work I have accomplished as a graduate student in Dr. Anthony Kossiakoff's laboratory in the Graduate Group in Biophysics at the University of California, San Francisco. The focus of my research has been to determine the neutron structure of perdeuterated *Staphylococcal* nuclease (Snase). All of the work involved in preparing perdeuterated Snase crystals for neutron data collection has been published and is included as an appendix to the thesis. In addition, collaborative work done on an active-site mutant of Snase containing a non-natural amino acid is also included as an appendix. The intention of this dissertation is to introduce the system and techniques used for the study, summarize the published results, and describe in detail all unpublished experiments and results.

1.2 Introduction to Neutron Diffraction

The experimental determination of hydrogen and deuterium atom positions is what distinguishes neutron from X-ray crystallography. Hydrogen atoms typically represent half of the atoms found in protein molecules, but are not identifiable in electron density maps. Although the majority of these atoms can be positioned stereochemically, there is a subset that cannot. It is this subset, made up of protons involved in protein mechanism, associated with charged residues or related to protein packing and flexibility, that are of biological interest and can be uniquely studied using neutron crystallography. This chapter will discuss the advantages and disadvantages of the method, explain why protein perdeuteration can enhance the technique, and describe the significance of previously determined neutron structures.

1.3 The Advantages of Neutron Crystallography

Neutron diffraction has two distinct and significant attributes in comparison with X-ray crystallography: 1) both hydrogen and deuterium atoms can be identified in neutron Fourier maps and, 2) hydrogen/deuterium exchange experiments are possible because hydrogen and deuterium scatter with opposite sign [1, 2]. Although protein crystals scatter both X-rays and neutrons, the resulting diffraction patterns are generated by very different means. X-rays are scattered by electrons by virtue of their charge which interacts with the incident radiation; this leads to a parallel between an atom's scattering capacity and its atomic number [3]. Hence, elements with few electrons are not easily identified in X-ray Fourier maps, except at extremely high resolution (>1Å) not usually attainable from protein crystals. In contrast, neutrons are scattered by atomic nuclei, causing neutron scattering amplitudes to depend on the nuclear properties of each element and its isotopes. The neutron scattering lengths for most protein atoms are similar, ranging from 2.8 x 10^{-13} cm for sulpher to 9.4 x 10^{-13} cm for nitrogen. Carbon and deuterium atoms have virtually identical scattering lengths (6.6 x 10^{-13} cm and 6.7 x 10^{-13} cm, respectively), allowing both elements to be equally identifiable in neutron Fourier maps. However, several elements, including hydrogen, cause a phase shift of 180° in the diffracted wave; such atoms have negative scattering lengths and correspond to negative peaks in neutron Fourier maps ($2F_O$-F_C). Despite its lower amplitude (-3.8 x 10^{-13} cm), the negative scattering length of hydrogen makes it distinguishable from deuterium [2, 4].

Along with the power to identify hydrogen and deuterium atoms, neutron diffraction has two other distinct advantages [5]. Because the scattering amplitude for nitrogen is ~40% higher than carbon, oxygen, or deuterium, the absolute orientation of residues such as histidine, glutamine and asparagine can be determined. In addition, exposure of protein crystals to X-rays causes crystal decay due to the formation of free

radicals; thermal neutrons are non-ionizing and do not cause such damage, allowing an entire neutron data set to be collected from a single crystal.

1.4 Technical Difficulties

Despite the advantages of neutron diffraction, lack of reliable neutron sources, low neutron flux, the difficulty of growing large crystals, and high incoherent background make common use of the technique prohibitive. Currently, the High Flux Beam Reactor at Brookhaven National Laboratory is the only nuclear reactor with the experimental set-up necessary for high-resolution single crystal protein crystallography. Due to the low intensity of thermal neutrons as a function of wavelength and the loss of neutrons through monochromation, the maximum obtainable flux at 1.5Å is ~10^3 times smaller than X-ray beams at equivalent wavelengths [3]. Long exposure times are necessary to collect neutron data causing experiments to last anywhere from a few months to a year. The most ironic drawback of neutron diffraction comes from the incoherent scattering of hydrogen atoms. Every element scatters neutrons both coherently and incoherently; coherent scatter adds to the diffraction pattern, while incoherent scattering only contributes to the noise in the data [5]. The cross-section for the coherent scattering of hydrogen is ~40 times lower than its incoherent cross-section, resulting in a inherently poor signal-to-noise ratio in neutron data [5]. One way to reduce this high background scatter is to soak the crystals in deuterated mother liquor before data collection; however, only the labile protons of the system (protein and solvent) will be replaced by deuterons and the remaining non-labile protons will continue to generate noise. To counter both low flux and poor signal-to-noise, it is necessary to use large protein crystals. Crystals ranging in size from 1.6mm^3 (trypsin) to 30 mm^3 (ribonuclease) have been used for neutron data collection [4], but even using pulsing and seeding techniques, such large crystals are extremely difficult to grow.

1.5 Protein Perdeuteration

To overcome some of the problems mentioned above, the protein under study can be completely perdeuterated. Protein perdeuteration goes beyond simple crystal soaking and replaces every non-labile hydrogen atom with deuterium. This can be accomplished by expressing the protein in bacteria grown on medium containing deuterated amino acids. Since incoherent scattering by deuterium atoms is negligible, protein perdeuteration should significantly reduce the amount of noise in the diffraction data and may reduce the minimum crystal size necessary for neutron work.

1.6 Historical Perspective

The first attempts at neutron crystallography were carried out in the late 1960s in an effort to locate hydrogen atoms in myoglobin [6] and to prove the feasibility of this new structure determination technique. The motivation for this approach came from two sources, the first was the lack of positional data for hydrogen atoms in proteins solved by X-ray diffraction and the second from the fact that the structures of small molecules had been successfully solved using neutron crystallography [7]. Over the next two decades, six unique protein structures, in addition to three variants of myoglobin, were solved using neutron crystallography (Table 1.1).

This body of research can be roughly divided into two categories, functional and structural. In several cases, the purpose of the work was to search for active-site protons and gain insight into substrate binding or enzyme mechanism. For example, neutron density maps (F_o-F_c and $2F_o$-F_c) were used to identify the protonation state of an active site histidine in oxymyoglobin [8, 9]; the presence of a deuterium atom bound at the NE2 rather than the ND1 position of the imidazole ring allows a hydrogen bond to form with the bound O_2 molecule. This hydrogen bond could contribute several kcal per mole to

Table 1.1: Protein structures determined by neutron crystallography

Myoglobin (~1981)	
Acid-metmyoglobin	Raghaven & Schoenborn [10]
Oxymyoglobin	Phillips [8]
Carbonmonoxymyoglobin	Hannon & Schoenborn [11]
Trypsin(1981)	Kossiakoff and Spencer [12]
Ribonuclease A (1983)	Wlodawer & Sjolin [13]
Lysozyme (1984)	Mason, Bentley & McIntyre [14]
Crambin (1984)	Teeter and Kossiakoff [15]
BPTI (1984) (Bovine Pancreatic Trypsin Inhibitor)	Wlodawer, Huber & Sjölin [16]
Subtilisin (1991)	Kossiakoff, Ultsch, White & Eigenbrot [17]

the enthalpy of oxygen binding and is speculated to affect oxygen affinity in both myoglobin and hemoglobin. In similar fashion, the protonation state of the histidine in the catalytic triad of the serine protease trypsin was unambiguously determined by neutron diffraction [12]. This family of enzymes was known to hydrolyze peptide bonds through general base-catalyzed nucleophilic attack on the carbonyl carbon of the substrate, but it was not known whether histidine or aspartic acid functioned as the general base in the reaction. By calculating difference density with alternative isotope occupation, the neutron structure of trypsin was used to identify the histidine as the general base of the reaction.

Neutron diffraction has also been used to study dynamic properties, packing characteristics, and hydration of proteins. The combination of neutron crystallography and hydrogen exchange has been used to correlate exchange patterns with structural and chemical properties using trypsin [18], subtilisin [17], bovine pancreatic trypsin inhibitor (BPTI) [16], crambin [15], acid metmyoglobin [10], oxymyoglobin [8, 9], ribonuclease [13], and lysozyme [19]. Hydrogen exchange patterns are used to assess the degree of flexibility throughout a protein, with rigid regions being more protected, and flexible regions being more highly exchanged. While hydrogen bonding itself does not guarantee protection from exchange, most protected amide protons participate in the hydrogen bonding schemes of alpha helices and beta sheets. In all of the studies mentioned above, a strong correlation was observed between protection from exchange and secondary structure. Even within the same types of secondary structure, interesting differences in exchange patterns have been seen. For example, in the anti-parallel beta sheet structure of trypsin, the most protected section of sheet is found in the central strands [18]; in contrast, subtilisin has its protected amide protons concentrated at one edge of parallel beta sheet [17]. This implies that different dynamic characteristics exist for these two types of beta sheet, which is likely to affect the overall stability of the protein. Likewise, different exchange patterns for alpha helices have been documented. Many proteins

(ribonuclease, myoglobin, trypsin) contain alpha helices that display amide proton protection on one side but not the other [10, 13, 18]. In these cases, protection has been attributed to intramolecular contact on the less accessible surface of the helices. However, in crambin, which is a small hydrophobic protein, full protection is seen for almost every amide proton in its alpha helices [15]. This difference may indicate the high rigidity of a tightly packed hydrophobic molecule.

Internal packing constraints can be examined by neutron diffraction by identifying the position of hydrogen and deuterium atoms affiliated with molecular rotors. Methyl rotors have been examined to determine whether cooperative packing effects in the folded protein can perturb a rotor from a low-energy "staggered" to a high-energy "eclipsed" conformation. Using the neutron structure of crambin, Kossiakoff and Shteyn found that the majority of methyl rotors assume the "staggered" conformation, suggesting that the inherent flexibility of the protein allows for packing accommodation even in the tightly packed protein interior [20]. In a similar fashion, hydroxyl rotors have been used to assess the relative importance of steric and electrostatic effects in the core of a protein. The correlation between possible hydrogen bond formation and actual rotor position was studied using the neutron structure of trypsin. In general, a complementarity was found between low-energy hydroxyl orientation and the formation of hydrogen bonds [21], demonstrating that neither steric nor electrostactic forces dominate internal packing.

Analysis of protein-solvent interactions have been studied using neutron data; questions concerning the primary and secondary protein hydration layer can be addressed by examining the orientation and hydrogen bonding networks of water molecules. Of particular interest has been the interaction of between polar and apolar residues and solvent molecules. Neutron D_2O-H_2O solvent difference maps were calculated for trypsin using two data sets, one collected from a crystal soaked in deuterated mother liquor, the other in hydrogenated mother liquor. These maps demonstrate that the large majority of the protein surface has direct contact with water molecules, but that the apolar

atoms of hydrophobic regions are circumvented by water molecules that are part of a larger water network [22]. Similar water analysis has been done using the protein crambin [23].

1.7 References

1. Kossiakoff, A.A., *The use of neutrons to show how proteins work,* in *Biotechnology and the human genome: Innovations and Impact,* A.D. Woodhead, B.J. Barnhart, and K. Vivirito, Editor. 1988, Plenum Press: New York. p. 63-78.

2. Kossiakoff, A.A., *The application of neutron crystallography to the study of dynamic and hydration properties of proteins.* Ann. Rev. Biochem., 1985. **54**: p. 1195-1227.

3. Blundell, T.L. and L.N. Johnson, *Protein Crystallography.* 1976, San Francisco: Academic Press.

4. Wlodawer, A., *Neutron diffraction of crystalline proteins.* Prog. Biophys. Molec. Biol., 1982. **40**: p. 115-159.

5. Bacon, G.E., *Neutron Diffraction.* Third ed. 1975, Oxford: Clarendon Press. 636.

6. Schoenborn, B.P., *Neutron diffraction analysis of myoglobin.* Nature, 1969. **224**: p. 143-146.

7. Moore, F.M., B.T.M. Willis, and D.C. Hodgkin, *Structure of a monocarboxylic acid derivative of vitamin B12.* Nature, 1967. **214**: p. 129-133.

8. Philips, S.E.V. and B.P. Schoenborn, *Neutron diffraction reveals oxygen-histidine hydrogen bond in oxymyoglobin.* Nature, 1981. **292**: p. 81-82.

9. Philips, S.E.V., *Hydrogen bonding and exchange in oxymyoglobin,* in *Neutrons in Biology,* B.P. Schoenborn, Editor. 1984, Plenum Press: New York. p. 305-322.

10. Raghaven, N.V. and B.P. Schoenborn, *The structure of bound water and refinement of acid metmyoglobin,* in *Neutrons in Biology,* B.P. Schoenborn, Editor. 1984, Plenum Press: New York. p. 247-259.

11. Hanson, J.C. and B.P. Schoenborn, *Real space refinement of neutron diffraction data from sperm whale carbonmonoxymyoglobin.* J. Mol. Biol., 1981. **153**: p. 117-146.

12. Kossiakoff, A.A. and S.A. Spencer, *Neutron diffraction identifies His57 as the catalytic base in trypsin.* Nature, 1980. **288**: p. 414-416.

13. Wlodawer, A. and L. Sjölin, *Hydrogen exchange in RNase A: Neutron diffraction study.* Proc. Natl. Acad. Sci. USA, 1982. **79**: p. 1418-1422.

14. Mason, S.A., G.A. Bentley, and G.J. McIntyre, *Deuterium exchange in lysozyme at 1.4Å resolution,* in *Neutrons in Biology,* B.P. Schoenborn, Editor. 1984, Plenum Press: New York. p. 323-334.

15. Teeter, M.M. and A.A. Kossiakoff, *The neutron structure of the hydrophobic plant protein crambin,* in *Neutrons in Biology,* B.P. Schoenborn, Editor. 1984, Plenum Press: New York. p. 335-348.

16. Wlodawer, A., *et al., Structure of bovine pancreatic trypsin inhibitor: Results of joint neutron and X-ray refinement of crystal form II.* J. Mol. Biol., 1984. **180**: p. 301-329.

17. Kossiakoff, A.A., *et al., Neutron structure of subtilisin BPN': Effects of chemical environment on hydrogen-bonding geometries and the pattern of hydrogen-deuterium exchange in secondary structure elements.* Biochemistry, 1991. **30**: p. 1211-1221.

18. Kossiakoff, A.A., *Protein dynamics investigated by the neutron diffraction-hydrogen exchange technique.* Nature, 1982. **296**: p. 713-721.

19. Bentley, G.A., *et al., Exchange of individual hydrogens for a protein in a crystal and in solution.* J. Mol. Biol., 1983. **170**: p. 243-247.

20. Kossiakoff, A.A. and S. Shteyn, *Effect of protein packing structure on side-chain methyl rotor conformations.* Nature, 1984. **311**: p. 582-583.

21. Kossiakoff, A.A., J. Shpungin, and M.D. Sintchak, *Hyroxyl hydrogen conformations in trypsin determined by the neutron diffraction solvent difference map method: Relative importance of steric and electrostatic factors in defining hydrogen-bond geometries.* Proc. Natl. Acad. Sci. USA, 1990. **87**: p. 4468-4472.

22. Kossiakoff, A.A., *et al., Analysis of solvent structure in proteins using neutron D2O-H2O solvent maps: Pattern of primary and secondary hydration of trypsin.* Proteins: Structure, Function, and Genetics, 1992. **12**: p. 223-236.

23. Teeter, M., *Water-protein interactions: Theory and experiments.* Annu. Rev. Biophys. Biophys. Chem., 1991. **20**: p. 577-600.

CHAPTER 2. *Staphylococcal* Nuclease as a Model System

Staphylococcal nuclease was chosen for this work for a variety of reasons, both biological and technical. Most important is the fact that a wealth of biochemical and biophysical information is known about the protein, creating a rich environment to interpret the protein's neutron structure. The protein's structure has been determined by both X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, along with its hydrogen exchange pattern in solution. Comparison of the neutron structure and exchange data with these previous studies can be used to analyze the effect of crystal packing on protein flexibility. A mechanism for Snase has been proposed based on the X-ray crystal structure of the inhibited protein; however, recent work has brought this proposal into contention. The ability to determine the protonation state of active site residues and water molecule orientation may shed further insight on the enzyme's activity. Technically, the fact that Snase can be expressed at high levels, crystallized with ease, and its X-ray structure used for phase information, make it a good candidate for neutron crystallography.

2.1 Biochemical and Biophysical Characteristics

Snase is a calcium dependent 5'-phosphodiesterase that cleaves both ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) yielding 3'-nucleotides as products [1-3]. The enzyme was discovered in 1956 [4] in cultures of *Staphylococcus aureus*; shortly thereafter its catalytic properties as a nuclease were revealed and it was used in now classical work on determination of nearest-neighbor frequencies in DNA before the era of sequencing [5]. It is a relatively small protein of 149 amino acids with a calculated molecular weight of 16,810 and pI of 9.6. It contains no disulfide bonds or free sulfhydryl groups [1-3]. The enzyme is strongly inhibited by deoxythymidine-3',5'-

bisphosphate (pdTp) in the presence of calcium [2]. The protein can be unfolded by low pH (< 3.5), high temperature (>55°C), and high concentrations of urea or guanidinium chloride; regardless of the mode of denaturation, it is completely reversible [5].

All of these characteristics have contributed to its appeal as a model system. Over the past 40 years Snase has been used to address general biophysical questions including the folding and denaturation pathways of proteins, the effect of individual and cooperative mutations on protein stability, secondary structure characterization, as well as the deduction of enzyme mechanism from three dimensional structure. Much of the work on folding pathways and characterization of the denatured state has been done by Shortle and co-workers; Snase is ideal for this type of work because it is highly soluble in both its native and denatured state, the transition between the folded and unfolded state is reversible and it displays a resistance to irreversible inactivation [6]. In addition, a 131 residue fragment of the wild-type sequence is a prototype for the denatured state because it has little residual secondary structure even in the absence of denaturants [7]. The cloning and expression of Snase in *Escherichia coli* [8] led to hundreds of mutants of the protein; insertion, substitution, and deletion mutants have been used to study the structural and energetic characteristics of the protein in both its folded and unfolded states [9-15], and most, if not all, residues in the active site have been mutated in an effort to correlate structure and function [16-21]. Structural features such as the alpha aneurism [16-22], the oligonucleotide/oligosaccharide-binding (OB) fold [23], and strained loop conformation [24] have been characterized in Snase.

2.2 Structures of *Staphylococcal* Nuclease

Both X-ray crystallography and NMR have been used to characterize the three-dimensional structure of Snase in its ligated and unligated forms [25-29]. The protein consists of a five-stranded antiparallel Greek key β barrel that packs against three approximately parallel α helices, categorizing Snase as a typical member of the α + β class of proteins [30] (Fig. 2.1). The active site is enclosed by one edge of the β barrel and a long extended loop. Only one calcium ion has been identified in the crystal structure of inhibited Snase despite its known requirement for two equivalents of Ca^{2+} for catalytic activity [25, 27].

The reported RMS deviation between the apo and inhibited X-ray structures is 0.24Å between backbone atoms, indicating that no large scale domain or secondary structure movement occur upon substrate binding [28]. Changes in the backbone are confined to a flexible loop centered on residue 50 and a short region which borders the active site. Both the N and C-termini are assumed to be disordered because density is only observed for residues 7 to 141 in the inhibited and 6 to 141 in the apo structures. Several side chains (apo = 8, inhibited = 11), most being lysine, lack sufficient density for confident placement. The majority of differences between apo and inhibited structures are seen in the active site side chains. 82 water molecules were built into the ligated (1.65Å resolution), and 85 into the unligated (1.7Å resolution), forms of the protein.

NMR was used to both determine the structure and gather hydrogen exchange data for the protein [26, 29]. The earlier work (Torchia, 1989) was done on the inhibited structure because a comparison of the crystal and solution conformations was desired; later (Loh, 1993), NMR data were used to calculate lifetimes and protection factors for the amide protons of both the ligated and unligated structures. Torchia et al. found the polypeptide backbone to be essentially the same in solution and crystalline states, and confirmed a high level of flexibility for residues 42-49, a region that has high B-factors in

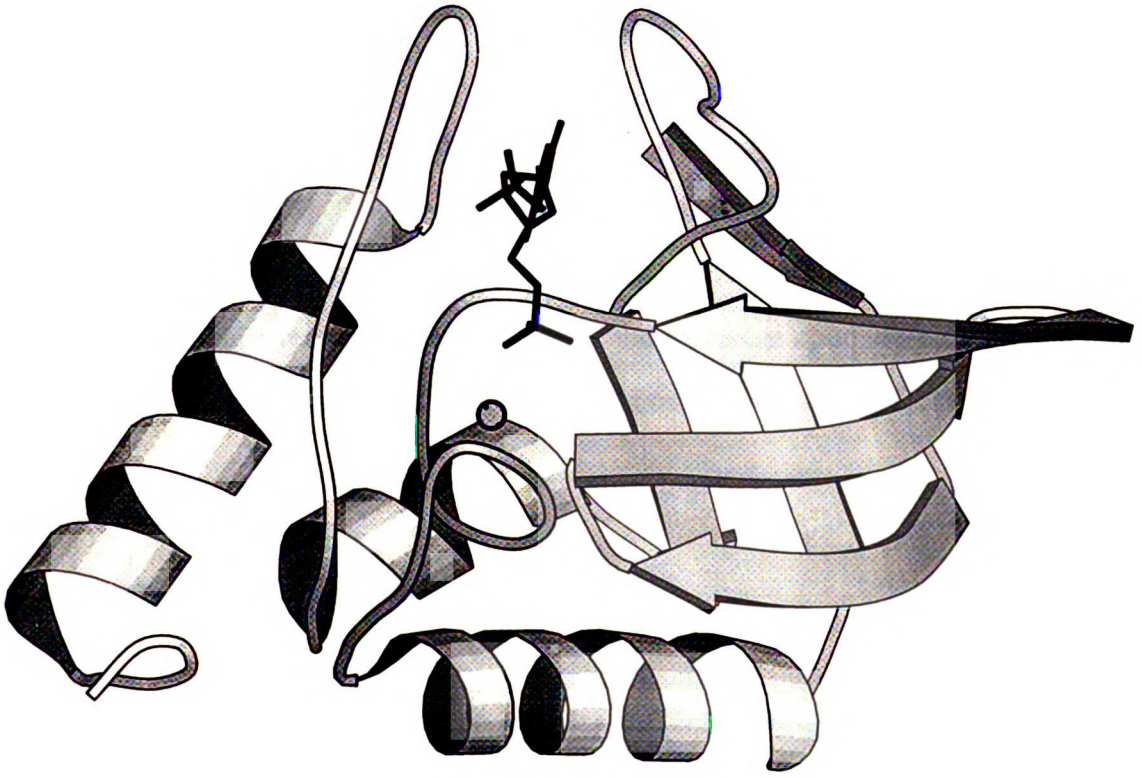


Fig. 2.1: Ribbon diagram of inhibited Snase. Coordinates used from the Protein Data Bank (1SNC).

the crystal structure [29]. 24 hours after dissolving Snase in D_2O (pH 6.5, 37°C), 45 amide protons remained unexchanged with the solvent. All of these protons form hydrogen bonds in the protein, however, the authors note that not all hydrogen-bonded amide protons are protected in this manner. For example, the exchange pattern significantly differs between helices; one is almost fully protected (helix 99-109), another is half protected (helix 58-69), and the third has only three protected amide sites (helix 122-134). These data demonstrate that although the helices are observed to be well-formed structures in both crystal and solution, they have different stabilities against hydrogen exchange.

The effect of calcium and inhibitor binding on hydrogen exchange was studied in the work of Loh et al [26]. The ligated protein has an increased number of protected amide protons, which are concentrated in the C-terminal half of the protein. Most striking is the global effect on protection factors and exchange upon inhibitor binding; the protection factors have increased for all protected sites regardless of their interaction with the substrate.

2.3 Mechanism of *Staphylococcal* Nuclease

2.3.1 Description of Enzyme Activity and Inhibition

As briefly mentioned in section 2.1, Snase catalyzes the hydrolysis of both DNA and RNA at the 5' postiion of the phosphodiester bond yielding a free 5'-hydroxyl group and a 3'-phosphate monoester [1-3, 5]. Two ions of Ca^{2+} are required for activity with all substrates and cannot be replaced by other ions; however, a number of other ions promote the binding of various inhibitors. This type of metal dependence suggests that Ca^{2+} is essential for both binding and cleavage in the active enzyme. The enzyme also displays a preference ($dA \geq dT >> dC >> dG$) for the type of base in the β-nucleoside

position in a series of dinucleotides ($dN^{\alpha}pdN^{\beta}$). The most potent Snase inhibitor known is pdTp (thymidine 3',5'-bisphosphate), which requires Ca^{2+} for inhibition. The enzyme-pdTp-Ca^{2+} complex is believed to closely resemble the actual enzyme-substrate-Ca^{2+} for two reasons: 1) if the 5' phosphate of the inhibitor is extended via a *p*-nitrophenyl ester, it becomes a good substrate for the enzyme, and 2) inhibition requires calcium rather than any other ion. It should be noted, however, that there is a discrepancy between the number of calcium ions required for activity (2) and the number found in the active site of the inhibited crystal structure (1), which may hinder correct mechanistic interpretation of the structural data. Water molecules are critically involved in the proposed mechanism of the enzyme, so it is important to note that there may be differences between the hydration of the inhibitor and the substrate due to the negative charge of the 5' phosphate (single negative charge for substrate, double negative charge for pdTp) [27]. The active site is depicted in Fig. 2.2.

2.3.2 Proposed Mechanism

In 1979, Cotton et al. proposed a mechanism to explain Snase's activity based primarily on its inhibited structure [27] and two assumptions based on the pH range for enzyme activity and diester containing substrate: 1) a direct nucleophilic attack of phosphorus forms a five-coordinate, trigonal bipyramidal transition state that breaks down into a new phosphate species, and 2) attacking and leaving groups only have access to the intermediate state via the apical positions of the bypyramid. The X-ray structure reveals multiple interactions between the protein, water molecules, the inhibitor and a single ion of Ca^{2+} in the active site. The calcium ion has three protein ligands (oxygen atoms from the carboxylate groups of Asp21 and Asp40, and the carbonyl oxygen of Thr41), one inhibitor ligand (an oxygen atom of the 5' phosphate), and three water ligands, giving it one more ligand than expected octahedral coordination. Arg35 and

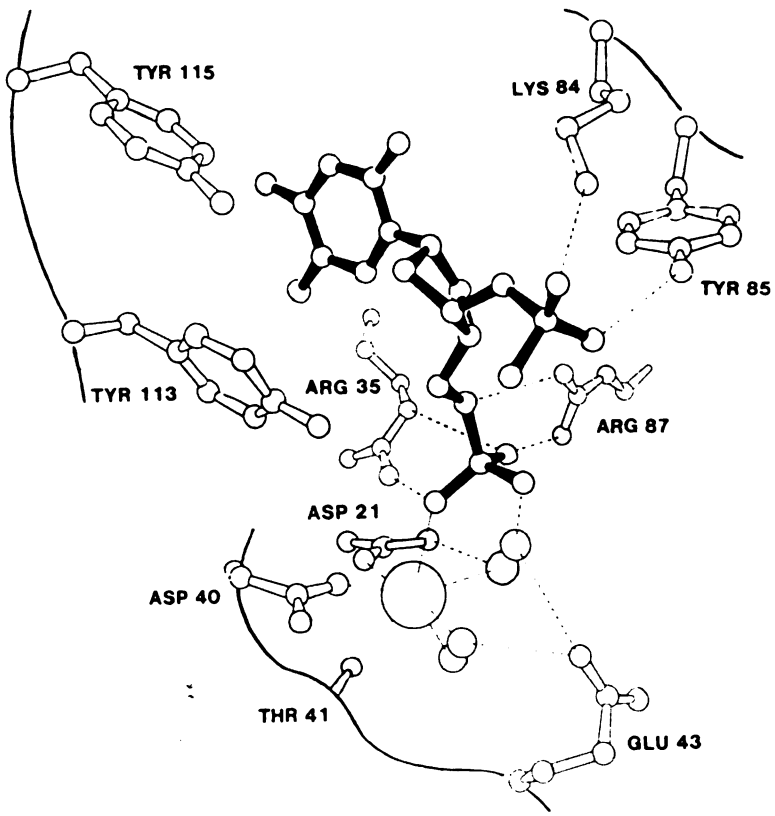


Fig. 2.2: Schematic drawing of the active site of Snase. Used with permission [25].

Arg87 form a total of four hydrogen bonds to the 5' phosphate group. A water molecule bridges the carboxylate group of Glu43 and one of the 5'-phosphoryloxygen atoms of the inhibitor.

Cotton and co-workers speculate that cleavage of the 5'--P--O bond is caused by nucleophilic attack at the phosphorus atom by a hydroxide ion in line with the 5'--C--O--P ester bond. Due to its position in the active site, the attacking nucleophile is postulated to be the bridging water molecule described above. In this model, Glu43 not only positions this water molecule but also acts as a general base, abstracting a proton from it. Cotton et al. suggest that the function of the calcium ion in the active site may be to neutralize the negative charge of the phosphate group, making the phosphorus atom more susceptible to nucleophilic attack. Efficient cleavage of the 5'--P--O bond requires protonation of the oxygen atom; although it is chemically unlikely at pH ≈ 8 for arginines to perform such a function, based on its position in the active site, Cotton et al. propose that Arg87 acts as a general acid. They also suggest that alternatively, Tyr113 or water molecules may play such a role [27] .

2.3.3 Conflicting Evidence

Over the years, the proposal of Cotton et al. has been the generally accepted mechanism for the enzyme; however, this proposed mechanism has now been called into question for several reasons. Mutations of Glu43 and other active site residues, as well as new information about the wild-type structure, have been used to argue both for and against the Cotton model. For example, when the structure was re-determined by Loll and Lattman [25], a second candidate for the attacking nucleophile was identified. A water molecule, which hydrogen bonds to an oxygen atom of the carboxylate group of Glu43 and acts as a calcium ligand, is correctly positioned and close enough to the 5' phosphorus atom to function as an alternative attacking nucleophile.

In an attempt to verify the proposed roles of active site residues, Glu43 has been replaced by aspartate, glutamine, asparagine, alanine, and serine [21, 31, 32], while mutations from arginine to lysine have been made at positions 35 and 87 [20]. Unfortunately, these substitutions have led to both decreases in activity and changes in protein structure, making it impossible to attribute the decrease in activity to changes in residue functionality or to simple displacement.

To get around this obstacle, Judice and coworkers incorporated non-natural amino acids at position 43 to test individual chemical properties of glutamic acid [33]. For example, a nitro analog of the residue, which is both isoelectronic and isosteric to the carboxylate group, was made to examine the role of Glu43 as a general base, while homoglutamate (glutamate extended by one methylene) was incorporated to assess the steric role of Glu43. This work, including the structure of the homoglutamate mutant, led to speculation that Glu43 does not function as a base but rather acts to stabilize the flexible loop found in the active site. (This work has been published and is included as Appendix A.)

This alternative role for Glu43 is supported by the work of Hale and co-workers; these researchers have used a combination of mutagenesis and enzyme kinetics to negate the hypothesis that Glu43 acts as a general base [34]. If Glu43 functions as a general base, then the dependence of log k_{cat} on pH should be described by an ascending line that reaches a limiting, pH -independent plateau value above the pK_a of Glu43. Although the actual pK_a value for Glu43 in the active enzyme is not known, the presence of a Ca^{2+} ion in the active site would be expected to decrease, not increase, the pK_a value from its unperturbed value (glutamic acid $pK_a = 4.3$). Measurements of k_{cat} were done for wild-type, an E43D mutant, and a mutant in which a large portion of the active site Ω–loop (residues 44-49) had been removed. In all cases, k_{cat} never reaches a plateau even at pH values greater than 10, leading to the conclusion that general base catalysis by Glu43 is unimportant [34]. Although compelling, this interpretation is questioned by Libson et al.

in their discussion of a Asp21 to Glu mutant structure, in which Glu43 is seen to be a direct ligand to the calcium ion. They argue that the decrease in activity of this mutant (1500-fold decrease in V_{max} from wild-type) is similar to that of mutants of Glu43, pointing to the direct involvement of this residue in the mechanism [35].

2.3.4 Role of Hydrogen Atoms in the Active Site

The use of neutron crystallography to identify hydrogen atoms has the potential to illuminate two aspects of the active site, the protonation state of the active site residues and the orientation of the water molecules. Even if such information will not determine the mechanism of Snase once and for all, it can still be used to re-evaluate the multiple conflicting theories, and hopefully lend credance to some and disclaim others. For example, there are two possibilities for the attacking nucleophile; the orientation of these two water molecules may distinguish one from the other as the more likely candidate. In addition, hydroxide ions may be differentiated from water molecules, if such are present in the time-averaged active site. Determination of the protonation state of all residues in the active site would be helpful in deciding whether particular residues can function as general acids or bases. More generally, the protonation states may reveal the extent to which local protein environment effect the pK_as of individual amino acids.

2.4 References

1. Tucker, P.W., E.E. Hazen, and F.A. Cotton, *Staphylococcal nuclease reviewed: A prototypic study in contemporary enzymology. I. Isolation; physical and enzymatic protperties.* Molecular and Cellular Biochemistry, 1978. **22**(2-3): p. 67-77.

2. Tucker, P.W., E.E. Hazen Jr., and F.A. Cotton, *Staphylococcal nuclease reviewed: A prototypic study in contemporary enzymology. II. Solution studies of the nucleotide binding site and the effects of nucleotide binding.* Molecular and Cellular Biochemistry, 1979. **23**(1): p. 3 - 16.

3. Tucker, P.W., E.E. Hazen, and F.A. Cotton, *Staphylococcal nuclease reviewed: A prototypic study in contemporary enzymology.* Molecular and Cellular Biochemistry, 1979. **23**(3): p. 131 - 141.

4. Cunningham, L., B.W. Catlin, and M.P. de Garilhe, *A deoxyribonuclease of Micrococcus pyogenes.* Journal of the American Chemical Society, 1956. **78**: p. 4642-4645.

5. Anfinsen, C.B., P. Cuatrecases, and H. Taniuchi, *Staphylococcal nuclease, chemical properties and catalysis,* in *The Enzymes,* P.D. Boyer, Editor. 1971, Academic Press: New York. p. 177-204.

6. Shortle, D., *Staphylococcal nuclease: a showcase of m-value effects.* Advances in Protein Chemistry, 1995. **46**: p. 217-247.

7. Alexandrescu, A.T. and D. Shortle, *Backbone Dynamics of a Highly Disordered 131 Residue Fragment of Staphylococcal Nuclease.* J. Mol. Biol., 1994. **242**: p. 527-546.

8. Shortle, D., *A genetic system for analysis of Staphylcoccal nuclease.* Gene, 1983. **22**: p. 181-189.

9. Sondek, J. and D. Shortle, *Accommodation of single amino acid insertions by the native state of Staphylococal nuclease.* Proteins: Structure, Function, and Genetics, 1990. **7**: p. 299-305.

10. Sondek, J. and D. Shortle, *Structural and energetic differences between insertions and substitutions in Staphylococcal nuclease.* Proteins: Structure, Function, and Genetics, 1992. **13**: p. 132-140.

11. Keefe, L.J., *et al., Accommodation of insertion mutations on the surface and in the interior of staphylococcal nuclease.* Protein Science, 1994. **3**: p. 391-401.

12. Green, S.M., A.K. Meeker, and D. Shortle, *Contributions of the polar, uncharged amino acids to the stability of Staphylococcal nuclease: Evidence for mutational effects on the free energy of the denatured state.* Biochemistry, 1992. **31**: p. 5717-5728.

13. Green, S.M. and D. Shortle, *Patterns of nonadditivity between pairs of stability mutations in staphylococcal nuclease.* Biochemistry, 1993. **32**: p. 10131-10139.

14. Flanagan, J.M., *et al.*, *Mutations can cause large changes in the conformation of a denatured protein.* Biochemistry, 1993. **32**: p. 10359-10370.

15. Miyazawa, S. and R.L. Jernigan, *Protein stability for single substitution mutants and the extent of local compactness in the denatured state.* Protein Engineering, 1994. **7**(10): p. 1209-1220.

16. Weber, D., J., A.K. Meeker, and A.S. Mildvan, *Interactions of the acid and base catalysts on Staphylococcal nuclease as studied in a double mutant.* Biochemistry, 1991. **30**: p. 6103-6114.

17. Weber, D.J., *et al.*, *Diverse interactions between the individual mutations in a double mutant at the active site of Staphylococcal nuclease.* Biochemistry, 1990. **29**: p. 8632-8642.

18. Serpersu, E.H., *et al.*, *Kinetic and magnetic resonance studies of the glutamate-43 to serine mutant of Staph ylococcal nuclease.* Biochemistry, 1989. **28**: p. 1539-1548.

19. Grissom, C.B. and J.L. Markley, *Staphylococcal nuclease active-site amino acids: pH dependence of tyrosines and arginines by 13C NMR and correlation with kinetic studies.* Biochemistry, 1989. **28**: p. 2116-2124.

20. Pourmotabbed, T., *et al.*, *Kinetic and conformational effects of lysine substitutions for arginines 35 and 87 in the active site of Staphylococcal nuclease.* Biochemistry, 1990. **29**: p. 3677-3683.

21. Loll, P.J. and E.E. Lattman, *Active site mutant Glu-43 -- Asp in Staphylococcal nuclease displays nonlocal structural changes.* Biochemistry, 1990. **29**: p. 6866-6873.

22. Keefe, L.J., *et al.*, *The alpha aneurism: a structural motif revealed in an insertion mutant.* Proc. Natl. Acad. Sci. USA, 1993. **90**: p. 3275-3279.

23. Alexandrescu, A.T., *et al.*, *NMR structure of a stable "OB-fold" sub-domain isolated from Staphylcoccal nuclease.* J. Mol. Biol., 1995. **250**: p. 134-143.

24. Hodel, A., R. Kautz A., and R.O. Fox, *Stabilization of a strained protein loop conformation through protein engineering.* Protein Science, 1995. **4**: p. 484-495.

25. Loll, P.J. and E.E. Lattman, *The crystal structure of the ternary complex of Staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65Å.* Proteins: Structure, Function, and Genetics, 1989. **5**: p. 183-201.

26. Loh, S.N., *et al.*, *Hydrogen exchange in unligated and ligated Staphylococcal nuclease.* Biochemistry, 1993. **32**: p. 11022-11028.

27. Cotton, F.A., E.E. Hazen Jr., and M.J. Legg, *Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme-thymidine 3',5'-bisphosphate-calcium ion complex at 1.5Å resolution.* Proc. Natl. Acad. Sci. U.S.A., 1979. **76**: p. 2551-2555.

28. Hynes, T.R. and R.O. Fox, *The crystal structure of Staphylococcal nuclease refined at 1.7Å resolution.* Proteins: structure, function, and genetics, 1991. **10**: p. 92-105.

29. Torchia, D.A., S.W. Sparks, and A. Bax, *Staphylococcal nuclease: sequential assignments and solution structure.* Biochemistry, 1989. **28**: p. 5509-5524.

30. Orengo, C.A. and J.M. Thornton, *Alpha plus beta folds revisited: some favoured motifs.* Structure, 1993. **1**(2): p. 105-120.

31. Hibler, D.W., *et al.*, *Site-directed mutants of Staphylcoccal nuclease. Detection and localization by 1H NMR spectroscopy of conformation changes accompanying substitutions for glutamic acid-43.* Biochemistry, 1987. **26**: p. 6278-6286.

32. Wilde, J.A., *et al.*, *Identification of residues involved in a conformation change accompanying substitutions for glutamate-43 in Staphylococcal nuclease.* Biochemistry, 1988. **27**: p. 4127-4132.

33. Judice, J.K., *et al.*, *Probing the mechanism of Staphylococcal nuclease with unnatural amino acids: Kinetic and structural studies.* Science, 1993. **261**: p. 1578 - 1581.

34. Hale, S.P., L.B. Poole, and J.A. Gerlt, *Mechanism of the reaction catalyzed by Staphylococcal nuclease: Identification of the rate-determining step.* Biochemistry, 1993. **32**: p. 7479-7487.

35. Libson, A.M., A.G. Gittis, and E.E. Lattman, *Crystal structures of the binary Ca2+ and pdTp complexes and the ternary complex of the asp21 to glu mutant of Staphylococcal nuclease. Implications for catalysis and ligand binding.* Biochemistry, 1994. **33**: p. 8007-8016.

CHAPTER 3. Crystal Preparation

This chapter briefly summarizes the preparation, crystallization and X-ray structure determination of perdeuterated *Staphylococcal* nuclease. All of this work has been published and is included in Appendix B [1]. The end of the chapter includes a description of the crystal soaking procedure performed prior to neutron data collection.

3.1 The Production and X-ray Structure Determination of Perdeuterated *Staphylococcal* Nuclease

Using *E. coli* stain AR120 containing a pAS1 plasmid, Snase was expressed on rich, deuterated media at levels equivalent to wild-type expression (80 mg/L). The purification method was based on previous Snase protocol that was furthered developed by Shortle and Meeker [2]. The mean perdeuteration level of the non-exchangeable sites in the protein was found to be 96% by electrospray ionization mass spectrometry. This result approaches the upper limit obtainable for perdeuteration based on the 97-98% deuterium incorporation in the amino acids of the media.

In order to demonstrate that perdeuteration itself did not conformationally alter Snase, the X-ray structure of the perdeuterated protein was determined and compared to the native conformation. The same basic protocol was used to crystallize native and perdeuterated *Staphylococca*l nuclease in the presence of inhibitor (thymidine-3',5'-bisphosphate) and calcium; conditions were optimized to grow large crystals. Both native and perdeuterated crystals grew in the same space group ($P4_1$) and had identical unit cell dimensions (48.4Å x 48.4Å x 63.4Å). X-ray data were collected on both crystal types to 1.9Å resolution and each data set had completeness greater than 95%. The R_{merge} for the two orientations of each crystal, total number of observations, and total number of reflections indicated that each crystal type behaved similarly during data

collection and processing. The two data sets were very similar as evidenced by an R_{scale} between them of 4%. In addition, both crystal types scatter anisotropically and had comparable anisotropic scale factors.

XPLOR was used to refine the two structures by using rigid-body, conjugate gradient, and individual temperature factor refinement. The final R factor for both structures, including water, the inhibitor, and calcium, was 19%. The most significant peaks found in a difference map comparing native and perdeuterated data were only 0.11 $e/Å^3$, indicating that perdeuteration did not significantly alter the structure or diffraction characteristics of the protein. Comparison of the final native and perdeuterated backbone atoms (Cα, C, N) along with the calcium atoms and the inhibitor yielded an RMS deviation of 0.08Å. The largest concerted differences between backbone atoms were found in a flexible loop that previous investigators have found to be partially disordered. Between side chain atoms, the largest differences were found at lysine sites located at the surface of the molecule. All but five of the water molecules had corresponding mates in the native and perdeuterated structures; no water position with a temperature factor less than 30Å differed from its mate by more than 0.3Å. The conclusion of this work was that perdeuteration did not cause significant conformational changes in the protein.

3.2 Neutron Crystal Preparation

All crystals used for neutron diffraction were soaked in deuterated mother liquor prior to data collection. The deuterated mother liquor consisted of 10.5 mM potassium phosphate (pH 8.15), 0.6 mM $CaCl_2$, 1.2 mM citric acid, 0.4 mM pdTp, D_20 and 60% deuterated MPD (2-methyl-2,4-pentane-d12-diol, 99.0 atom %D, Isotec Inc, Miamisburg, OH). The crystals were transferred from their original sitting drops to drops containing increasing amounts of precipitant starting with 30% MPD and reaching 60% MPD in four steps. The crystals were carefully monitored throughout the procedure to prevent

excessive cracking; despite caution, some crystals, especially the largest ones, suffered surface damage.

The benefits of this procedure are two fold: 1) incoherent scattering of the crystal and thus noise in the data are reduced and 2) hydrogen exchange information can be obtained from the final neutron structure. Both perdeuterated and native protein were purified and crystallized in hydrogentated solvents. During purification, the protein is completely denatured, thus, every labile proton site in the crystallized protein will be occupied by an hydrogen atom. Soaking in deuterated mother liquor will replace most exchangeable protons with deuterons; however, a small subset of amide protons will be protected from exchange by intramolecular hydrogen bonds and crystal contacts. The resulting hydrogen exchange pattern can then be interpreted in terms of protein structure and crystal environment.

3.3 References

1. Gamble, T.R., K.R. Clauser, and A.A. Kossiakoff, *The production and X-ray structure determination of perdeuterated Staphylococcal nuclease.* Biophysical Chemistry, 1994. **53**: p. 15-26.

2. Shortle, D. and A.K. Meeker, *Residual structure in large fragments of Staphylococcal nuclease: effects of amino acid substitutions.* Biochemistry, 1989. **28**: p. 936-44.

CHAPTER 4. Data Collection

All neutron data for this project were collected at the High Flux Beam Reactor (HFBR) at Brookhaven National Laboratory (BNL) (Upton, New York). The crystallography station (H3A) there is a single crystal diffractometer coupled with a position-sensitive detector. This chapter will describe the physical parameters of H3A and how they were optimized for data collection. Data collection was designed with two distinct goals in mind; one was to collect comparable data on both native and perdeuterated Snase crystals in order to determine the effect of perdeuteration on signal-to-noise ratios, the other was to collect a complete neutron data set of a perdeuterated Snase crystal in order to solve its neutron structure. A description of the experimental strategy and a summary of all collected data will be presented.

4.1 The HFBR and H3A Beamline

The HFBR is a steady-state fission reactor that uses highly enriched Uranium-235 fuel and a heavy water (D_2O) moderator to sustain a controlled nuclear reaction [1]. The HFBR was built to meet an increasing demand for high flux research reactors and specifically designed to maximize the number of neutrons available at each beamline. It became operational in 1965, and eventually provided a peak thermal flux of 1.05×10^{15} neutrons/cm^2-sec at a power of 60 megawatts [2]. Due to perceived safety issues the reactor is now operated at half power (30 megawatts), effectively reducing the flux by half (peak thermal flux = 0.5×10^{15} neutrons/cm^2-sec). Normally, the reactor is operated for 24 hours per day for 24 days; at the end of each cycle, the reactor is shutdown for five to seven days for refueling and routine maintenance. There are nine experimental beamlines at the HFBR that have been developed to cover a wide range of scientific research: nuclear physics is studied at the on-line isotope separator facility,

powder diffraction and reflectometry are available for chemistry research, and facilities for small and large angle scattering studies are provided for biological work [2].

The H3A beamline (Fig. 4.1) provides high quality thermal neutrons that are uncontaminated by fast neutrons and gamma rays [3]. It is coupled to a protein crystallographic station featuring a two-dimensional, position-sensitive detector and a four-circle goniostat for crystal orientation (Fig. 4.2). A converging beam tube "focuses" neutrons onto a single zinc crystal monochromator whose 002 reflection is used to select a wavelength (λ) of 1.61Å. At this wavelength, the flux at the sample is ~1.0 x 10^6 neutrons/cm^2-sec at a reactor power of 30 megawatts. The beam size at the sample is 6 mm in diameter with a 20' divergence. The beam is collimated with circular cadmium apertures [1]. The 2Θ swing arm of the goniostat holds a tower, which supports the detector [4]. The physical parameters of the station (e.g., detector position), as well as the experimental parameters (ω rotation, sample exposure, etc.), are controlled by the program XTAL2D on a VAX 4200 computer using the VMS operating system [1].

4.1.1 Physical Parameters and Detector Characteristics

Although the essential equipment necessary for neutron data collection is equivalent to that used for X-ray data collection (beam, goniometer, detector), the scale is much larger. With the exception of the Xuong-Hamlin multiwire area detector, in an X-ray setup, the distance between the crystal and the detector is typically 35 - 300 mm. At the H3A beamline, however, a range of 600 - 1000 mm for the crystal to detector distance is possible, designed to achieve optimal spot separation for data from large crystals. The protein crystal is aligned in a quartz capillary tube and mounted onto a goniometer. Quartz capillaries are used because of their low absorbance and scattering of neutrons [5]. The crystal is rotated in discrete steps around its ϕ axis, which is the axis perpendicular to

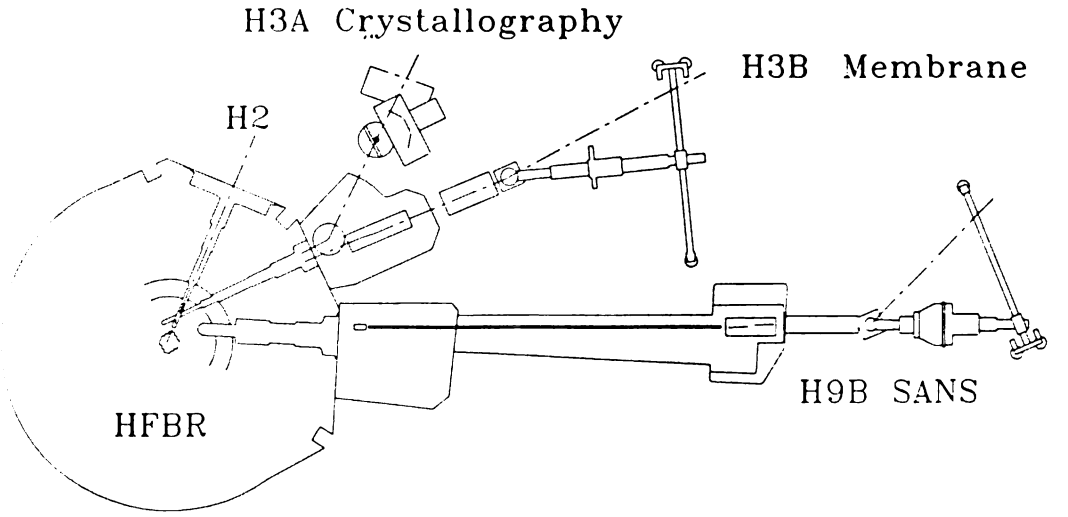


Fig. 4.1: The biological beamlines at HFBR (H3A, H3B and H9B). The reactor core is depicted on the left of the figure. It is surrounded by several meters of biological shielding composed of concrete, steel, and lead. Neutrons are generated in the core and transported to the individual beamlines via neutron guides. The H3 beamline extracts neutrons from the core tangentially in order to minimize contamination with fast (epithermal) neutrons and gamma rays.

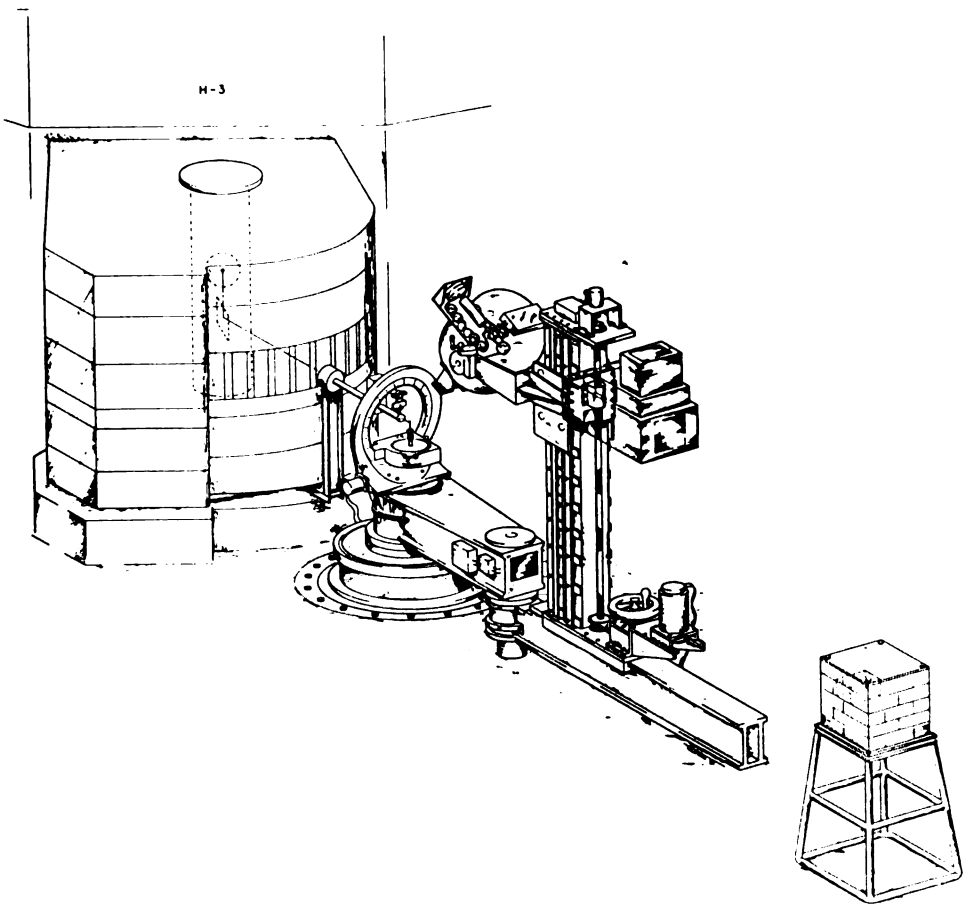


Fig. 4.2: Schematic drawing of the H3A protein crystallography station.

the beam. The detector has four degrees of freedom: the crystal-to-detector distance (DIST), the tower height (ZZ), a horizontal 2Θ swing arm (2Θ), and a detector tilt angle (T).

The detector used at H3A is a He^3 filled position-sensitive detector built in-house at BNL with a counter resolution of 1.5mm [4]. The detector face itself is 20 by 18 cm and divided up into 256 x 128 pixels (FAST area detector = 6.4 x 4.8 mm, 512 x 512 pixels; Mar image plate = 300 mm in diameter, 2000 pixels). In practice the limited size and resolution of the detector severally restricts the solid angle and necessitates a set of different detector positions for a complete data set. Thermal neutrons (0.001 - 10 eV, ≅ 9 - 0.09 Å) are only weakly absorbed by most materials making them difficult to detect directly. Although the interaction between a neutron and an atom's nucleus produces a change in the energy or momentum of the neutron, the low incident energy of a thermal neutron results in very little energy transfer during the process. A secondary reaction must be initiated in order to create radiation with sufficient energy to be detected directly. In the case of the H3A detector, the secondary reaction (Equation 4.1) gives rise to the production of protons , which in turn cause secondary ionization events that can be easily detected [6]. The counting efficiency of the detector is ~80% at a wavelength of 1.6Å [4]. In addition, a helium filled snout is attached to the detector face designed to reduce background air scatter in the data. The snout is adjustable and can be moved closer to or further from the crystal.

Equation 4.1 $$^{3}He + {}^{1}n \Longrightarrow {}^{3}H + {}^{1}p + 0.77\ MeV$$

4.1.2 Data Frames

A frame of data consists of an array of 128 x 256 pixel values corresponding to the number of neutrons that have impinged on the detector face at each pixel position during a given interval (Fig. 4.3). Frames of data were collected as a function of crystal exposure time; this measurement of the number of neutrons to hit the detector is referred to as the monitor count. In general, low resolution data (infinity to 5Å) were collected using a monitor count of 8000, mid resolution data (5Å to 3Å) at a monitor count of 16,000, and high resolution data (3Å to 2.2Å) at a monitor count of 48,000. These monitor counts correlate roughly to collection times of 3 minutes/frame, 6 minutes/frame, and 18 minutes/frame respectively (Table 4.1). These raw data are stored in binary format and can be read directly into the MADNES (Munich Area Detector Non-Enraf System) [7] program for processing.

Once MADNES was successfully installed (this occurred after several data sets had already been collected), it was discovered that several rows and columns of pixels on the detector surface were nonfunctional (Fig. 4.4). Because this was not known when planning the data collection strategy (see below), it was not originally taken into consideration. However, when it became clear that a percentage of data was not being collected due to this problem, the detector settings were adjusted accordingly.

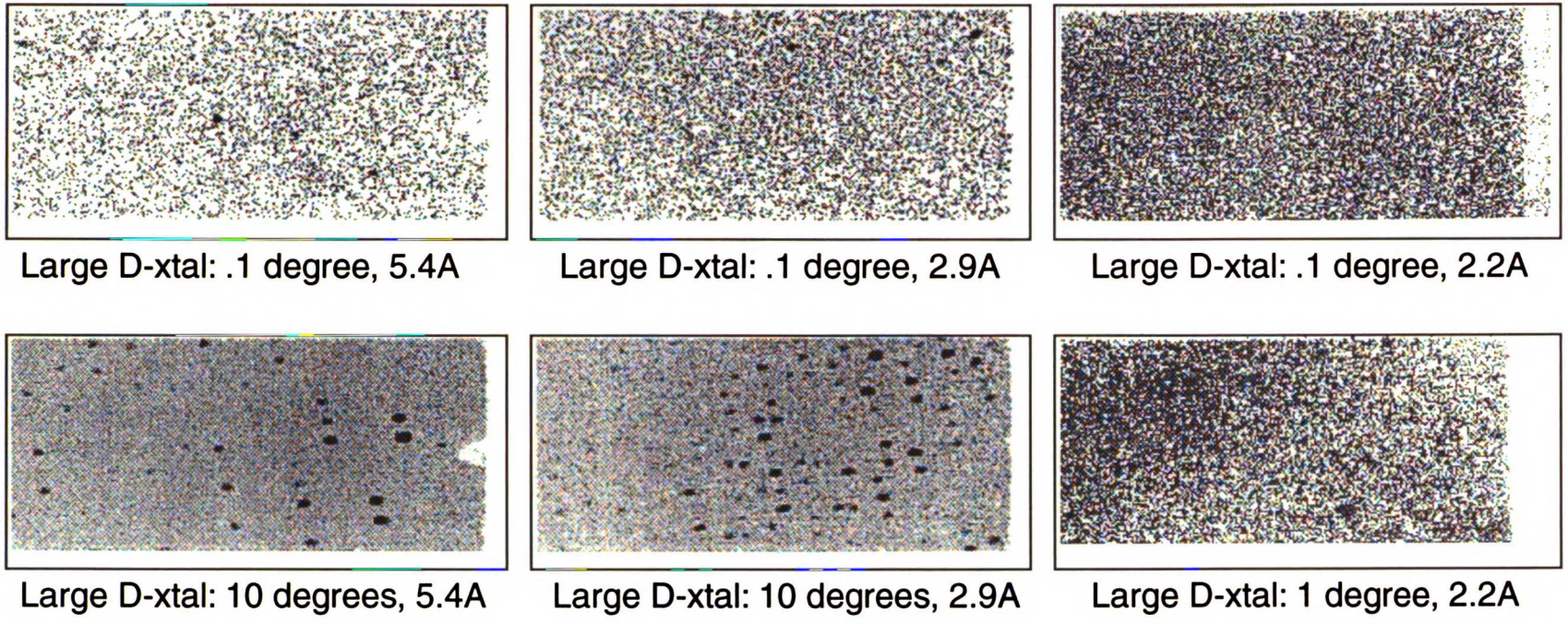


Fig. 4.3: Examples of raw data frames at low (infinity to 5.4Å), mid (5.4 to 2.9Å), and high (2.9Å to 2.2Å) resolution data.

Table 4.1: Approximate Data Collection Time

Monitor Counts	Minute/Data frame	Time to collect one data set (hrs/days)
8000	3	50/2
16000	6	100/4.2
20000	7.5	125/5.2
24000	9	150/6.25
32000	12	200/8.3

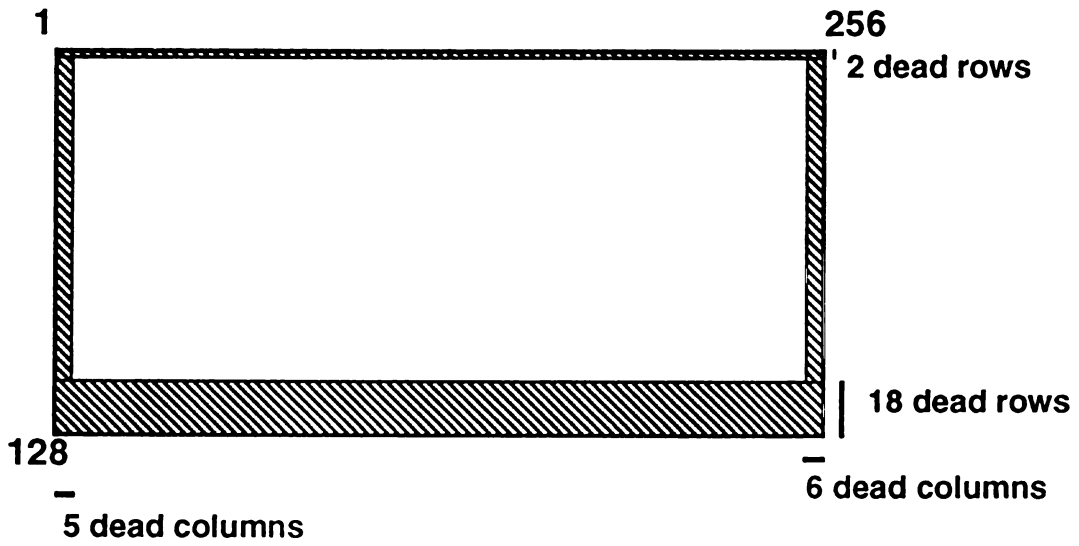


Fig. 4.4: Diagram mapping out dead pixels on the detector face.

4.2 Collecting Comparable Data from both Native and Perdeuterated Crystals

4.2.1 Experimental Strategy

The hypothesis put forth in this work is that protein perdeuteration will increase the signal-to-noise ratio by reducing background due to incoherent scattering by the protein. To ascertain the effect of protein perdeuteration on background scatter and signal-to-noise for the data, comparable data on both native and perdeuterated Snase crystals were collected. To achieve this it was critical that all parameters be held as constant as possible for the two data sets. In the ideal case, data should be collected under identical conditions using native and perdeuterated crystals of identical size. As is the nature of crystal growth, such ideal conditions don't exist, thus, data were collected from native and perdeuterated crystals of various sizes. In each case, the crystal was placed in a quartz capillary and mounted onto the goniometer. The crystal was then rotated in discrete steps of 0.1° in ϕ with the detector at a low resolution setting. In addition, it was necessary to assess the ambient background scatter. This was done by collecting data frames from a quartz capillary containing only mother liquor (no crystal), a dry quartz capillary tube, and recording frames with nothing in the neutron beam at all.

4.2.2 Summary of Data

Table 4.2 summarizes comparable data collected on native and perdeuterated crystals. As shown in the table, data were collected from four crystals: two native crystals (6.1 mm^3 and 1.2 mm^3) and two perdeuterated crystals (2.2 mm^3 and 0.5 mm^3). In most cases the tower height (ZZ) and the tilt angle (T) remained at zero, the 2Θ swing angle (2Θ) set to -8.0°, and the neutron counting rate constant (monitor counts = 8000). In two exceptions, the 2Θ swing angle (2Θ) and monitor counts were altered to determine

Table 4.2: Data Collected to Study the Effects of Protein Perdeuteration

Files	Crystal (cubic mm)	Monitor Counts	2Θ (degrees)	Colimator (cm)	Notes
NATIVE CRYSTALS					
HSN8_1_001.2a;2-101	1.2	8000	-8	0.1	*
HSNHEL_001.2a;2-101	1.2	8000	-8	0.1	*, same as HSN8
					with helium in snout
SMALLH_2_001.2a;2-788	1.2	8000	-8	0.2	*
HSNASE_001.2a;1-998	6.1	8000	-8	0.3	*
HSN38T_001.2a;1-130	6.1	24000	-38	?	*
HSN45_001.2a;1-231	6.1	24000	-45	?	*
PERDEUTERATED CRYSTALS					
DSN8T_001.2a;1-62	2.2	8000	-8	0.3	*
DSN8_1_001.2a;1-100	2.2	8000	-8	0.1	*
DSN8_2_001.2a;2-999	2.2	8000	-8	0.2	*, helium in snout
SMALLD_001.2a;2-336	0.5	8000	-8	0.2	*, helium in snout
					* ZZ = 0, TA = 0

the highest diffraction resolution for the large native crystal. A range (0.1 to 0.3 cm) of cadmium apertures were used to collimate the beam; this was done in an attempt to match beam size with crystal dimensions. One experiment performed on a native crystal was designed to test the effect of filling the detector snout with helium on background (HSN8 vs. HSNHEL). In addition, experiments were done to measure background scattering; these data are shown in Table 4.3. In all background measurements, the monitor counts were held constant at 8000 and the 0.1 cm cadmium aperture was consistently used. The data collected with no capillary in the beam (NCP) were taken to measure background scattering when no direct beam falls on the detector; for the case of $2\Theta = -8.0$, the beam falls on the very edge of the detector and is attenuated by the beamstop.

4.3 Collecting a Complete Neutron Data Set for a Perdeuterated Crystal

4.3.1 Experimental Strategy

Designing an optimal strategy for collecting a complete neutron data set requires considering several additional parameters in comparison to a typical X-ray data collection. The parameters under consideration for X-ray data collection include crystal-to-detector distance, the 2Θ swing angle, as well as crystal orientation and rotation variables. The issues addressed by these parameters, spot (reflection) separation, resolution, and efficiency (i.e. how to best "capture" a complete set of unique reflections), are identical in neutron data collection. However, the puzzle is made more complicated by the following factors. Neutron flux is low (10^6 vs. 10^9 counts/cm^2-s) compared to those obtained in typical X-ray experiments and is compensated for by long exposure times and crystal size. In addition, the detector is relatively small in relation to all of reciprocal space at a given resolution, and can only "capture" a small percentage of the

Table 4.3: Data Collected to Study Background Scattering

Files	Crystal (cubic mm)	Monitor Counts	2Θ (degrees)	Collimator (cm)
WETCAP_001.2a;1-100	none	8000	-8	0.1
DRYCAP_001.2a;2-101	none	8000	-8	0.1
NOCAP_001.2a;2-101	none	8000	-8	0.1
NCP_11_001.2a;1-58	none	8000	-11	0.1
NCP2_11_001.2a;1-32	none	8000	-11	0.1
NCP3_11_001.2a;1-10	none	8000	-11	0.1
Note: In all cases, ZZ = 0, TA = 0				

total reflections in any given position. These two factors must be carefully balanced when devising a strategy to effectively collect a complete neutron data set.

The reasons that cause such a limited amount of reciprocal space coverage by any one detector positions include large beam size, large crystal size, (neutron crystals > 2 mm^3 , X-ray crystals > 0.1 mm^3), and the resolution of the detector. This necessitates using multiple, overlapping detector positions. In theory, the goniostat allows for unlimited crystal orientation requiring a small number of detector positions. However, this strategy leads to a much larger number of total data sets needed to cover reciprocal space (roughly twice as many) than a strategy in which the crystal remains at a fixed rotation axis and the detector has multiple positions in three-dimensional reciprocal space. Both a horizontal and vertical 2Θ swing angle must be computed to orient the detector; however, the physical set up of the H3A station forces the vertical 2Θ swing arm to be broken down into two components, the height of the detector on its supporting tower (ZZ) and the detector tilt angle (T) (see section 4.3.1.1). The fact that improvement in signal-to-noise increases as the square root of exposure time must be taken into consideration when choosing the time to collect for each data frame.

To account for all of the parameters described above, eight overlapping detector positions were calculated to collect a complete neutron data set for Snase to 2.2Å (see section 4.3.1.1, Fig. 4.5). If Snase is aligned with the four fold (c* axis) along the axis of rotation (space group = $P4_1$), only a 90° rotation of the crystal is necessary to capture unique reflections at any given detector position. For all data collection, the crystal-to-detector distance was fixed at 623 mm; however, this parameter was not optimized for the crystal cell dimensions ($P4_1$, a = 48.0Å, c = 63.5Å). A shorter distance would have been more desirable allowing more reflections to be collected per detector position. As described in section 4.1.2, exposure times were chosen to find an acceptable compromise between signal-to-noise and minimal data collection time. Under ideal conditions, where

data are collected 24 hours a day, with no mishaps or down time, a complete Snase neutron data set would take approximately 2 months to collect.

4.3.1.1 Calculating Detector Position

Currently there is no program at the H3A station that will automatically calculate the optimal placement of multiply detector positions . However, this can roughly be estimated graphically using Bragg's law. Fig. 4.5 represents a volume of reciprocal space; the blue curves show increasing spheres of resolution. The detector, which is 20 x 18 cm, is approximated with a 4 x 4 unit square. The red curve represents the blind region, or cusp, which is not collectable with a vertical rotation axis. The green shading at the bottom of each detector square depicts the dead region of the detector. It is essential that each detector position overlap with at least one other one so that the data can be merged together.

Once the graph is drawn, each detector position can be calculated with respect to crystal-to-detector distance (DIST), tower height (ZZ), horizontal 2Θ swing angle (2Θ), and detector tilt angle (T). The horizontal 2Θ swing angle (2Θ) can be calculated using Bragg's law.

The solid angle (SA), which subtends the center of the detector and its edge, is 9.12°. From the diagram (Fig. 4.5), the resolution at the horizontal edge of each detector position is determined; the 2Θ swing angle is then calculated (Equation 4.1, Fig. 4.6a)

Equation 4.1: $$2\Theta = 2\sin^{-1}[n\lambda/2d]$$

The tilt angle (T) is used to approximate a vertical swing angle by tilting the detector face tangential to the sample at a given tower height (ZZ). The angle is

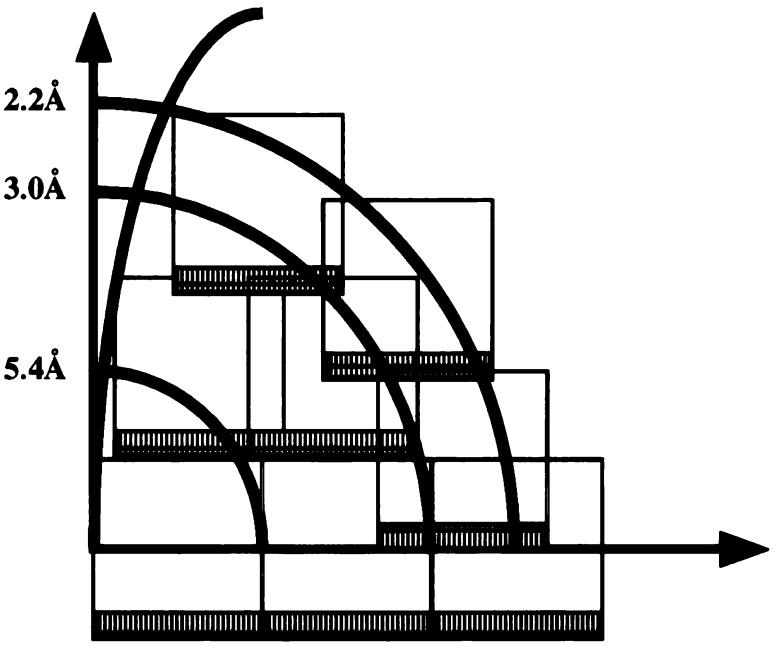
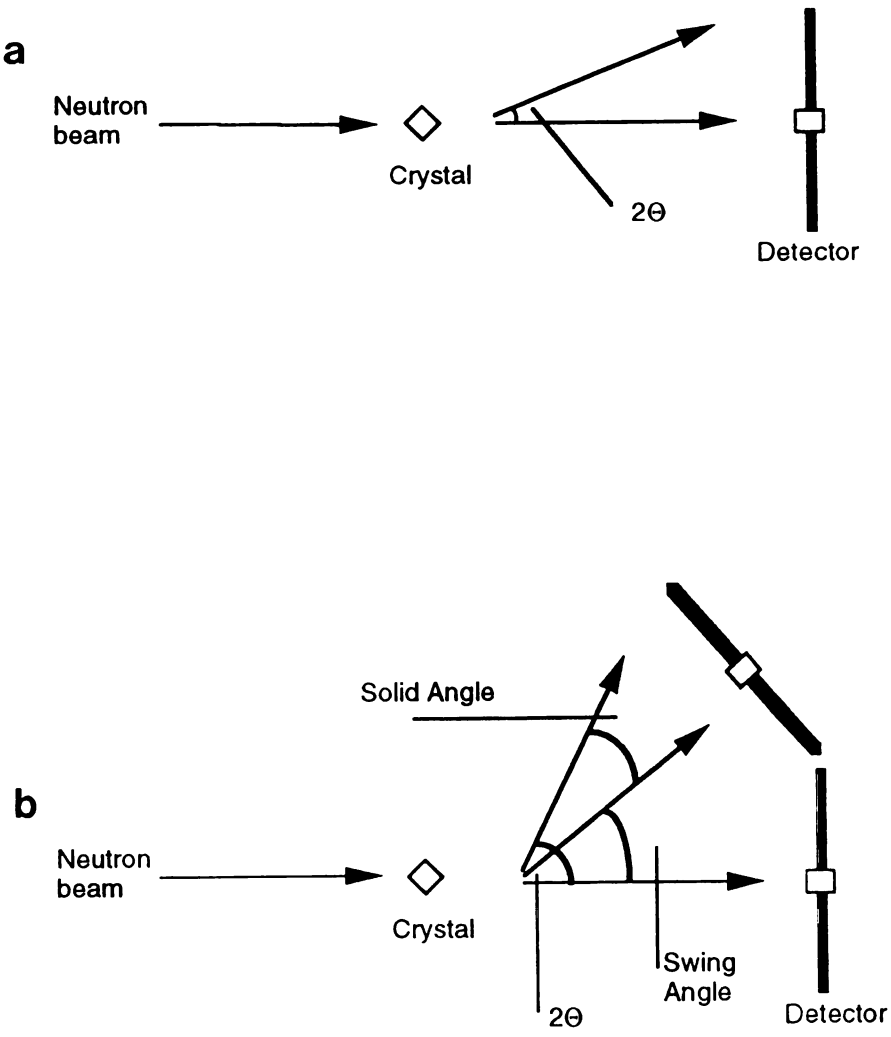


Fig. 4.5: Overlapping detector positions optimized to cover reciprocal space



4.6: Diagram depicting (a) horizontal swing angle and (b) vertical tilt angle calculation

calculated for a given crystal-to-detector distance and tower height (Equation 4.2, Fig. 4.6b).

Equation 4.2: $T = \tan^{-1}[ZZ/DIST]$

The tower height for each detector position is estimated from the diagram (Fig. 4.5). Table 4.4 displays the parameters for the original detector positions calculated to collect a complete neutron data set to 2.2Å resolution.

4.3.1.2 Correcting for Crystal Misalignment

The original calculations for detector positions were dependent on aligning the crystal with its c* axis parallel to the axis of rotation. Only with this alignment and these detector positions, calculated to capture only one reflection of each Friedel pair, will solely unique data be amassed. Collecting a maximum of unique reflections per setting will minimize the amount of time required for data collection. If the crystal is tilted off the c* axis, some number of duplicate symmetry related reflections will be collected. Although the crystal was aligned optically along its c* axis when it was placed in the quartz capillary, it slipped off axis sometime between collecting data for the first and second detector position. (See details in Chapter 5, Data Processing.) This was not recognized until all the originally calculated detector settings had been collected. In combination with optimal crystal rotation, the rest of the neutron data was collected using three new detector settings.

Table 4.4: Data Collected for Neutron Structure of Perdeuterated Snase

Files	Monitor Counts	2Θ (degrees)	ZZ (cm)	T (degrees)	Phi rotation (degrees)
DSN8_2_001.2a;2-999	8000	-8.00	0.00	0.00	80 -- 179.8
DSN23_2_001.2a;2-999	16000	-23.00	0.00	0.00	80 -- 179.8
DSN38_2_001.2a;2-167	20000	-38.00	0.00	0.00	80 -- 96.6
DSN38_3_001.2a;2-712	20000	-38.00	0.00	0.00	96 -- 167.1
DSN25_2_001.2a;1-973	16000	-8.24	25.00	21.90	80 -- 177.2
DSN10_2_001.2a;2-999	48000	-33.53	10.00	9.12	80 -- 179.8
DSN275_2_001.2a;2-657	48000	-29.82	27.50	23.82	80 -- 145.6
DSN275_5_001.2a;2-400	48000	-29.82	27.50	23.82	140 -- 179.9
DSN375_001.2a;2-782	48000	-15.85	37.50	31.04	80 -- 158.1
DSN375_2_001.2a;2-222	48000	-15.85	37.50	31.04	158 -- 180.11
DSN20_2_001.2a;2-851	16000	-18.09	20.00	17.79	80 -- 165
**DSN8_001.2a;2-200	16000	-18.09	20.00	17.79	160 -- 179.9

All data collected using a single perdeuterated crystal (2.2 cubic mm) and a cadmium collimator of 0.2 cm diameter.

**Clearly an error in name convention; however it is listed here to be consistant with the data tapes.

4.3.2 Summary of Data

The data collected for the eight initial detector positions is listed in Table 4.4. In all cases a single perdeuterated crystal (2.2 mm^3) and a 0.2 cm cadmium collimator were used. In four of the data sets, data were collected in two segments (DSN38, DSN275, DSN375, DSN20). A variety of reasons caused collection of these data sets to be truncated, all of which can be attributed to either equipment failure at the station (hardware or software) or premature termination of the reactor cycle. Table 4.5 displays the parameters for the detector positions calculated to compensate for the crystal misalignment. The reason that six data sets were eventually collected was because the ϕ rotation was not taken into account when correcting for the misalignment; the first three data sets contained a significant percentage, but not all of the missing data.

Table 4.5: Data Collected to Compensate for Crystal Misalignment

Files	Monitor Counts	2Θ (degrees)	ZZ (cm)	T (degrees)	Phi rotation (degrees)
DSN100_1_001.2a;2-900	8000	-9.12	10.00	9.12	(-110) -- (-20.2)
DSN250_001.2a;2-710	16000	-22.00	25.00	21.86	(-160) -- (-89.2)
DSN400_001.2a;2-668	48000	-22.65	40.00	32.70	(-170) -- (-103.4)
DSN23N_001.2a;2-998	16000	-23.00	10.00	9.10	140 -- 239.6
DSN250N_001.2a;2-805	32000	-18.00	25.00	21.90	(-20) -- 60.3
DSN250j_001.2a;2-905	32000	-18.00	25.00	21.90	(-20) -- 70.3
DSN250t_001.2a;2-360	32000	-18.00	25.00	21.90	?
DSN250z_001.2a;2-321	32000	-18.00	25.00	21.90	?
DSN375N_001.2a;2-999	48000	-18.90	37.50	31.00	(-110) -- (-10.3)

4.4 References

1. Korszun, Z.R., A.M. Saxena, and D.K. Schneider, *Structural Biology Facilities at Brookhaven National Laboratory's High Flux Beam Reactor.*

2. Axe, J.D. and R. Greenberg, *HFBR Handbook.* 1992

3. Greenberg, R., *Faxed information on HFBR.* 1995

4. Schoenborn, B.P., *A Protein Structure is Only as Good as the Data,* in *Neutrons in Biology,* B.P. Schoenborn, Editor. 1984, Plenum Press: New York. p. 261-279.

5. Bacon, G.E., *Neutron Diffraction.* Third ed. 1975, Oxford: Clarendon Press. 636.

6. Convert, P. and J. Forsyth, ed. *Position-Sensitive Detection of Thermal Neutrons.* 1983, Academic Press: San Diego.

7. Pflugrath, J. and A. Messerschmidt, *Munich Area Detector NE System (MADNES).* Cold Spring Harbor Laboratory.

CHAPTER 5. The Effect of Protein Perdeuteration on the Background Levels of Neutron Diffraction Data

Protein perdeuteration is expected to enhance the signal-to-noise inherent to neutron diffraction data by reducing the hydrogen incoherent scattering contribution in the sample. This chapter describes the theory on which this prediction is based and the experiments performed to test the hypothesis. Under cursory inspection, the data demonstrate that no difference is seen in background levels under a variety of hydrogenated and perdeuterated conditions. However, the factors that contribute to the background level of neutron diffraction data include more than solely the incoherent scatter due to the crystallized protein. Mother liquor components, protein ligands, and elements of the experimental conditions (e.g. air scatter, capillary composition, electronic detector noise, etc.) can make significant contributions to the background level of the data. When considered in the context of the complete experimental conditions, the data suggest that *for these experiments*, factors other than incoherent scattering by the sample constitute the majority of the background level, effectively swamping out any possible improvement in signal-to-noise due to reducing the incoherent scattering by the protein itself.

5.1 Background: Coherent vs. Incoherent Scattering

Every atomic element interacts uniquely with neutrons creating both coherent and incoherent scattering. Coherent scattering creates constructive and destructive interference, giving rise to a diffraction pattern when a crystal is exposed to a neutron source. Incoherent scattering, which is completely random, cannot interfere with other diffracted particles, and contributes only to the background of neutron diffraction data. Isotopes of a single element may have radically different scattering profiles, the most

biologically relevant example being hydrogen and its isotope deuterium [1]. Because the absolute value of their scattering lengths are similar, deuterium atoms (82% coherent, 18% incoherent) contribute much less to the background than hydrogen atoms (2% coherent, 98% incoherent).

It is possible to calculate the extent to which the incoherent scattering will decrease when replacing hydrogen atoms with deuterium atoms in a crystal. It is assumed that the total scattering is the linear addition of coherent and incoherent scattering. Equation 5.1 can be used to calculate a factor that indicates the reduction of incoherent scattering as a function of the fraction of hydrogen and deuterium atoms in a protein crystal. At zero percent hydrogen, the incoherent scatter contributing to the background should be decreased by a factor of 40.

Equation 5.1: **Reduction Factor =** $$\frac{80}{f_D * 2 + f_H * 80}$$

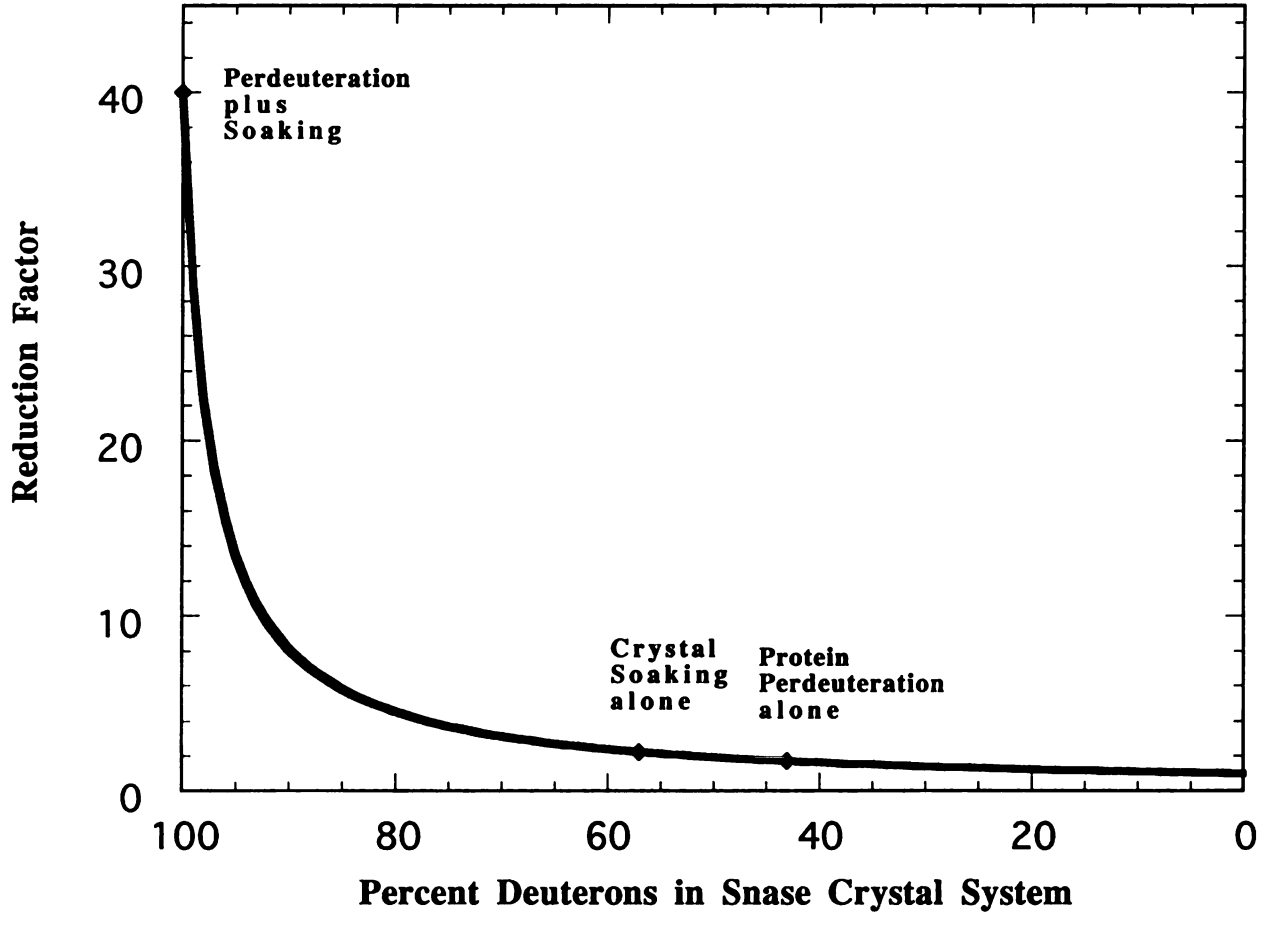
where f_D/f_H = fraction of deuterium/hydrogen atoms in the system (note: on a per atom basis, the number of incoherently scattered neutrons per unit exposure by hydrogen is 80 and by deuterium is 2)

When considering the effect of protein perdeuteration on incoherent scattering it is important to keep in mind that neutron diffraction data are a result of diffraction from the crystal system *as a whole.* The crystal system is made up not only of protein molecules, but ligands, solvent, and mother liquor components as well. In the particular case of Snase, the unit cell (48.0Å x 48.0Å x 63.5Å, space group = $P4_1$) has a solvent content of 38%. In the simplified example of a mother liquor of 55.6 M water and 4 perdeuterated protein molecules, one unit cell will contain 4852 labile (3720 from the water, 1132 from the protein) and 3688 non-labile (protein only) protons. Using these parameters and equation 5.1, soaking Snase crystals in D_2O will substitute 57% of the protons in the crystal system with deuterons, while protein perdeuteration will replace the

remaining 43%. In this simplified case, soaking will replace over half of the protons with deuterons, yielding a reduction in incoherent scatter by a factor of 2.3. The combination of protein perdeuteration and crystal soaking maximally reduces incoherent scatter. When used alone, 100% protein perdeuteration will reduce incoherent scatter by a factor of 1.7, while in conjunction with soaking, the reduction in incoherent scatter will be an additional factor of 17.4. Fig. 5.1 shows the continuum of incoherent scattering reduction from a crystal system containing 100% hydrogen atoms to one containing 100% deuterium atoms.

In the actual experiments, Snase crystals were not soaked in pure D_2O, but rather a mother liquor solution made up of 22.2 M D_2O, 4.7 M MPD, 10.5 mM potassium phosphate, 0.6 mM calcium chloride, 1.2 mM citric acid and 0.4 mM inhibitor (pdTp). The incoherent scattering of potassium, phosphate, calcium, and chloride will be insignificant for two reasons: 1) they scatter so few neutrons incoherently (all cases < 3 incoherently scattered neutrons per atom per unit exposure) and 2) there are so few of these atoms in the unit cell. The potassium phosphate, calcium chloride and citric acid stock solutions were made up in D_2O; based on their molar concentrations, hydrogen contamination due to their non-labile hydrogen atoms will be negligible. The bound inhibitor contributes 48 hydrogen atoms to the unit cell and the MPD (99.0% perdeuterated) adds another 22 hydrogen atoms. In combination, the inhibitor and the MPD cause a 1% hydrogen contamination of the unit cell. In addition, Snase was reproducibly perdeuterated to a level of only 96%, causing 2% hydrogen contamination. Given the known hydrogen contaminants, the Snase crystal system should contain 97% deuterons and 3% protons, leading to the expectation that the incoherent scatter of a soaked, perdeuterated crystal will be reduced by a factor of 18.4.

Fig 5.1: Effect of Replacing Protons with Deuterons on Incoherent Scattering



5.2 The Source of Average Background Scattering Levels in Neutron Diffraction Data

The background level in neutron diffraction data is the result of several factors: neutrons that are part of the general background in the neighborhood of the reactor, diffuse scattering (both coherent and incoherent) of the direct beam by any air or solid material in its path, incoherent scattering by the sample, and electronic properties of the detector. Precautions have been taken in the design of the station to minimize background such as a beam tube that shields the experiment from the non-monochromated beam and a helium-filled snout on the detector that cuts down on air scatter. Apart from the crystallized protein itself, hydrogen atoms found in the mother liquor, the capillary, and the air between the beam tube and the sample or the sample and the detector snout will contribute to the average background level.

Protein perdeuteration *only* affects the portion of the average background level that is caused by incoherent scattering by the sample. Specifically, it decreases the incoherent scattering caused by the protein itself and will have no relevance to other sources of background noise. It is difficult to predict the contribution of the incoherent scattering by the sample to the overall background; however, based on the fact that measurable decreases in background are observed upon soaking a crystal in D_2O (for example Mason's lysozyme work, [2]), incoherent scattering by the sample itself is expected to be a significant contributor to the total background. Given the estimate that replacing non-labile protons with deuterons will reduce incoherent scattering by the crystal almost as much as soaking alone, protein perdeuteration should also cause a significant decrease in the average background of neutron diffraction data.

5.3 Experimental Strategy

In order to evaluate the contribution of protein perdeuteration to total background, experiments were designed to document the level of background due to sources other than incoherent scattering by the protein. These factors include electronic noise, ambient neutrons, diffuse scatter of the direct beam, and the effect of placing various samples in the beam. The level of ambient neutrons was detected by collecting data frames with the neutron beam blocked at the end of the beam tube by cadmium. Detector noise was recorded by covering the snout with a material made from boron epoxy. (Both cadmium and boron absorb neutrons.) To establish the level of diffuse scattering by the direct beam, data frames were collected with no sample in the beam. In these experiments, the detector was swung out so that the direct beam did not fall on the detector. The contribution of the quartz capillary was determined by collecting data with an empty capillary in the beam. Finally, to obtain measurements of total backgrounds, data were collected under identical experimental conditions from both native (containing hydrogenated protein) and perdeuterated crystals that had been soaked in deuterated mother liquor.

5.4 Results of Background Measurements

Tables 5.1, 5.2, and 5.3 present the conditions under which data were collected along with the average background value for each data set. A program was written to calculate the average background pixel value over an entire data frame and compute the standard deviation. This average background pixel value per frame was then used to calculate an average background value and average standard deviation for an entire data set. This can be done in a direct manner if the data set contains no diffracted intensities; however, in the case where reflections are present in the data, they must not be included

Table 5.1: Average Background Measurements: Native vs. Perdeuterated Crystals

Files	Description	Crystal (cubic mm)	Collimator (cm)	2Θ (degrees)	Monitor Counts	Degrees of Data	Helium	Average Background	Sigma	Number of Pixels
NATIVE CRYSTALS										
HSN8_1_001.2a;2-101	Small native crystal	1.2	0.1	-8	8000	9.9	no	0.3608	0.0004	2169300
HSNHEL_001.2a;2-101	Same as HSN8_1 plus helium	1.2	0.1	-8	8000	9.9	yes	0.3399	0.0004	1974269
SMALLH_2_001.2a;2-788	Same as HSN8_1/larger collimator	1.2	0.2	-8	8000	78.6	yes	0.9202	0.0002	17070384
HSNASE_001.2a;1-998	Large native crystal	6.1	0.3	-8	8000	99.7	no	2.4143	0.0004	21630778
HSN38T_001.2a;1-130	Same as HSNASE/higher resolution	6.1	0.3	-38	24000	12.9	no	3.5996	0.0012	2821920
HSN45_001.2a;1-231	Same as HSNASE/higher resolution	6.1	0.3	-45	24000	23.0	no	3.8219	0.0009	5014276
PERDEUTERATED CRYSTALS										
DSN8_1_001.2a;1-100	Perdeuterated crystal	2.2	0.1	-8	8000	9.9	no	0.4065	0.0004	2146700
DSN8_2_001.2a;2-490	Same as DSN8_1/larger collimator	2.2	0.2	-8	8000	48.8	yes	0.9529	0.0002	21647040
DSN23_2_001.2a;2-999	Same as DSN8_2/higher resolution	2.2	0.2	-23	16000	99.7	yes	1.3880	0.0004	10608583
DSN38_2_001.2a;2-167	Same as DSN8_2/higher resolution	2.2	0.2	-38	20000	16.5	yes	1.5475	0.0007	3602598
DSN38_3_001.2a;2-739	Continuation of DSN38_2	2.2	0.2	-38	20000	73.7	yes	1.5681	0.0003	14001600
SMALLD_001.2a;2-336	Small perdeuterated crystal	0.5	0.2	-8	8000	33.4	yes	0.7909	0.0003	7269533
Note: In all cases, ZZ = 0, TA = 0										

Table 5.2: Average Background Measurements

Files	Description	Crystal (cubic mm)	Collimator (cm)	2θ (degrees)	Monitor Counts	Degrees of Data	Helium	Average Background	Sigma	Total Number of Pixels
APER_001.2a;2-4	No sample in beam	NA	0.1	-11	40000	0.0	yes	4.4620	0.0083	65124
CAP_001.2a;2-4	Empty capillary	NA	0.1	-11	40000	0.0	yes	4.4267	0.0082	65124
BLACK_001.2a;2-4	Capillary filled with D20	NA	0.1	-11	40000	0.0	yes	4.3784	0.0081	65124
RED_0012a;2-4	Capillary filled with H20	NA	0.1	-11	40000	0.0	yes	4.4302	0.0082	65124
SNOUT_001.2a;2-4	Snout covered	NA	0.1	-11	40000	0.0	yes	0.4829	0.0027	65124
ZERO_001.2a;2-4	0 mm aperature	NA	0.0	-11	40000	0.0	yes	0.4464	0.0026	65124
NOCAP_001.2a;2-101	No capillary in beam	NA	0.1	-8	8000	9.9	no	0.2905	0.0004	2167108
DRYCAP_001.2a;2-101	Brand new capillary, empty	NA	0.1	-8	8000	9.9	no	0.2891	0.0004	2166373
WETCAP_001.2a;1-100	Perdeuterated crystal out of beam/ mother liquor on sides of capillary	NA	0.1	-8	8000	9.9	no	0.3294	0.0004	2169853
HSN8_1_001.2a;2-101	Small native crystal	1.2	0.1	-8	8000	9.9	no	0.3605	0.0004	2169300
								(0.3639)*		
DSN8_1_001.2a;1-100	Large perdeuterated crystal	2.2	0.1	-8	8000	9.9	no	0.4065	0.0004	2146700

NOTE: APER, CAP, BLACK, RED, SNOUT, and ZERO were collected one year later than the rest of the background data

* Average background corrected for crystal size [1.11(.3605 - .3294) + .3294]

Table 5.3: Average Background Measurements of Neutron Data Collected for the Perdeuterated Snase Structure

Files	Crystal (cubic mm)	Collimator (cm)	2Θ (degrees)	DIST (mm)	T (degrees)	Monitor Counts	Degrees of Data	Helium	Average Background	Sigma	Number of Pixels
				REFINED VALUES							
DSN8_2_001.2a;2-999	2.2	0.2	-8.09	621.97	0.01	8000	99.70	yes	0.9529	0.0002	21647040
DSN100_1_001.2a;2-877	2.2	0.2	-8.99	631.05	8.98	8000	87.50	yes	1.0941	0.0002	19005828
DSN20_2_001.2a;2-851	2.2	0.2	-17.00	649.87	18.06	16000	84.90	yes	1.2439	0.0003	18440970
*DSN8_001.2a;2-200	2.2	0.2	-17.12	654.71	18.01	16000	19.80	yes	1.3308	0.0006	4317868
DSN23_2_001.2a;2-490	2.2	0.2	-22.99	621.88	0.01	16000	48.80	yes	1.3880	0.0004	10608583
DSN25_2_001.2a;1-973	2.2	0.2	-7.62	665.96	21.57	16000	97.20	yes	1.1424	0.0002	21105592
DSN250_001.2a;2-710	2.2	0.2	-20.41	669.15	22.71	16000	70.80	yes	1.5449	0.0003	15385992
DSN23N_001.2a;2-998	2.2	0.2	-22.52	630.12	9.60	16000	99.60	yes	0.9709	0.0002	21630226
DSN38_2_001.2a;2-167**	2.2	0.2	-38.00	0.00	0.00	20000	16.50	yes	1.5475	0.0007	3602598
DSN38_3_001.2a;2-712**	2.2	0.2	-38.00	0.00	0.00	20000	71.00	yes	1.5681	0.0003	14001600
DSN250N_001.2a;2-805	2.2	0.2	-10.02	669.57	21.90	32000	80.30	yes	1.8902	0.0003	17441712
DSN250J_001.2a;2-905	2.2	0.2	-15.89	668.58	21.90	32000	90.30	yes	1.5816	0.0003	19613424
DSN250T_001.2a;2-360	2.2	0.2	-10.93	621.26	0.00	32000	35.80	yes	1.0158	0.0004	7788744
DSN250Z_001.2a;2-321	2.2	0.2	-9.83	670.40	21.97	32000	31.90	yes	1.8117	0.0005	6941169
DSN10_2_001.2a;2-999	2.2	0.2	-33.27	627.14	10.80	48000	99.70	yes	3.0283	0.0004	21662192
DSN275_2_001.2a;2-657	2.2	0.2	-27.23	674.82	26.25	48000	65.50	yes	2.7480	0.0004	14238866
DSN275_5_001.2a;2-400	2.2	0.2	-27.24	678.93	26.23	48000	39.80	yes	3.3852	0.0006	8660695
DSN375_001.2a;2-782	2.2	0.2	-13.73	720.83	31.16	48000	78.00	yes	2.9474	0.0004	16951104
DSN375_2_001.2a;2-222	2.2	0.2	-13.75	720.88	31.20	48000	22.00	yes	2.9979	0.0008	4796650
DSN400_001.2a;2-651	2.2	0.2	-18.89	735.48	33.74	48000	64.90	yes	2.9640	0.0005	14108996
DSN375N_001.2a;2-999	2.2	0.2	-10.88	724.42	31.00	48000	99.70	yes	2.2545	0.0003	21661520

*Error in name convention, see Table 4.4

**Data not processed, detector position not refined

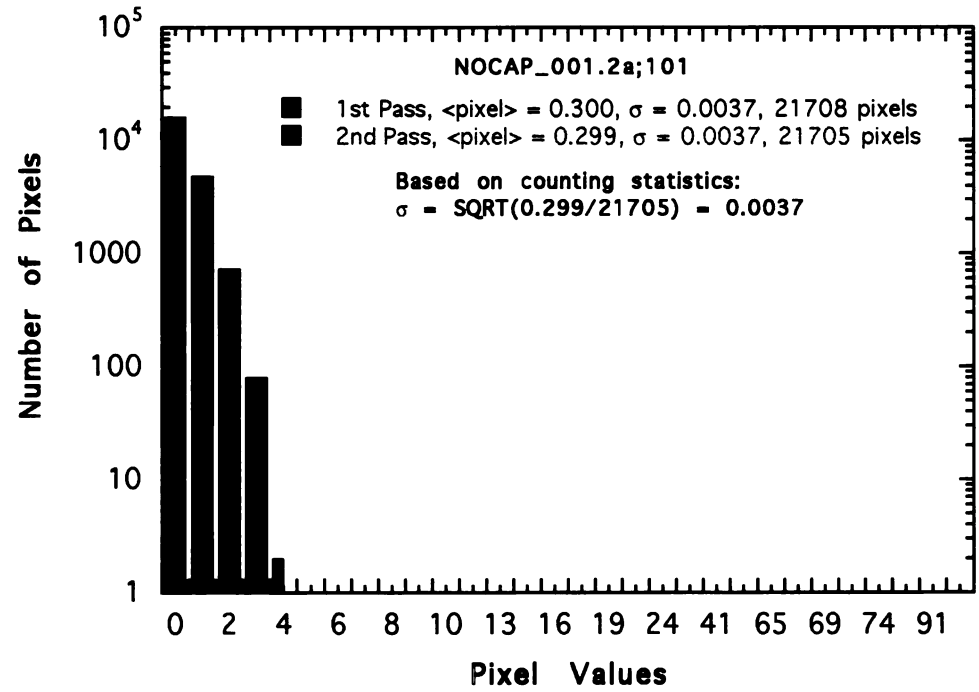
NOTE: With the beam centered on the detector, DIST = 623 mm

in background calculations. In these cases, reflections were removed from the data frames in the following manner: the average pixel value and standard deviation were calculated over one data frame, pixels containing values five standard deviations above the average were rejected and the process was repeated until no pixels were rejected. At this point a final average and standard deviation were computed using only pixels containing background. In all cases, the final standard deviation was the same as the standard deviation based purely on counting statistics. Fig. 5.2 depicts the results of calculating the average background value for data frames with and without reflections.

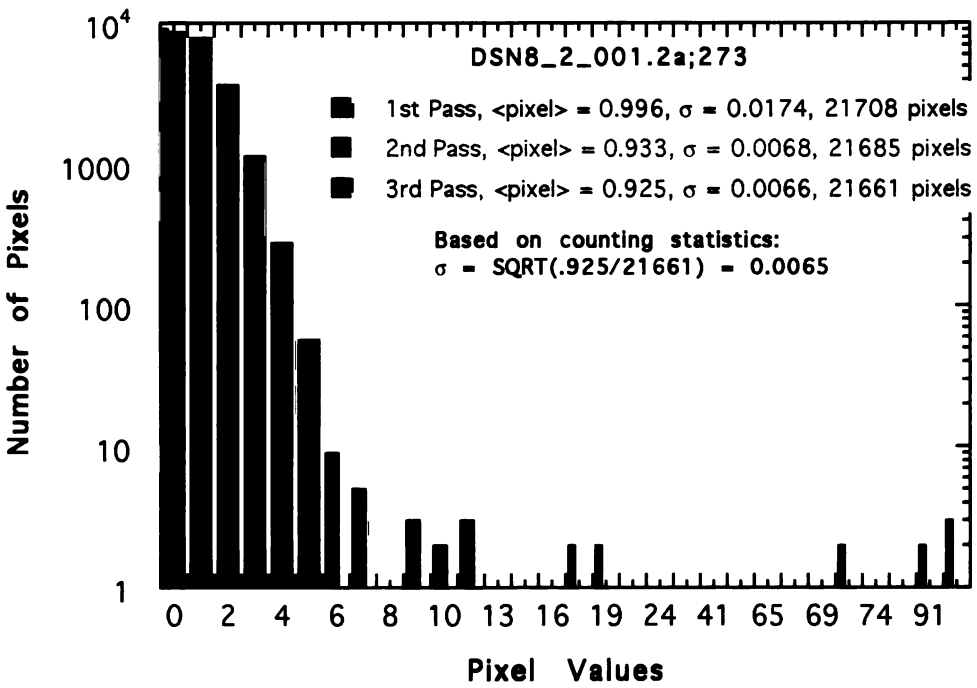
Several experimental parameters will affect the average background value: the collimator diameter, the monitor counts, the detector position, crystal size and the atomic elements in the path of the direct and diffracted beam. An increase in collimator diameter or monitor counts will result in an increase in background. As the horizontal or vertical swing angle (2Θ) of the detector increases, the background due to air scatter will decrease. A larger crystal will increase the amount of recorded background scatter. The addition of helium to the snout attached to the front of the detector will decrease air scatter between the sample and the detector.

The experiments documented in Table 5.1 depict expected changes in average background values in relation to the experimental parameters. A comparison between helium in the snout vs. no helium (HSN8_1 vs. HSNHEL) reveals a small (6%), but significant (52σ), decrease in average background level with the addition of helium. The effect of crystal size on background scatter is demonstrated by comparing experiments DSN8_2 and SMALLD in which all parameters were held constant. A 17% lower average background value is observed with the smaller crystal (SMALLD). Comparing experiments HSNHEL and SMALLH shows the effect of doubling the collimator diameter. As expected, the larger the collimator diameter, the higher the observed background measurement; however, the increase is only a factor of 2.7 rather than the expected factor of 4.

Fig. 5.2a: Calculating Average Background Data Frames without Reflections



5.2b: Calculating Average Background Data Frames with Reflections



A correction for differences in crystal size must be made in order to directly compare the background level between native and perdeuterated crystals. Such a size correction can be made by calculating a scale factor that relates native and perdeuterated data. This was done by first scaling observed data from both a native and a perdeuterated crystal to calculated data derived from appropriate models. Specifically, data from SMALLH were scaled to calculated reflections from a model consisting of protons at non-labile sites and deuterons at labile sites (scale = 527.68), while a model containing only deuterons was used for DSN8_2 (scale = 473.49). Once all reflections were on an absolute scale with one another a scale factor of 1.11 was calculated for the background. Since both crystals were completely bathed in the neutron beam, this scale factor is applicable to any comparison of these two crystals, even under different experimental parameters.

As described in Section 5.3, experiments were done to assess the contribution of incoherent scattering by the protein to the average background value. These background measurements were repeated once - the sets of experiments differ in two ways: 1) the neutron exposure is significantly different (8000 vs. 40,000 monitor counts), and 2) the horizontal 2Θ swing angle of the detector differs by 3°. In addition, the experiments were done over a year apart. By comparing the experiment in which the snout is covered (SNOUT) to one where it is not (APER), the electronic noise of the detector is shown to comprise about 10% of the average background value when no sample is in the beam. The impact of ambient neutrons on total background is seen to be negligable when comparing background average when the snout is covered (SNOUT) versus when the beam tube is blocked (ZERO). The experiment in which no sample is in the beam (APER) determines the level of diffuse scatter from the direct beam. When the beam tube is completely blocked (ZERO), average background is 0.4464; however, this value jumps to 4.4620 when an aperture with a 0.1 cm diameter pinhole is used. By comparing the series of experiments, no sample (APER), empty capillary (CAP), capillary with D_2O

(BLACK), and capillary with H_2O (RED), a progressive increase in background is expected due to an increase of incoherent scattering by the sample. However, the data do not show this trend; in fact, they yield the surprising result that the background level is highest when nothing is in the beam.

Overall, the earlier series of experiments to assess the contribution of incoherent scattering by the protein to the overall background level appear to follow expectations (Table 5.2). Placing a capillary wet with mother liquor (WETCAP, <background> = 0.3294, σ = 0.0004) results in a higher average background value than having no capillary in the beam (NOCAP, <background> = 0.2905, σ = 0.0004); likewise, placing a crystal in the beam (HSN8_1, <background> = 0.3605, σ = 0.0004 and DSN8_1, <background> = 0.4065, σ = 0.0004) produces a higher background average than a wet capillary. However, placing a capillary in the beam *decreases* the average background value (DRYCAP, <background> = 0.2891, σ = 0.0004 vs. NOCAP, <background> = 0.2905, σ = 0.0004) and a hydrogenated crystal, even when correcting for its smaller size, yields a *lower* average background than a perdeuterated crystal (HSN8_1, <background> = 0.3639, σ = 0.0004 and DSN8_1, <background> = 0.4065, σ = 0.0004). Similar to the later series of experiments, compared to cases where crystals were placed in the beam, an unexpectedly high background is observed for the case when no sample is in the beam.

Although background level is influenced by several predictable factors, the data in Table 5.3 suggests that other, unknown factors come into play. Several of the data sets (DSN20, DSN38, DSN275, DSN375) were collected at two different times. Because the experimental conditions were held constant for the first and second half of data collection, these data sets reveal that the background level of distinct data sets collected under identical conditions can differ up to 20%.

5.5 Discussion

Major discrepancies between expected and observed results make it difficult to assess the impact of protein perdeuteration on background level in the neutron diffraction data described. The largest problem arises from the observation that altering the properties of the diffracted beam by adding incoherent scatterers to the sample did not necessary lead to an increase in background level. For example, the background level generated by a capillary full of H_2O (RED) is lower than the background level produced when no sample is placed in the direct beam (APER). In similar fashion, the corrected average background value for a native crystal (HSN8_1) is lower than a perdeuterated one (DSN8_1). By its very definition, incoherent scattering can *only increase* the background of diffraction data, leading to the conclusion that an alternative explanation must exist to describe these data.

Given the data collected, it is possible to asses the contribution of the incoherent scattering by a crystal to the total background. Because average background levels can vary by 20% between identical experiments, it is reasonable to conclude that the differences between having no sample in the beam (NOCAP), an empty capillary (DRYCAP), and a capillary containing mother liquor (WETCAP) are inconsequential. By averaging these three background levels, an average background value of 0.3030 is representative of all background scatter except what is contributed by the crystal itself. Using the same argument, the difference between the average background levels of native and perdeuterated crystals (HSN8 and DSN8) is also insignificant. Therefore, the two crystals' background levels can be averaged together for a value of 0.3852. After subtracting out all other contributions to background scatter, it is seen that a crystal contributes 21% to the total average background. Most notable is the fact that by far the largest percentage (79%) of the average background level is not caused by the crystal, but

rather by factors such as diffuse scatter of the direct beam, ambient neutrons, and electronic detector noise.

The variability in background levels under identical conditions (up to 20%) combined with the observation that a crystal is estimated to contribute only 21% to the total background level makes it impossible to assess the effect of protein perdeuteration under these conditions. The observed fluctuations in background levels between comparable data sets are masking the contributions made by the crystal, and therefore the protein itself. This conclusion is supported by a second set of experiments in which no significant difference was found between having no sample in the beam (APER), an empty capillary (CAP), D_2O (BLACK) and H_2O (RED).

5.6 Conclusion

Protein perdeuteration should affect the background level of neutron diffraction data by reducing the amount of incoherent scattering caused by the protein. A reduction in incoherent scattering directly correlates to better neutron data by virtue of the fact that the signal-to-noise will be improved. However, the background level is made up of multiple factors, of which scattering by the protein itself is only a part. The largest background contributors to the data collected for *Staphylococcal* nuclease are not the protein, or even the crystal system, but rather factors involving diffuse scatter from the direct beam. In order for protein perdeuteration to make a significant difference in signal-to-noise, the protein itself must be a major source of the background level. In the future, sources of neutron contamination must be minimized so that protein perdeuteration can have significant impact on the quality of neutron data.

5.7 References

1. Bacon, G.E., *Neutron Diffraction*. Third ed. 1975, Oxford: Clarendon Press. 636.

2. Mason, S.A., G.A. Bentley, and G.J. McIntyre, *Deuterium exchange in lysozyme at 1.4Å resolution,* in *Neutrons in Biology,* B.P. Schoenborn, Editor. 1984, Plenum Press: New York. p. 323-334.

CHAPTER 6. Data Processing

Processing and merging of the neutron diffraction data is a critical step in obtaining a highly accurate protein structure. Software originally designed for use with X-ray diffraction data was modified to accommodate the parameters unique to the H3A beam line; this program (MADNES) is capable of distinguishing reflections from background, using these found reflections to determine an orientation matrix, and then evaluating intensity for all reflections in a given data set. MADNES parameters were adapted to the nature of the neutron data (for example low signal-to-noise) in order to optimize data processing results. Each data set (all reflections collected at a given detector position) was processed independently, the individual settings were then merged together. Before merging, all of the observed data were scaled to a calculated data set as a means to correct for both anisotropic scaling and the low redundancy of the data. An analysis of the final data set in terms of completeness and quality is presented at the end of the chapter.

6.1 Modified Version of MADNES

MADNES (Munich Area Detector Non-Enraf System) is a diffraction data collection software program designed to be used with electronic two-dimensional position-sensitive detectors [1]. It has the capacity to: 1) search for and find peak centers when the crystal orientation is unknown, 2) determine the unit cell parameters and crystal orientation by auto-indexing based on the reflections found during the search, 3) refine the unit cell and orientation of the crystal as well as the detector parameters, and 4) collect reflections for a complete data set, updating the orientation matrix during data collection if necessary [1]. Although the program was originally intended to be used with X-ray diffraction data, collected on an Enraf-Nonius FAST television area detector,

there is nothing inherent to the software to prevent its use with neutron diffraction data collected on a neutron position-sensitive detector.

Certain hardware and software specifications are required in order for MADNES to function properly. The hardware must be able to integrate diffracted quanta (in this case neutrons) in memory to form an image of the detector surface. The software must be able to transfer an integrated image to a FORTRAN array, convert from millimeters to device pixels and vice versa, and correct for the inhomogeneity of pixel response [1]. The program is intended to work "on-line," that is, to integrate reflections as they are collected on the detector; however, it can also be used "off-line," that is, to process an entire data set that has been previously collected. For all of the neutron data, MADNES was used "off-line" during data processing.

In order to adapt MADNES to process neutron data collected at BNL, it was modified to accommodate the physical parameters of the H3A station. These modifications have been made and documented in a new version of the program called MADNEU. The following changes and definitions were made according to the instructions given in the MADNES documentation [1]. First, the detector coordinate system corresponding to the physical setup of the H3A beam line was defined. This coordinate system defines the position of a reflection on the detector by its x and y coordinates, adding a phi coordinate to describe the position of the crystal rotation axis. The center of a reflection will occur at Xms, Yms, and Phi, respectively [1]. A second coordinate system centered on the goniostat was also defined. Once again, there are three axes used: X is parallel to the crystal rotation axis (the positive direction downwards), Z is anti-parallel to the primary beam (the positive direction towards the source), and Y is normal to the XZ-plane forming a right-handed coordinate system [1]. Next, the software was modified so that raw data frame output from the data collection program XTAL2D [2] could be read directly into MADNES; this involved information concerning the format, dimensions and header of the data files. Finally, a non-uniformity file was

created to compensate for the dead pixels found along the each edge of the detector (see Chapter 4), and the center of the beam was defined by taking stills of the attenuated beam.

6.2 Using MADNES to Process Neutron Data

The general principles of data processing are identical between X-ray and neutron diffraction data; however, the parameters within MADNES must be adjusted to accommodate properties inherent to neutron data as well as the specifics of data collection. For example, the amplitude of the neutron reflections, as well as the signal-to-noise, is much lower in neutron than X-ray data, necessitating the criteria for peak picking to be adjusted to accept low intensity reflections. Although the steps of data processing (finding reflections, auto-indexing, orientation matrix refinement and data collection) are normally sequential, quality data results from iterative passes through the cycle. For instance, reflections from data collection can be treated as found reflections and often result in an improved starting matrix. The most difficult step of processing these neutron data proved to be determining good orientation matrices for each detector setting; poor signal-to-noise, weak data at high resolution, and unknown detector parameters caused multiple problems at the auto-indexing step. In the end, all but one data set was processed through a combination of iteratively using MADNES and manual intervention.

6.2.1 Finding Reflections

The subroutine FIND uses several parameters to determine reflections from background: a minimum pixel value (CUTOFF), the number of standard deviations above background (SIGMA), the limitations set by the edge of the detector or bad pixels

(LIMITS), a minimum number of pixels above background in a peak and the number of pixels used for background determination (PKCRIT), and the distance between neighboring peaks (NEIGH) (Table 6.1). The parameters are interdependent of one another; for example SIGMA uses PKCRIT2 and PKCRIT3 to set an area for background average and standard deviation calculations. In turn, PKCRIT determines peaks to be PKCRIT1 pixels above the threshold specified by SIGMA. FIND assigns an error code to each peak that it accepts (Table 6.2); these error codes were designed for use in COLLECT, making some of them inapplicable for FIND reflections. For example, error code 512, too big in phi, has no meaning in FIND because at this stage there is no realistic estimate of crystal mosaicity.

Typical FIND parameters are shown in Table 6.1; with the exception of CUTOFF, identical parameter values were used for each detector setting. The CUTOFF parameter, which sets a minimum pixel value for reflections, must be adjusted as a function of exposure time. Fig. 6.1 shows slices through two FIND peaks, one from a low resolution detector setting (dsn8) and the other from a high resolution detector setting (dsn10); in both cases, these peaks were flagged by FIND to be good reflections. At low resolution, many peaks are well above background; however, at high resolution, where reflections are weaker, the CUTOFF parameter must be set judicially so that peaks can be distinquished from the higher background. For some settings, a problem was discovered with the phi rotation. As noted in Table 6.1, the crystal rotation was usually made in increments of 0.1°; however, for some of the detector settings there was a problem with the phi encoder. Through trial and error, the phi increment which most closely matched the experimental value was found to be 0.112° for these data sets (dsn23n, dsn250n, dsn250j, dsn250t, dsn250z, dsn375n).

It is important to set the FIND parameters so that only real reflections are found; ideally, these reflections will have the error code of 0 (GOOD). For example, if the minimum pixel value or sigma level is set too low, noise peaks will be picked up. Often

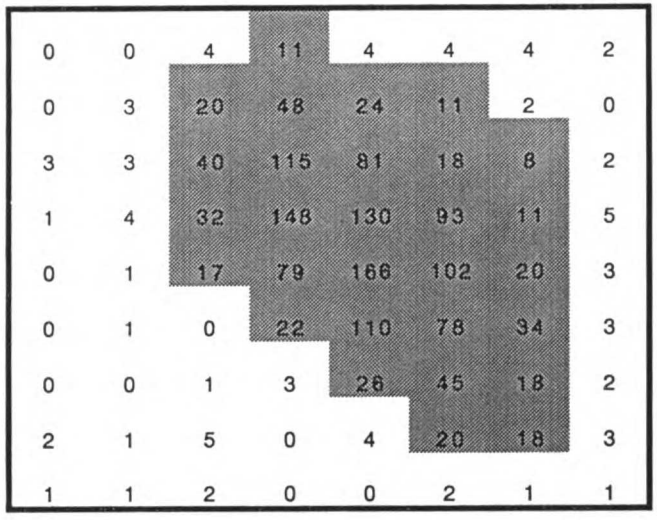
Table 6.1: FIND Parameters

FIND Parameter	Description	Value
CUTOFF	Minimum pixel value for a reflection	5 to 10
INC	Rotation angle increment (Phi)	0.1 **
LIMITS	Data frame limits	
	min Y	5
	max Y	250
	min Z	5
	max Z	110
NEIGH	Set neighbor closeness	
	neigh1	12
	neigh2	12
PKCRIT	Peak criteria and background size in pixels	
	pkcrit1	2
	pkcrit2	24
	pkcrit3	24
SIGMA	Pixels must be (sigma * SD + <background>)	3.5
	to be eligible for a reflection	
** INC = 0.112 was used for dsn23n, dsn250n, dsn250j, dsn250t,		
dsn250z and dsn375n because the phi rotation slipped for these data sets		

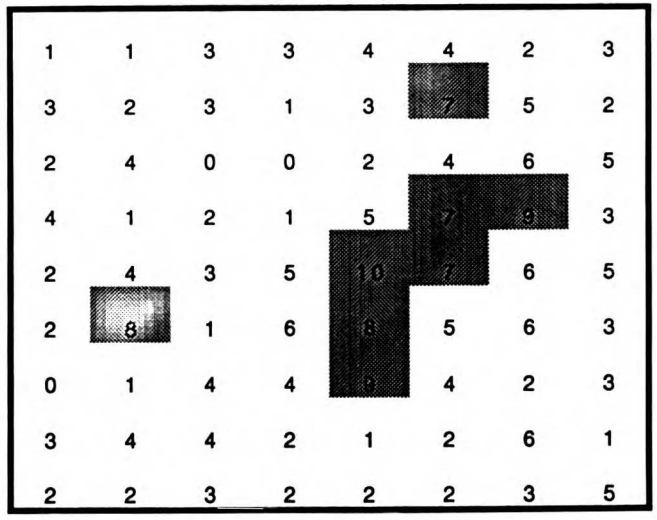
Table 6.2 Error Codes

ERROR CODE	INTERPRETATION
0	Good, strong reflection
1	Weak reflection
2	On edge of data array in Yms
4	On edge of data array in Zms
8	On edge of data array in Phi
16	Too far from predicted Yms
32	Too far from predicted Zms
64	Too far from predicted Phi
128	Too big in Yms (refln wider than expected)
256	Too big in Zms
512	Too big in Phi
1024	Background bad
2048	Background standard deviation bad
4096	Reserved for future use
8192	Fobs less than or equal to 0.0
16384	Bad non-uniformity pixels in data array

Figure 6.1: FIND peaks from low and high resolution detector settings



Slice through FIND peak from low resolution setting
(dsn8_2_001.2a;208)
ERROR CODE = GOOD
CUTOFF = 5
Average background for dsn8 = 0.9529



Slice through FIND peak from high resolution setting
(dsn10_2_001.2a;859)
ERROR CODE = GOOD
CUTOFF = 6
Average background for dsn10 = 3.0283

NOTE: The shaded boxes contain pixel values above CUTOFF

these peaks can be recognized by their error code, but inevitably, some of these bad "reflections" will be passed to the auto-indexing data array, causing difficulty in determining the orientation matrix for the data set. Likewise, if the neighboring reflection parameter is too big, reflections will be missed. In practice, even after careful adjustment of the FIND parameter, reflections with bad error codes are found. One way around this problem is to select only GOOD reflections out of the FIND array and feed them into the auto-indexing step - this proved to be useful provided that a sufficient number of GOOD reflections had been found.

6.2.2 Auto-indexing

When a sufficient number of reflections have been found (ideally 50 - 70; however, as few as 15-20 have been used), MADNES uses one of three functions to determine the orientation of the crystal given the physical parameters of the system and the set of reflection positions. Of the three functions, AUTI was never used, AUTJ worked successfully if the detector parameters were known accurately, and ENDEX proved to be the most robust and reliable after it was altered to include TAU3 (equivalent to a vertical 2Θ swing angle). Both AUTJ and ENDEX require the parameters describing the detector position as well as wavelength information.

The key criterion to successful auto-indexing is the extent to which the detector parameters in MADNES match the physical reality of the experiment. When using AUTJ, the direct beam position (CCX, CCY), the distance from the crystal to the detector (DIST), and the three detector angles (TAU3 = tilt angle, TAU2 = 2Θ swing angle, TAU1 = rotation around the beam) cannot significantly diverge from the actual experimental values; if they do, auto-indexing fails. However, even when the detector parameters are relatively far off, ENDEX will present the user with several possible orientation matrices, albeit poor ones, which can often be used as starting points for matrix refinement.

ENDEX reports those reflections (out of a maximum of 50) used for auto-indexing, a figure of merit (0.0 being the best) for each possible candidate solution, and the cell parameters and standard deviations it has determined. When ENDEX found the expected Laue symmetry (tetragonal), the solution was always correct and was accepted even if it did not have the highest figure of merit. Sometimes the detector parameters were so inaccurate that ENDEX could not even find the correct Laue symmetry (dsn250n, dsn250t, dsn250z, and dsn375n). In these cases, the candidate solutions were improved by manual adjustment of TAU2 or TAU3 and the solutions were monitored by the figure of merit, the number of reflections used, and the proximity of the cell angles to 90°. In all cases, this procedure eventually led to successful matrix refinement. The observation that essentially the same missetting angles were found for each independently processed detector proves that none of the settings were mis-indexed.

6.2.3 Refinement

Once an approximate orientation matrix is found with ENDEX or AUTJ, the subroutine REFINE is used to refine it; REFINE does this by adjusting several parameters including the unit cell dimensions (a,b,c, α, β, γ), the crystal missetting angles (PHIX, PHIY, PHIZ), the crystal to detector distance (DIST), and the detector position parameters (Table 6.3).

In principle, the missetting angles and cell dimensions can be refined together with the detector parameters. However, there are two inherent characteristics of REFINE that compromise its functionality if the starting matrix is poor: 1) the REFINE parameters are correlated to one another and 2) the radius of convergence is relatively small. To minimize the correlation problem, only a small number of parameters were refined simultaneously, keeping all others fixed. For example, in the cases when TAU2 was known to be off, only this parameter was allowed to change in the early stages of

Table 6.3: Refinement Parameters of MADNES

REFINE Parameters	Description
A*, B*, C*	Reciprocal cell dimensions
ALP*, BET*, GAM*	Reciprocal cell angles
TAU1, TAU2, TAU3	Rotations of the detector plane
TAU1	Rotation around the positive Z axis
TAU2	Horizontal swing angle
TAU3	Veritcal swing angle (= Tilt angle)
EPS	Effective mosaic spread of the crystal
CCX, CCY	Detector translations along x and y
DIST	Detector to crystal distance
WAVL	Radiation wavelength
MU1, MU2	Inclination angles between the primary beam and positive Z axis
	(0.0 for all experiments)
PHIX, PHIY, PHIZ	Missetting angles of the crystal

refinement, with all parameters set to reasonable values. As the orientation matrix improved, additional parameters were included in the refinement. REFINE will not accept reflections whose positions are poorly predicted by the matrix; when the starting matrix was particularly poor, REFINE would reject almost all reflections making refinement impossible. The tolerance for acceptance is controlled by the REFINE parameter LIMITS, which sets the tolerance for differences between predicted and found reflection positions. If these tolerances were set high enough (>5 pixels for spot location, >5 degrees for phi rotation), REFINE would accept reflections predicted by poor, but essentially correct, matrices, thus enlarging the radius of convergence such that a poor auto-indexing solution could be refined into a good orientation matrix.

Two quantities used to monitor the progress of refinement were the difference in predicted and observed spot positions and the calculated standard deviations in the crystal missetting angles (PHIX, PHIY, PHIZ). RMSCHK, a user-defined variable, was employed to define upper limits for the RMSD between a spot position and its predicted value (RMSCFD, upper limit = 1.2) and a reciprocal lattice vector and the predicted one (RMSDEG, upper limit = 0.5). Over the course of processing all of the neutron data, RMSCFD varied from 0.33 to 0.69 and RMSDEG from 0.03 to 0.11; the lower a RMSD value, the more accurate the orientation matrix. The standard deviations of the crystal missetting angles are very sensitive indicators of the quality of the matrix. In all cases, well refined matricies had standard deviations on the order of 0.02°.

Once a good matrix was obtained, it was used as the starting matrix to process the entire data set using the subroutine COLLECT. During data collection, the matrix was updated by refinement every time 30 (low resolution settings) or 20 (medium or high reolution settings) good reflections had accumulated. For most of the neutron data collection, only DIST, TAU1, EPS (crystal mosaisity), and the missetting angles (PHIX, PHIY, and PHIZ) were allowed to change during the matrix updates.

6.2.4 Data Collection

Normally, the COLLECT routine is the last data processing step in MADNES. This subroutine uses a continuously updated orientation matrix to index reflections and integrate intensities. Fig. 6.2a and Fig. 6.2b show the COLLECT parameters used in the processing of dsn100 as an example. Many of the parameters are irrelevent when processing raw data frames off-line: these include TIME, NOSYNC, DARK, WAIT, SUB, EXPTIM, BORD, SMOOTH, BKTYPE, and DUMP. NUFTYP is set to (0, 11) so that no nonuniformity correction is made to the data. This is consistent with the response of the neutron detector, which is uniform except for pixels that are dead. RECENT (2,-30) forces matrix refinement to occur every time 30 GOOD reflections have accumulated. The SIZE of the reflections are limited to 17 x 15 pixels. The WIDTH parameters indicate that the reflections fit into this allotted space. The choice of EVAL 5 forces the reflections to be profile fitted and the background to be fitted to a plane and subtracted from the data. Just as in FIND, COLLECT assigns each reflection an error code (Table 6.2), these flags are used in later programs to accept and reject reflections.

Since all of the data were collected from a single crystal, an orientation matrix obtained for low resolution data can, in theory, be used to index high resolution data. This is particularly useful if a high resolution data setting contains few strong reflections. If accurate values are known for the physical parameters of the high resolution detector position, the low resolution matrix can be used to COLLECT reflections from the weak data set, which can then be used to determine a new, and more accurate matrix for the high resolution data. Occasionally even the best matrix determined using reflections from the FIND and REFINE subroutines is very poor. In these cases, reflections out of COLLECT can be introduced back into REFINE and an improved matrix obtained.

COLLECT SAVE FILE
(DSN8.SAV)

```
SCAN          PHI
ROT           0.000          99.000
INC           0.100
TIME          180.00
NOSYNC
DARK          0              0
TEMPLATE      dsn8_2_001.2a;###
FILNUM        2
NUFTYP        0              11
EVAL          5              0.600          0.300          1.000
RECENT        2              -30
SIZE          17             15
BATCH         10.000
PAD           3              0
WAIT          750.00
SUB           0.000
EXPTIM        148.405
BORD          2              2              6
ERRLIM
              3.00           3.00           0.40
              !Max allowed Yms, Zms, Rot displacements
              5.0            100.0
              !Max allowed Background, Background SD
              5
              !Max allowed Bogus pixels
              3.00
              !Max allowed Rot width ratio
              31             31
              !Max allowed spot dimensions in pixels
SMOOTH        5
MASK          30             3              15             15
BKTYPE        0              0
DUMP          0
SDFACT        1.000
```

Fig. 6.2a: An example of COLLECT parameters

COLLECT SAVE FILE
(DSN8.SAV)

OFFSET	1	0.000	0.000	0.000
OFFSET	2	-0.002	-0.060	0.009
OFFSET	3	-0.046	-0.046	0.010
OFFSET	4	-0.057	-0.018	0.009
OFFSET	5	0.008	-0.006	0.000
OFFSET	6	0.027	-0.015	0.011
OFFSET	7	-0.023	-0.028	0.025
OFFSET	8	-0.050	-0.032	0.016
OFFSET	9	0.000	0.000	0.000
OFFSET	10	0.053	-0.006	0.021
OFFSET	11	0.029	-0.021	0.005
OFFSET	12	-0.044	0.001	0.014
OFFSET	13	0.000	0.000	0.000
OFFSET	14	0.000	0.000	0.000
OFFSET	15	0.028	0.000	0.007
OFFSET	16	0.000	0.000	0.000
WIDTH	1	12.967	10.538	1.000
WIDTH	2	13.212	10.006	1.018
WIDTH	3	12.729	10.873	0.977
WIDTH	4	12.827	10.517	0.993
WIDTH	5	13.017	10.749	1.000
WIDTH	6	12.826	11.005	0.977
WIDTH	7	12.666	10.180	0.977
WIDTH	8	12.488	10.257	0.995
WIDTH	9	13.204	10.955	1.000
WIDTH	10	13.250	11.175	0.988
WIDTH	11	12.725	10.814	0.993
WIDTH	12	12.034	10.446	1.009
WIDTH	13	13.007	10.825	1.000
WIDTH	14	13.229	10.119	1.000
WIDTH	15	12.650	10.885	1.004
WIDTH	16	12.673	10.656	1.000
EXIT				

Fig. 6.2b: An example of COLLECT parameters

6.3 Processing Individual Settings

The strategy designed to collect a complete neutron data set involved multiple overlapping detector positions, each capturing of a portion of reciprocal space, maximizing the number of unique reflections collected. This section describes the parameters and results of processing the data from each detector setting independently, including the data collected to compensate for crystal misalignment.

6.3.1 Orientation Matrices

The data from each detector position were processed in the manner described in section 6.2. Tables 6.4a and 6.4b summarize the parameters used for each detector position. Table 6.4a gives the observed (refined) values for TAU1, TAU2, and TAU3 as well as the target values for TAU2 and TAU3. In every case, there is always at least a small discrepancy between the actual and predicted values; however, in some cases (dsn250n, dsn250t, dsn250z, dsn375n for TAU2, dsn250t for TAU3) there is a large difference between the target value and the actual parameter used during data collection. For TAU2, these errors, all of which are approximately 8°, arose because of a malfunction of the parameter encoders of the physical experimental setup. Although the correct values were entered via the software and then verified by the digital encoder readout, the 2Θ swing arm was positioned incorrectly. This problem was not detected until it became obvious that it was impossible to process the data with the expected TAU2 values. In the case of the TAU3 discrepancy (dsn250t), the detector was neither tilted nor raised on the tower by mistake, as can be seen from the refined DIST value (at the zero position, the crystal-to-detector position is 623 mm). The actual values for TAU2 and TAU3 values were determined by the procedure described in Section 6.2.2. Although several of the refined detector parameters are different from what was

Table 6.4a: Matrix Refinement Statistics

Data Set	tau1 actual	tau2 actual	tau2 target	tau3 actual	tau3 target	dist actual	dist target	height	ccx	ccy
dsn8	-0.06	-8.09	-8.00	0.01	0.00	621.97	623.00	0.00	0.09	-0.94
dsn100	1.31	-8.99	-9.12	8.98	9.12	631.05	631.00	100.00	-0.99	-1.02
dsn20	5.79	-17.00	-18.09	18.06	17.79	649.87	654.00	200.00	-2.38	1.53
dsn20_end	5.44	-17.12	-18.09	18.01	17.79	654.71	654.00	200.00	-0.99	-1.03
dsn23	-0.01	-22.99	-23.00	0.01	0.00	621.88	623.00	0.00	-0.17	-0.96
dsn25	3.24	-7.62	-8.24	21.57	21.86	665.96	671.00	250.00	0.04	-0.51
dsn250	8.08	-20.41	-22.00	22.71	21.86	669.15	671.00	250.00	0.42	-0.51
dsn23n	3.66	-22.52	-23.00	9.60	9.10	630.12	631.00	100.00	-2.53	-4.12
dsn23n_end	3.81	-22.53	-23.00	9.60	9.10	628.51	631.00	100.00	-2.51	-3.29
dsn250n	4.34	-10.02	-18.00	21.90	21.90	669.57	671.00	250.00	3.06	0.23
dsn250j	6.13	-15.89	-18.00	21.90	21.90	668.58	671.00	250.00	-2.40	0.79
dsn250t	-0.11	-10.93	-18.00	0.00	21.90	621.26	671.00	250.00	-1.88	0.04
dsn250z	4.24	-9.83	-18.00	21.97	21.90	670.40	671.00	250.00	4.03	0.28
dsn10	6.12	-33.27	-33.53	10.80	9.12	627.14	631.00	100.00	-1.40	-3.18
dsn275	12.95	-27.23	-29.82	26.25	23.82	674.82	681.00	275.00	-2.00	-2.00
dsn275_end	12.87	-27.24	-29.82	26.23	23.82	678.93	681.00	275.00	-2.00	-2.00
dsn375	8.32	-13.73	-15.85	31.16	31.04	720.83	727.00	375.00	-1.02	-2.99
dsn375_end	8.56	-13.75	-15.85	31.20	31.04	720.88	727.00	375.00	-1.04	-3.00
dsn400	12.49	-18.89	-22.65	33.74	32.70	735.48	740.00	400.00	-1.74	1.87
dsn375n	6.83	-10.88	-18.90	31.00	31.00	724.42	727.00	375.00	-1.24	-1.38

Table 6.4b: Matrix Refinement Statistics

Data Set	misset phix	phix stdev	misset phiy	phiy stdev	misset phiyz	phiz stdev	cell	rms mm	rms deg	eps
dsn8	78.39	0.00	28.57	0.01	76.36	0.01	48.07/63.5	0.46	0.07	0.71
dsn100	65.54	0.00	13.02	0.01	67.96	0.00	48.13/63.5	0.44	0.04	0.65
dsn20	65.56	0.00	12.96	0.01	68.14	0.01	48.02/63.5	0.44	0.03	0.71
dsn20_end	65.70	0.00	13.20	0.01	68.22	0.01	48.39/63.5	0.41	0.05	0.79
dsn23	65.88	0.00	12.17	0.01	68.52	0.01	48.18/63.5	0.38	0.03	0.79
dsn25	65.57	0.00	12.98	0.01	68.12	0.00	48.00/63.5	0.50	0.05	0.50
dsn250	65.70	0.00	13.22	0.02	68.45	0.01	48.04/63.5	0.55	0.06	0.52
dsn23n	67.37	0.01	8.91	0.03	68.39	0.03	48.25/63.3	0.42	0.12	0.64
dsn23n_end	67.66	0.00	8.04	0.02	68.58	0.02	48.11/63.3	0.50	0.05	0.74
dsn250n	67.76	0.01	7.48	0.01	68.21	0.01	48.04/63.5	0.67	0.15	0.53
dsn250j	67.41	0.02	8.01	0.01	68.34	0.01	48.03/63.5	0.69	0.11	0.61
dsn250t	67.45	0.01	7.96	0.04	68.64	0.03	48.07/63.5	0.61	0.11	0.83
dsn250z	67.79	0.02	8.06	0.02	68.48	0.00	48.34/63.5	0.33	0.06	0.54
dsn10	65.56	0.00	12.98	0.01	68.23	0.01	48.01/63.3	0.51	0.04	0.72
dsn275	65.54	0.00	12.99	0.02	68.20	0.02	48.06/63.5	0.64	0.05	0.68
dsn275_end	65.64	0.00	12.94	0.01	68.10	0.02	48.00/63.5	0.64	0.06	0.75
dsn375	65.57	0.00	12.95	0.01	68.07	0.01	48.02/63.5	0.56	0.07	0.44
dsn375_end	65.76	0.01	12.66	0.01	68.03	0.01	48.00/63.5	0.48	0.06	0.44
dsn400	65.70	0.01	12.96	0.02	68.08	0.01	48.02/63.5	0.60	0.62	0.55
dsn375n	66.17	0.04	10.99	0.01	68.09	0.01	48.00/63.5	0.94	0.24	0.38

originally calculated to optimize data collection, the values in Table 6.4a and 6.4b accurately reflect the actual experimental conditions. The overall completeness will be compromised by these problems, but there is nothing inherently wrong with the data from these settings, and all of it was integrated into the final data set.

As can be seen in Table 6.4a, the refinement parameters CCX and CCY, representing small offsets from the predicted position of the direct beam, were used to compensate for inaccuracies in the overall processing parameters. Because the entire neutron data collection system is so large, the precision of the experimental parameters is diminished (for example the accuracy with which the detector can be placed at a specific height in millimeters is not very precise). In addition, typical neutron reflections are large and often irregularly shaped, making it difficult to define their centers precisely. In combination, these errors compromise the accuracy of the orientation matrix. During matrix refinement, CCX and CCY were allowed to vary, effectively compensating for some of these errors and improving the orientation matrix.

As described in section 6.2.3, two RMSDs were used to assess the quality of each matrix. The cutoff values for RMSCFD and RMSDEG were 1.2 and 0.5, respectively. The RSMD values for these neutron data differ significantly from typical X-ray values (RMSCFD < 0.150, RMSDEG < 0.010); once again this can be attributed to the dimensions and hence inaccuracies of the entire experimental system. Among other correlations, the RMSCFD variable is dependent on the crystal-to-detector distance, causing a weak correlation with the height of the detector; the value is lower, hence better, when the detector position is lower on the tower (Table 6.4b). The precision of the missetting angles is quantified by the standard deviation reported for each value. In all cases, PHIX, PHIY, and PHIZ are highly precise with a range of standard deviation from 0.000 to 0.040.

6.3.2 Compensation for Crystal Misalignment

As described in Chapter 4, after the first data set was collected, the perdeuterated *Staphylococcal* nuclease crystal slipped in the capillary. Because all of the detector positions have been processed using a standard set of missetting angles, the change in both PHIX (78.39 vs. 65.64) and PHIY (28.57 vs. 12.92) between the first data set (DSN8) and subsequent data sets is indicative of a change in crystal orientation The crystal misalignment problem was discovered after the original detector positions had been merged together, resulting in a data set for which the completeness was only 58% complete to 2.2Å (61% complete to 2.4Å). The region of reciprocal space that was missing from this data set was determined by plotting all the collected reflections on a three dimensional grid (Fig. 6.3). The strategy for acquiring the missing region involved minimizing the number of additional detector settings and maximizing the collection of missing reflections by optimizing the crystal rotation for each data set. MADNES, along with information about the crystal missetting angles and the expected detector parameters, was used to predict reflections and optimize the outcome of the second series of data collection.

A further change in PHIX (65.64 vs. 67.37) and PHIY (12.92 vs. 8.49) is observed between the original detector positions and the sets collected to correct for the crystal misalignment (dsn23n, dsn250n, dsn250j, dsn250t, dsn250z, dsn375n). This change was caused by the removal of the goniometer between the first and second sets of measurements. In this case, it is the capillary, rather than the crystal, that has been slightly tilted in the beam.

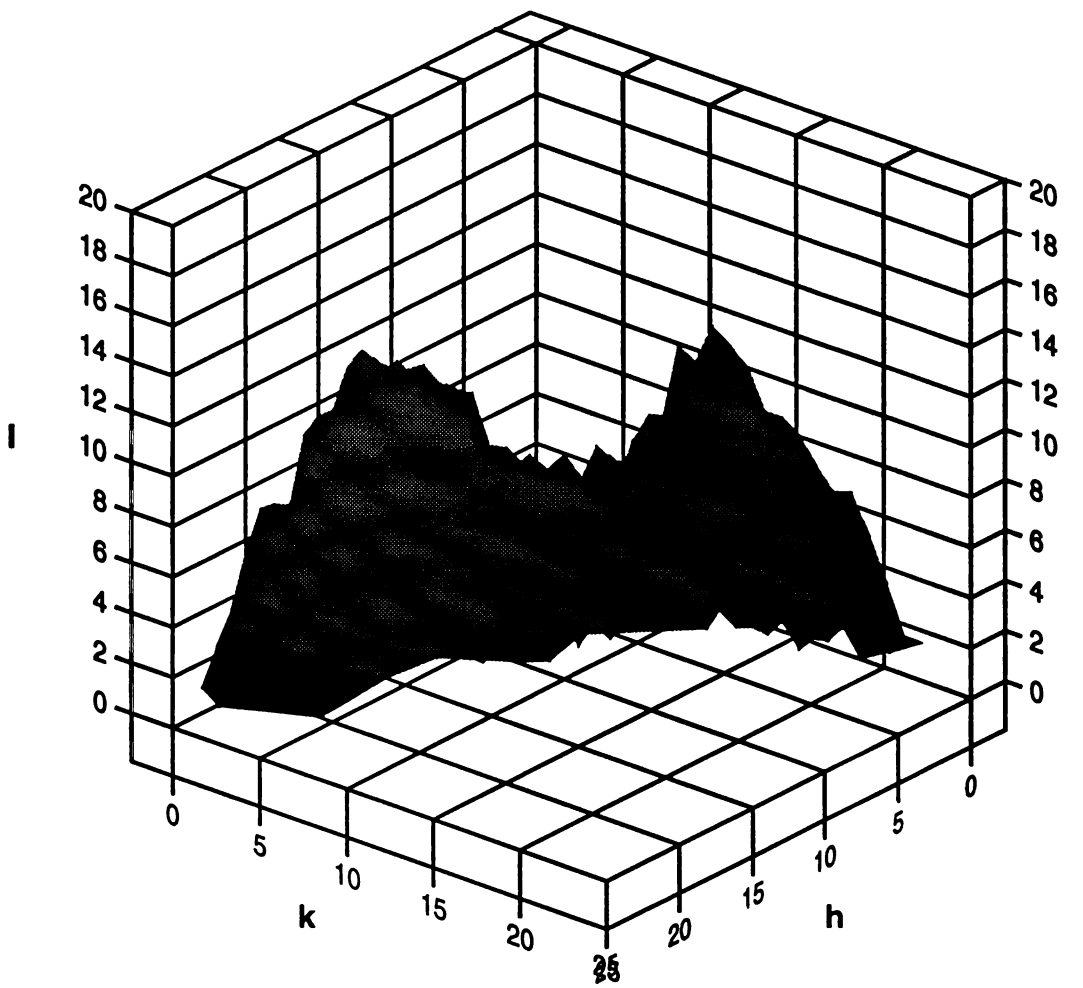


Fig. 6.3: Plot of all collected reflections from original detector settings

6.3.3 Results of Data Collection

After processing, the output of MADNES (h,k,l,F,σF, and error codes) is fed into the program SYMMAV [3]. SYMMAV performs three functions: 1) based on error codes, reflections are retained or rejected, 2) the input data are reduced to the standard asymmetric unit, and 3) symmetry related observations are compared to their weighted average; if any observation deviates more than REJ1% from the average, it is rejected. In addition, if REJ2% of the observations are rejected, the entire reflection is rejected. Any reflection possessing an error code of 1024, 2048, 4096, 8192, or 16384 is automatically rejected. SYMMAV allows the user to reject reflections based on other error codes; reflections with flags indicating on edge, too far and too big errors were rejected from this data. Although a cutoff for weak reflections can be set in SYMMAV, all weak reflections were accepted. Weak reflections are treated with a more generous cutoff than strong reflections. SYMMAV computes an average F value for all reflections with error code 1 (WEAK), any reflection with F less than this value will be rejected if it differs from its symmetry mates by this average F value multiplied by REJ1. The rejection criteria for the symmetry related observations were set at 45-50% (REJ1) and 41% (REJ2). Table 6.5 summarizes the data processing results. The first column (Observations Read) represents the total number of observations collected in MADNES, the next four columns contain information about the rejection and acceptance of redundant and unique data in SYMMAV. When a data set contains redundant information, SYMMAV will calculate an R_{sym} for the data. Because detector positions were chosen to maximize the collection of unique data, an R_{sym} could not be calculated for every data set. The general trend for the R_{sym} values is as expected (increase with resolution) and given the low I/σI for the data (3.5 - 16.6), indicates data of acceptable quality. The low number of redundant observations in dsn100 and dsn23n_end compromised the use of their R_{sym} values for data quality assessment.

Table 6.5: Processing Results

Data Set	Observations Read	Observations Accepted on Error Code	Accepted Observations	Total Reflections	Accepted Reflections	Rsym %	Observations Used In Calculation
dsn8	520	513	511	291	290	8.00	415
dsn100	564	558	558	552	552	17.16	12
dsn20	1026	889	889	889	889	0.00	0
dsn20_end	203	167	167	166	166	0.29	2
dsn23	826	599	595	440	438	9.62	314
dsn25	505	467	467	467	467	0.00	0
dsn250	887	784	784	784	784	0.00	0
dsn23n	62	60	60	60	60	0.00	0
dsn23n_end	1539	1516	1512	1498	1496	20.13	32
dsn250n	560	538	538	538	538	0.00	0
dsn250j	983	729	715	512	512	13.78	385
dsn250t	313	309	309	274	274	12.81	70
dsn250z	192	189	188	115	115	10.25	119
dsn10	2343	2136	2106	1852	1837	19.15	538
dsn275	868	800	800	800	800	0.00	0
dsn275_end	504	466	466	465	465	43.13	2
dsn375	611	587	587	587	587	0.00	0
dsn375_end	148	146	146	146	146	0.00	0
dsn400	493	484	484	484	484	0.00	0
dsn375n	699	555	555	555	555	0.00	0

6.4 Merging Individual Settings

As the data from the individual detector settings became available, they were merged into a combined data set. This preliminary merging was done in order to monitor the quality and completeness of the data. When data collection was complete, all data were scaled to calculated neutron data and then merged to form the final data set.

6.4.1 Preliminary Merging

As each individual detector setting was successfully processed, the reflections were merged into a combined data set. This step involved the use of three programs: SYMMAV [3], PROTEIN [4], and SCALESYM [5]. SYMMAV takes unsorted, redundant observations and produces a sorted, unique data set while providing the user with statistics about the data. This program was used throughout data processing to preliminary merge the reflections collected from each detector position. A list of the unique data was generated for the multiple detector settings along with an overall R_{merge} value. PROTEIN, which is a suite of programs written to work with protein crystallography data as well as the resulting structures, was used to calculate a scale factor for each of the multiple detector positions and to compute the completeness of the emerging data set. SCALESYM applies the scale factors generated by PROTEIN to averaged SYMMAV output of each individual setting. SYMMAV was then used to merge the scaled data together.

6.4.2 Local Scaling using Calculated Data

It was decided to scale the observed neutron data to calculated neutron data because the amount of overlap between any one detector position and any other is limited

by the physical setup of data collection. The refined X-ray structure of perdeuterated *Staphylococcal* nuclease, complete with bound inhibitor and deuterons at every proton position, was the model used to calculate structure factors (F_{calc}). No water molecules were included in this model.

Before scaling observed data to calculated data, an overall B-factor correction must be applied to the calculated data. This is done to take into account the lattice imperfections and/or the amount of disorder in the particular crystal from which data were collected. A modified "Wilson plot" calculation was made taking into account the fact that there is no atomic fall-off with resolution in neutron data. Because the B-factor generated for the neutron data (25.9Å^2) was not significantly different from the average B-factor (23.3Å^2) for the refined X-ray structure of perdeuterated *Staphylococcal* nuclease, no B-factor correction was made to the calculated data.

To correct for anisotropic diffraction, slowly varying local scale factors were applied to each setting, scaling every reflection with a scale factor based on its neighbors. This was done using a program called LOCSCL [6], which used a grid of 11 x 11 x 5 lattice points and a minimum of 100 neighboring reflections to determine the local scale factor to apply to each reflection. If a reflection had less than 100 neighbors, it was scaled with an overall scale factor. The overall scale factor for each setting, as well as the overall and local R_{scale} values are listed in Table 6.6. Because all of the data sets were preliminarily scaled as soon as they were collected, the overall scale factor is similar for most detector settings. One apparent exception is dsn250t; in reality, this detector setting was not collected in its target mid-resolution position, but rather at a lower resolution (see section 6.3.1) with the same monitor count as other low resolution settings (8000). Very little change was seen between overall and local R_{scale} values; however, this was to be expected. The size of the detector is small in comparison to the overall size of the Ewald's sphere such that the effect of anisotropic diffraction will not be great in any individual setting.

Table 6.6: Overall Scale Factors and Rscale Values

Data Set	Overall Scale Factor	Rscale overall	Rscale local
dsn8	61.2	22.9	22.7
dsn100	50.0	16.4	16.0
dsn20	36.4	17.7	16.7
dsn20_end	35.5	14.9	14.8
dsn23	35.8	18.5	17.3
dsn25	37.8	14.9	14.5
dsn250	31.9	16.8	16.5
dsn23n	39.4	12.3	12.3
dsn23n_end	39.5	16.1	15.4
dsn250n	39.7	14.5	14.2
dsn250j	32.8	15.7	15.1
dsn250t	54.6	18.8	18.7
dsn250z	33.0	13.7	13.7
dsn10	28.3	19.4	19.2
dsn275	29.3	19.2	19.0
dsn275_end	28.4	18.0	17.9
dsn375	31.0	18.1	18.2
dsn375_end	32.5	20.3	20.2
dsn400	29.4	19.5	19.4
dsn375n	32.9	19.2	19.0

6.5 Final Data Set

The final neutron data set used to solve the structure of perdeuterated *Staphylococcal* nuclease was the accumulation of data from 20 individual data sets, stemming from 16 unique detector positions. These merged data contained reflections to 2.2Å resolution; however, only data to 2.4Å had average I/σI greater than 3.5 (see Fig. 6.4). Based on this low signal-to-noise at higher resolution, only reflections to 2.4Å were used in the final data set. Normally the best signal-to-noise would be associated with low-resolution data; however, the signal-to-noise of these data was best at mid-resolution (~4Å), and fell off with increasing resolution. This phenomenon stems from the fact that the mid-resolution detector settings were exposed to more neutrons (8000 vs. 16000 monitor counts) than the low-resolution settings. These high counting rates overcompensated for the normal decrease in signal-to-noise expected with increased resolution. The overall R_{merge} for the data was 16.4% with an R_{merge} (on I) of 21.1% for the 2.5Å to 2.4Å resolution shell. Table 6.7 shows the number of overlapping reflections each data set had with neighboring detector settings; these overlapping reflections were used in the R_{merge} calculation. A total of 10911 observations, corresponding to 5765 unique reflections, yields an overall completeness to 2.4Å of 83% with 64% completeness for the highest resolution shell (2.5Å to 2.4Å resolution). 4358 (76%) of the unique reflections were greater than 2σ on I, with 270 between 2.5Å and 2.4Å resolution. These numbers were a significant improvement over the completeness of the data obtained using the original 8 (one of which could not be processed) detector positions (61% complete to 2.4Å, see Section 6.3.2). Fig. 6.5 shows the SYMMAV output for the final data set, and Fig. 6.6 shows the variation of the completeness with resolution. With no further alterations, these data were put directly into XPLOR for structure refinement.

Fig. 6.4: I/σ vs Resolution

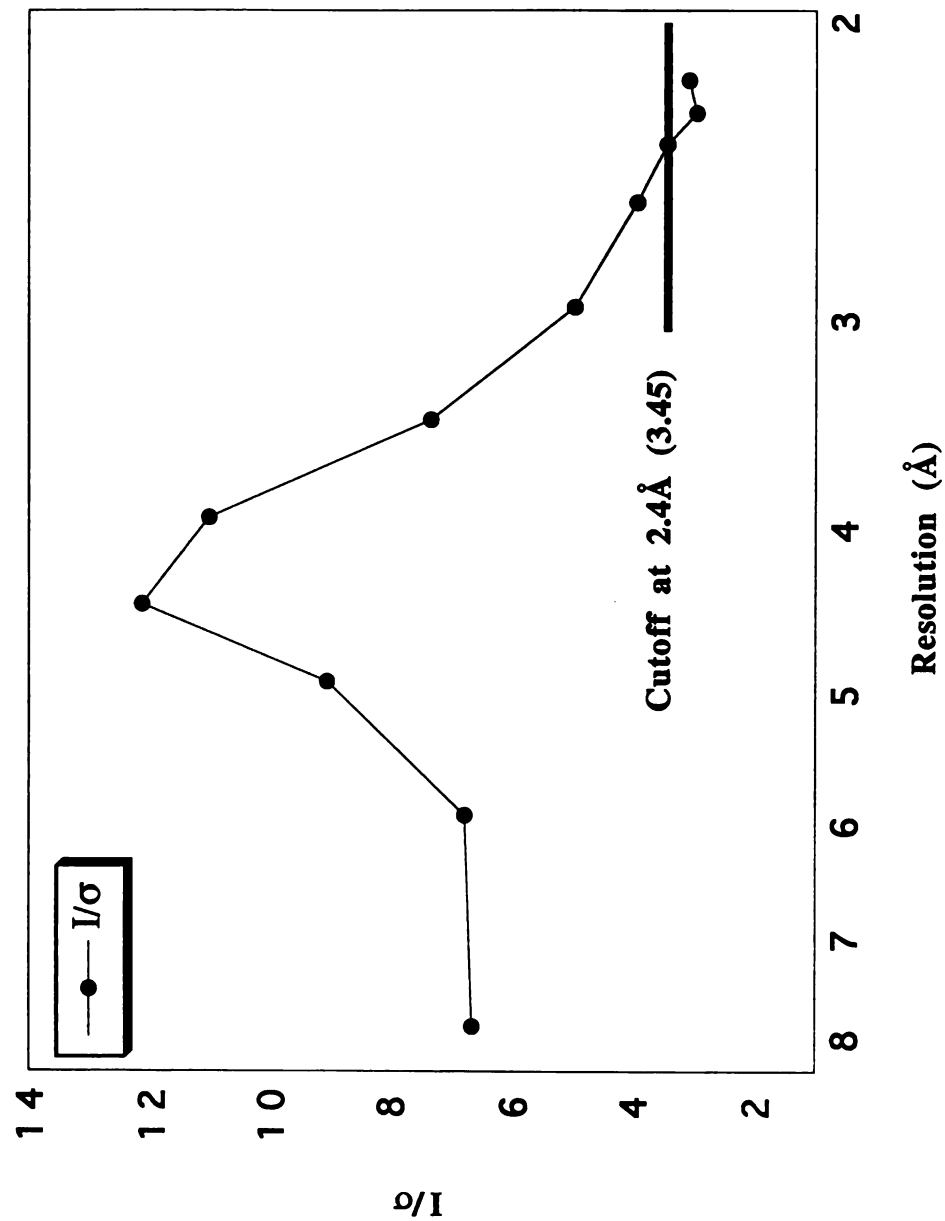


Table 6.7: Overlapping Reflections From Each Detector Position

Date Set	Total # Reflections	Overlapping Reflections
dsn8	290	191
dsn100	552	365
dsn20	889	631
dsn20_end	166	146
ds23	438	288
ds25	467	385
dsn250	784	588
dsn23n	60	54
dsn23n_end	1496	1070
dsn250n	538	422
dsn250j	512	206
dsn250t	274	203
dsn250z	115	105
dsn10	1837	712
dsn275	800	474
dsn275_end	465	297
dsn375	587	373
dsn375_end	146	100
dsn400	484	295
dsn375n	555	341

```
P R O G R A M   S Y M M A V


REDUCE A "FAST" DATA SET
VERSION P-03-NOV-1988


A FULL DATA REDUCTION RUN INVOLVES FOUR STEPS :

1. TRANSLATE THE OBSERVATIONS TO THE ASYMMETRIC UNIT AND PRODUCE A SORTED
OUTPUT FILE

2. REJECT OBSERVATIONS ON THE BASIS OF ERROR CODE AND/OR PHI RANGE

3. CALCULATE AND APPLY PHI-DEPENDENT SCALE FACTORS

4. REJECT "BAD" OBSERVATIONS AND AVERAGE ACCEPTED ONES

STEPS 2. AND 3. ARE OPTIONAL. THE BINARY FILES PRODUCED BY EACH OF THE FOUR
STEPS CAN BE READ BY EACH OF STEPS 2., 3. AND 4.


USER : GAMBLE
DATE : 13-FEB-95
TIME : 16:18:44
```

Fig. 6.5a: SYMMAV output for the final data set

SUBROUTINE REDUCE

CURRENT SET-UP :

Laue group :	4 (4/m)
Input file name :	temp.loc (1 files)
Output file :	REDUCED.BIN
Read format :	(3i4,2f8.2,10x,i1,4f2.1)
Cell parameters :	48.000 48.000 63.500
	90.000 90.000 90.000

8 EQUIVALENT REFLECTIONS :

1 :	+H	+K	+L
2 :	-H	-K	+L
3 :	-K	+H	+L
4 :	+K	-H	+L

... and Friedel mates

11455 OBS. READ FROM VERSION 1 (11455 GOOD)
AND SCALED WITH K = 1.00000, B = 0.00

REFLECTION COUNT :

PHI RANGE (PHI_MIN, PHI_MAX)	0.00	0.00
TOTAL NUMBER OF OBSERVATIONS	11455	
NUMBER WITH F.LE.0 (RESET TO .5 SIG)	0	
NUMBER OF "GOOD" OBSERVATIONS	11455	
NUMBER OF "WEAK" OBSERVATIONS	0	(AV. F = 0.00)
NUMBER OF OBSERVATIONS "ON THE EDGE"	0	
NUMBER OF OBSERVATIONS "TOO FAR OFF"	0	
NUMBER OF OBSERVATIONS "TOO BIG"	0	

COUNT OF ERROR CODES :

IERR:	0	1	2	4	8	16	32	64	128	256
NOBS :	11455	0	0	0	0	0	0	0	0	0

IERR:	512	1024	2048	4096	8192	16384
NOBS :	0	0	0	0	0	0

>>>> 11455 OBSERVATIONS WRITTEN TO FILE REDUCED.BIN <<<<

Fig. 6.5b: SYMMAV output for the final data set

SUBROUTINE AVERAGE

CURRENT SET-UP :

Rejection for obs,refs:	0.45	0.41	
Weak refls have F .LT.	200.000		
Average Friedels :	Y		
Input file :	REDUCED.BIN		
Output file :	all_madness_loc_rsym.avg		
Laue group nr.	4 :4/m		
Cell parameters :	48.00	48.00	63.50
	90.00	90.00	90.00
#refls to print :	-10		

REFLECTION COUNT

TOTAL NUMBER OF OBSERVATIONS	11455
TOTAL NUMBER OF REFLECTIONS	5866
NUMBER OF OBSERVATIONS ACCEPTED	10911
NUMBER OF REFLECTIONS ACCEPTED	5765
TOTAL NUMBER OF WEAK REFLS ACCEPTED	2736
TOTAL NUMBER ACCEPTED WITH I.LT.2*SIG	1407
NUMBER OF ACENTRIC REFLECTIONS	5486
NUMBER OF CENTRIC REFLECTIONS	279
NUMBER OF FRIEDEL PAIRS	0

R-FACTORS (ALL ON I)

	ALL	F+	F-	ANO
R (%)	16.43	16.47	0.00	0.00
NOBS	8595	8482	0	0

VARIATION OF RSYM WITH FSQ

FSQ	2500	5000	10000	20000	40000	80000	160000	320000	640000	999999
R (%)	62.05	58.03	47.72	40.88	34.20	18.30	14.86	14.06	13.74	13.30
NOBS	18	80	320	1153	2162	1659	1617	947	465	174
NREFS	80	105	294	823	1434	1141	993	522	274	99
I > 2*SIG	0	0	16	225	1123	1115	984	522	274	99
<FSQ/SIG>	0.22	0.46	0.99	1.72	2.81	4.64	7.34	10.95	16.62	27.18
REJECTED	0	0	0	0	1	67	27	6	0	0

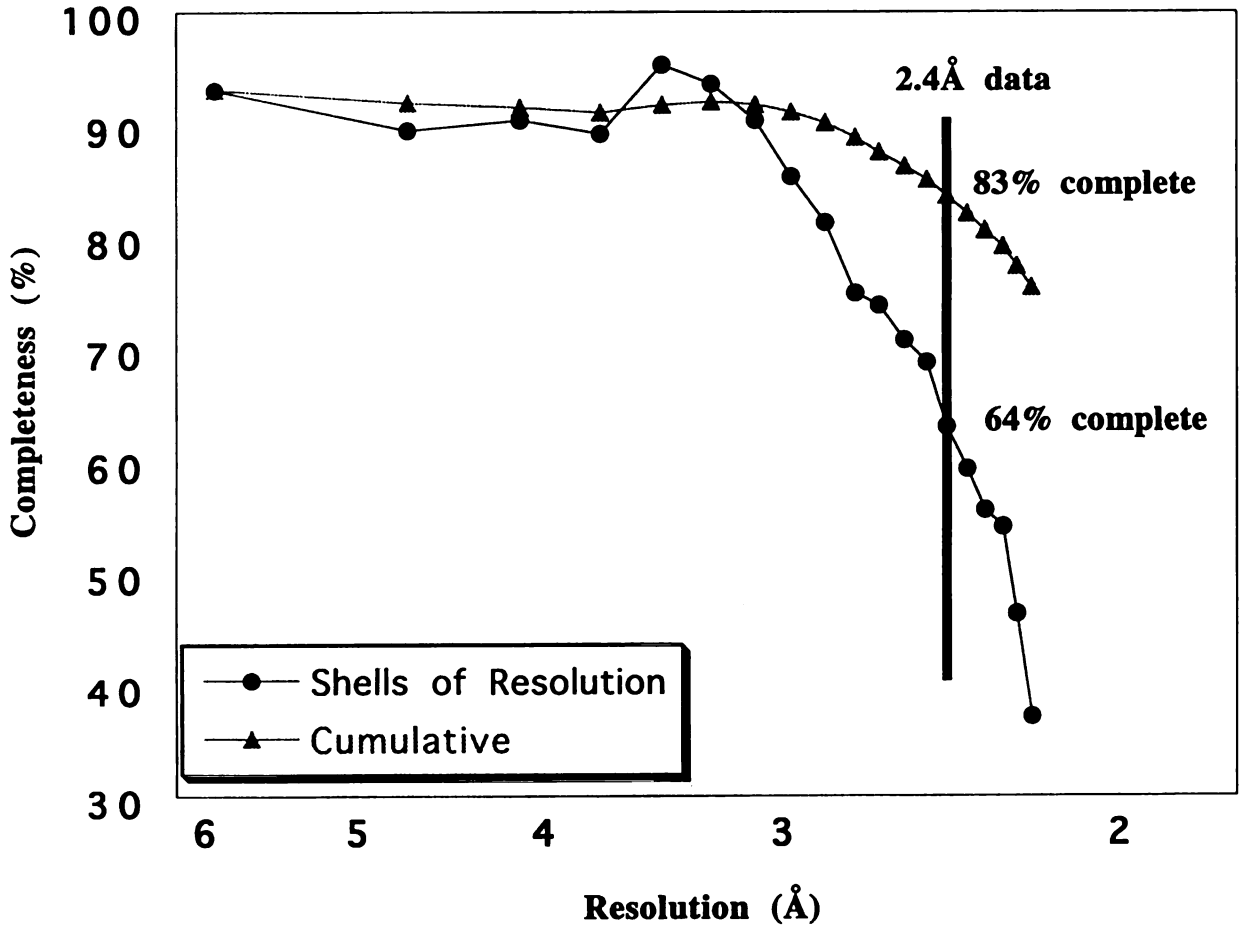
VARIATION OF RSYM WITH RESOLUTION

ANG.	10.00	8.00	6.00	5.00	3.50	3.00	2.60	2.40	2.30	2.20
R (%)	12.31	11.15	16.38	15.60	14.56	18.09	22.37	21.08	22.22	27.29
NOBS	112	92	237	380	1954	2039	2022	1049	431	279
NREFS	80	68	202	256	1079	990	1232	796	459	603
I > 2*SIG	76	64	174	241	969	802	875	527	273	357
<FSQ/SIG>	16.59	6.67	6.79	9.06	9.35	4.98	3.95	3.46	2.97	3.10
REJECTED	1	4	2	3	24	23	20	7	5	12

>>>> 5765 REFLECTIONS WRITTEN TO FILE all_madness_loc_rsym.avg <<<<

Fig. 6.5c: SYMMAV output for the final data set

Fig 6.6: Completeness vs Resolution (All neutron data)



6.6 References

1. Pflugrath, J. and A. Messerschmidt, *Munich Area Detector NE System (MADNES)*. 1987, Cold Spring Harbor Laboratory.

2. *XTAL2D*. Brookhaven National Laboratory, Upton.

3. de Vos, A.M., *SYMMAV*. 1988, South San Francisco.

4. Steigemann, W., *PROTEIN: A Package of Crystallographic Programs for Analysis of Proteins*. 1982, Marinsreid, West Germany.

5. Gamble, T.R., *SCALESYM*. 1994, South San Francisco.

6. de Vos, A.M., *LOCSCL*. 1990, South San Francisco.

CHAPTER 7. Structure Refinement

Once the data were successfully collected and processed, the next step was to determine the neutron structure of perdeuterated Snase. Generally, the procedure of neutron structure refinement parallels X-ray refinement; however, the presence of hydrogen and deuterium atoms in the neutron case causes both fundamental and practical differences. This chapter describes the inherent differences between X-ray and neutron refinement, the adaptations made to existing X-ray refinement programs for use with neutrons, the refinement process itself, and an assessment of the quality of the final structure.

7.1 Crystallographic Refinement

Crystallographic refinement is the accepted method for improvement of the model structure. This method uses least-squares procedures to minimize the differences between observed (F_O) and calculated (F_C) structure factor amplitudes by adjusting the model [1]. The parameters adjusted in this process include both atomic coordinates and temperature factors. A problem that arises during protein structure refinement is the poor observation-to-parameter ratio. This ratio, which is the ratio between the number of unique reflections and the number of independent parameters, should be high (> 10) for robust refinement [2]. However, in a typical protein crystal structure determination at 2Å to 3Å resolution, the ratio of the number of unique reflections to independent parameters is only 0.5 to 1.0, too low for successful least squares refinement. This problem is circumvented by introducing stereochemical information in the form of restraints between atoms, thus decreasing the number of parameters. These restraints force bond lengths and angles to be close to target values based on idealized geometry. Instead of simply minimizing the difference between observed and calculated data, an "energy function"

(Equation 7.1), which includes appropriately weighted restraints, is the minimized function.

Equation 7.1: $$E_{total} = w_a\Sigma E_{crystallographic} + w_b\Sigma E_{bond\ distances} + w_c\Sigma E_{bond\ angles} + w_d\Sigma E_{torsion\ angles} + w_e\Sigma E_{non\text{-}bonded\ contacts} + w_f\Sigma E_{planar\ groups} + w_g\Sigma E_{chiral\ volumes}$$

(where w is the weight each restraint is given during refinement)

In the refinement of the neutron structure of perdeuterated Snase, the combination of only moderately high resolution and the increased number of parameters due to the inclusion of hydrogen and deuterium atoms, made the observation to parameter ratio even worse than in typical X-ray protein refinement. The number of atoms contained in a neutron Snase model was approximately double that of an X-ray model (1187 H/D, 1135 non-hydrogen atoms), while the number of unique reflections collected between 10Å and 2.4Å was 4625 (out of a possible 5606). Therefore, the observation to parameter (x,y,z,B x number of atoms) ratio for the neutron refinement of Snase would only be 0.50, compared to an observation to parameter ratio of 1.02 in the equivalent X-ray case. In order to minimize this problem as much as possible, geometry parameters involving hydrogen/deuterium atoms were very tightly restrained during refinement.

The resolution of the data created another problem affecting the ability to identify deuterium atoms in the structure. At 2.4Å resolution, the distance that two atoms can be resolved from each other (two peaks in $2F_O$-F_C electron density, as opposed to one) will be approximately 1.8Å [3]. While this distance is only slightly longer than most bonds in the structure, it is almost double the length of any bond involving deuterium. This will make it particularly difficult to distinguish deuterium atoms from their neighboring nitrogen atoms on the peptide backbone. In order to assess the effect of resolution and data incompleteness, test calculations were done. Fig. 7.1 shows calculated neutron

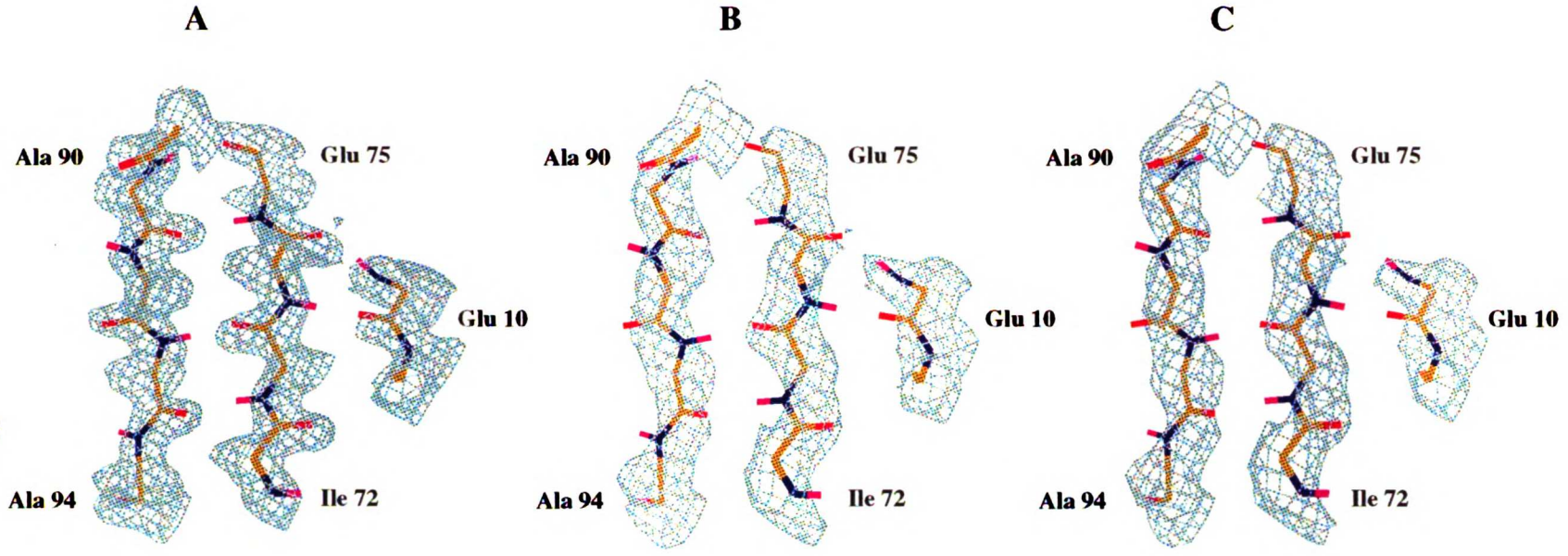


Fig. 7.1: Calculated data using final model and all data at 1.8A (A) and 2.4A (B) , and at 2.4A (C) , using only data that matched the actual reflections collected. Residue 10 was determined to have a deuterium atom at the H/D site, residues 73-75, 90-91, and 94 to have hydrogen atoms, and residues 72 and 92-93 were undetermined. In the final data, carbonyl oxygens often lacked density; these calculated maps show that this problem arises mainly from the resolution.

density at 1.8Å and 2.4Å resolution using all data, and at 2.4Å resolution using only reflections that match the actual data. As can be seen in the figure, these problems significantly decrease the quality of the neutron density, thus decreasing the amount of visible detail that can be derived from the maps. Difference Fourier maps have been partially successful in circumventing this problem.

7.2 XPLOR and O Modifications

XPLOR uses the combination of crystallographic diffraction data and molecular dynamics to refine a molecular structure; empirical, geometric and effective energy terms describing the experimental data are minimized by conjugate gradient descent, simulated annealing, and conformational search procedures [4]. The method of conjugate gradient minimization uses derivatives of the energy function to determine a direction of decreasing energy. Because the total energy is a function of the difference between observed (F_O) and calculated (F_C) structure factor amplitudes, it is independent of the sign of the scattering factors. This feature allows the method to be used without modification for both X-ray and neutron data. In determining the neutron structure of perdeuterated Snase, XPLOR was used for positional and both overall and individual B-factor refinement.

As is the case with the program MADNES, XPLOR and O [5] are software packages originally designed to be used to determine protein structure via X-ray crystallography. Both programs were modified for use with neutron diffraction data structural information. Alterations to XPLOR took into account atomic scattering length, including the difference in sign between hydrogen and deuterium, and an atom type for deuterium was defined in the program's reference files. O, which was used for manual adjustment to the protein, was changed to include all hydrogen atom positions.

XPLOR requires four types of input files to characterize the model; topology reference files, parameter reference files, a molecular (or protein) structure file (psf), and a protein data bank (pdb) file. Changes were made to the references files so that together the pdb and psf files could accurately describe the protein's structure complete with hydrogen and deuterium atoms.

The topology file contains information about the amino acid residues and ligands of the structure, including atom descriptions, covalent bonds, and bond angles (simple three atom, dihedral and improper). The standard topology file used for X-ray diffraction data, TOPHCSDX.PRO, contains no non-labile hydrogen atom information for use in refinement. Therefore, two changes were required for neutron structure refinement. Hydrogen positions were built into all residue definitions using the file TOPALLHDG.PRO (normally used for NMR structure calculations) as a guide, and a mass was added for chemical atom type D so that both hydrogen and deuterium atoms could be built into the structure. Although it appears counter intuitive, the chemical atom type D was used to represents hydrogen atoms (mass = 1.008) and the chemical atom type H to represent deuterium atoms (mass = 2.016); this was done for practical reasons. In the final neutron structure of perdeuterated Snase, the vast majority of hydrogen atom positions will be occupied with deuterium atoms due to biosynthetic replacement (all non-labile sites) or rapid exchange with deuterated solvent (hydroxyl rotors, most amide groups); therefore most positions will contain deuterium atoms, with hydrogen being the exception. Because XPLOR has many functions that recognize hydrogen atoms, for example the HBUILD function that generates appropriate hydrogen atom coordinates, by simply changing the mass of chemical atom type H to deuterium's mass, the program made deuterium the default value for any hydrogen position. Hydrogen atoms, with chemical atom type D, were then treated as the exception to the default value. As an example, Fig. 7.2 shows what changes were made to the topology statement for threonine in order to incorporate hydrogen atoms.

```
RESIdue THR
    GROUp
        ATOM    N     TYPE=NH1    CHARge=-0.35        END
        ATOM    HN    TYPE=H      CHARge= 0.25 END
        ATOM    CA    TYPE=CH1E   CHARge= 0.10 END
        ATOM    HA    TYPE=HA     CHARge= 0.10 END
        ATOM    CB    TYPE=CH1E   CHARge= 0.25 END
        ATOM    HB    TYPE=HA     CHARge= 0.10 END
        ATOM    OG1   TYPE=OH1    CHARge=-0.65        END
        ATOM    HG1   TYPE=H      CHARge= 0.40 END
        ATOM    CG2   TYPE=CH3E   CHARge= 0.00 END
        ATOM    HG21  TYPE=HA     CHARge= 0.10 END
        ATOM    HG22  TYPE=HA     CHARge= 0.10 END
        ATOM    HG23  TYPE=HA     CHARge= 0.10 END
        ATOM    C     TYPE=C      CHARge= 0.55 END
        ATOM    O     TYPE=O      CHARge= 0.55 END

        BOND    N     CA
        BOND    CA    C
        BOND    C     O
        BOND    N     HN
        BOND    CA    CB
        BOND    CB    OG1
        BOND    CB    CG2
        BOND    OG1   HG1
        BOND    CA    HA
        BOND    CB    HB
        BOND    CG2   HG21
        BOND    CG2   HG22
        BOND    CG2   HG23

        DIHEdral  N     CA    CB    OG1
        DIHEdral  CA    CB    OG1   HG1
        DIHEdral  CA    CB    CG2   HG21

        IMPRoper  CA    N     C     CB
        IMPRoper  CB    OG1   CG2   CA
        IMPRoper  HA    N     C     CB
        IMPRoper  HB    CA    OG1   CG2
        IMPRoper  HG21  HG22  CB    HG23

        DONOr     HN    N
        DONOr     HG1   OG1
        ACCEptor  OG1   " "
        ACCEptor  O     C
```

Fig. 7.2: Topology statement for threonine including hydrogen atoms

The parameter file specifies various constants for the conformational and nonbonded energy terms. In the file PARHCSDX.PRO, these parameters are based on the work of Engh and Huber [6], who derived bond lengths, bond angles, and force constants from a compilation of small molecule crystallographic structures contained in the Cambridge Structure Database. Fig. 7.2 shows the additional bond and angle constants added to the parameter file to describe exchangeable amide hydrogen atoms, as well as hydrogen atoms bound to side chain nitrogen atoms (for ARG, ASN, GLN, HIS, LYS, TRP). The force constants involving hydrogen and deuterium atoms for bonds, and simple, dihedral, and improper angles were increased by ten fold over the default values contained in PARALLHDG.PRO (normally used for NMR calculations). Separate topology and parameter files were used to describe the D_2O associated with the structure (DWATER.TOP and DWATER.PARM) and the bound inhibitor (TOPALLHDG.PDTP and PARALLHDG.PDTP). The TIPS3P model was used to create the topology and parameter files for D_2O. The inhibitor pdTp (deoxythymidine 3',5'-bisphosphate) was built based on the topology and parameters files TOPALLHDG.DNA and PARALLHDG.DNA, respectively. The bond length and angle constants necessary to characterize the non-labile hydrogen atoms associated with the inhibitor are denoted with an "*" in Fig. 7.3.

XPLOR approximates atomic scattering factors with a resolution-dependent expression consisting of the summation of four Gaussian functions and a constant [4]. The parameters for this equation are defined in the SCATTER statement, which is part of XPLOR's subroutine XREFINE. Because neutrons are scattered by an atom's nucleus, which is very small compared to the neutron wavelength, there is essentially no angular dependence in the scattering factors of neutrons [3]. Therefore, the equation describing a neutron scattering factor is independent of resolution and simplifies to a single constant. Fig. 7.4 shows the contrast between the atomic scattering factors normally used in XPLOR for x-ray diffraction, and the values used for neutron refinement.

BOND	D	NH1			405.0		0.98	
BOND	D	NH2			405.0		0.98	
BOND	D	NH3			405.0		1.04	
BOND	D	NC2			405.0		1.00	
BOND	D	C2			1000.0		1.09*	
BOND	D	C3			1000.0		1.09*	
BOND	D	CH			1000.0		1.09*	
BOND	D	CF			1000.0		1.09*	
ANGLE	D	C2	D		500.0		109.47*	
ANGLE	D	C3	D		500.0		109.47*	
ANGLE	D	CH	D		500.0		109.47*	
ANGLE	D	CH	C3		500.0		109.47*	
ANGLE	D	CH	CH		500.0		109.47*	
ANGLE	D	CH	C2		500.0		109.47*	
ANGLE	D	C2	C3		500.0		109.47*	
ANGLE	D	C2	OS		500.0		109.47*	
ANGLE	D	C3	CS		500.0		109.47*	
ANGLE	CH	C2	D		500.0		109.47*	
ANGLE	D	CH	OS		500.0		109.47*	
ANGLE	D	CH	NS		500.0		109.47*	
ANGLE	D	CF	NS		500.0		120.0*	
ANGLE	D	CF	CS		500.0		120.0*	
ANGLE	C	NH1	D		30.0		120.0	
ANGLE	CW	NH1	D		30.0		120.0	
ANGLE	C	NH2	D		30.0		120.0	
ANGLE	D	NH2	D		40.0		125.0	
ANGLE	C5	NH1	D		30.0		120.0	
ANGLE	CRH	NH1	D		35.0		120.0	
ANGLE	CR1E	NH1	D		35.0		120.0	
ANGLE	CH1E	NH1	D		35.0		120.0	
ANGLE	CH2E	NH1	D		35.0		120.0	
ANGLE	CH2G	NH1	D		35.0		120.0	
ANGLE	CH2E	NH3	D		35.0		109.5	
ANGLE	D	NH3	D		40.0		109.5	
ANGLE	C	NC2	D		35.0		120.0	
ANGLE	D	NC2	D		40.0		120.0	
DIHE	D	NH1	C	CH1E	500.0	0	0.0	
DIHE	D	NH1	C	CH2G	250.0	0	0.0	
IMPR	D	CH	C2	OS	500.0	0	65.0*	
IMPR	D	C2	CH	OS	500.0	0	65.0*	
IMPR	D	C2	OS	NS	500.0	0	-65.0*	
IMPR	D	NS	CS	CF	500.0	0	0.00*	
IMPR	NH1	C	CH1E	D	135.0	0	0.0	
IMPR	NH1	C	CH2G	D	135.0	0	0.0	
IMPR	NH1	C5	CRH	D	135.0	0	0.0	
IMPR	NH1	CR1E	CW	D	135.0	0	0.0	
IMPR	NH2	D	D	C	135.0	0	0.0	
IMPR	NH1	CH2E	C	D	135.0	0	0.0	
IMPR	D	D	CH2E	D	500.0	0	-66.0	
NONBONDED	D			0.0498	2.2272	0.0498	2.2272	
NBFIX	D	NH1			44.2	1.0	44.2	1.0

* Definitions necessary for the non-labile hydrogen atoms found in the inhibitor pdTp. All others specify parameters for backbone amides or exchangeable hydrogen atoms bound to side chain nitrogen atoms.

Fig. 7.3: Additional parameters for hydrogen atoms

X-ray

```
SCATTER     (CHEMICAL C*)
    2.3100  20.8439  1.0200  10.2075  1.5886  0.5687  0.8650  51.6512  0.2156

SCATTER     (CHEMICAL N*)
   12.2126   0.0057  3.1322   9.8933  2.0125 28.9975  1.1663   0.5826 -11.5290

SCATTER     (CHEMICAL O*)
    3.0485  13.2771  2.2868   5.7011  1.5463  0.3239  0.8670  32.9089  0.2508

SCATTER     (CHEMICAL S*)
    6.9053   1.4679  5.2034  22.2151  1.4379  0.2536  1.5863  56.1720  0.8669

SCATTER     (CHEMICAL P*)
    6.4345   1.9067  4.1791  27.1570  1.7800  0.5260  1.4908  68.1645  1.1149

SCATTER     (CHEMICAL ALC*)     !calcium
   15.6348  -0.0074  7.9518   0.6089  8.4372 10.3116  0.8537  25.9905 -14.8750
```

Neutron

```
SCATTER     (CHEMICAL C*)
    6.6500   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL N*)
    9.4000   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL O*)
    5.8000   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL S*)
    2.8000   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL P*)
    5.1000   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL ALC)      !calcium
    4.7000   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL H*)       !deuterium
    6.6700   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000

SCATTER     (CHEMICAL D*)       !hydrogen
   -3.7400   0.0000  0.0000   0.0000  0.0000  0.0000  0.0000   0.0000  0.0000
```

$$f_i(\vec{h}) = \sum_{k=1}^{4} a_{ki} exp(-b_{ki}(\mathcal{F}^* \vec{h})^2/4) + c_i + \mathrm{i}d_i$$

Fig. 7.4: X-ray and neutron atomic form factors and XPLOR equation

O was modified to incorporate the display of hydrogen atoms in the structure by including hydrogen atoms into the residue definitions. In addition, changes were made to the dihedral angle definitions so that rotation around the O-H of hydroxyl rotors could be made.

7.3 The Starting Model

The X-ray structure of perdeuterated Snase [7] was used as the starting model for refinement of the neutron structure. This structure had been refined to an R-value of 19.3% (15Å - 1.9Å data) with good stereochemistry and included the protein (residues 7-141), the inhibitor (pdTp), and one calcium ion. A total of 1108 non-hydrogen atoms were included in the model. All exchangeable hydrogen atoms were given zero occupancy so that exchangeable sites would be characterized in an unbiased fashion during the progression of the refinement. Table 7.1 shows the amino acid types that contain exchangeable hydrogen atoms associated with their side chains.

7.4 Refinement Methodology

XPLOR was used for positional and temperature factor refinement. Figs. 7.5 and 7.6 show the files used as XPLOR input for these functions.

7.4.1 Positional Refinement

Positional refinement is the process by which the overall energy of the structure is minimized by applying shifts to atomic positions that give the lowest total energy [8]. Fig. 7.5 shows the input used for positional refinement. PRINT THRESHOLD statements are found at the beginning and end of the procedure to give the user an idea of

Table 7.1: Exchangeable Hydrogen Sites

RESIDUE	EXCHANGEABLE HYDROGEN SITES
ARG	HE, HH11, HH12, HH21, HH22
ASN	HD21, HD22
GLN	HE21, HE22
HIS	HD1, HE2
LYS	HZ1, HZ2, HZ3
SER	HG
THR	HG1
TYR	HH
TRP	HE1
PTP	H3
(inhibitor)	

```
remarks file positional_hdg.inp

topology @tophcsdx_hdg_neu.pro  end          { contains H/D info }
topology @topallhdg.pdtp        end          { information about inhibitor }
topology @dwater.top            end          { information about D2O }
parameter @parhcsdx_hdg_neu.pro              { contains H/D info }
        nbonds
        atom cdie shift eps=1.0  e14fac=0.4  {nonbonded interaction energy}
        cutnb=7.5 ctonnb=6.0 ctofnb=6.5      {options. Note the reduced }
        nbxmod=5 vswitch                     {nonbonding cutoff to save }
        end                                      {some CPU time }
end
 parameter @parallhdg.pdtp          { information for the inhibitor}
        nbonds
        atom cdie shift eps=1.0  e14fac=0.4  {nonbonded interaction energy}
        cutnb=7.5 ctonnb=6.0 ctofnb=6.5      {options. Note the reduced }
        nbxmod=5 vswitch                     {nonbonding cutoff to save }
        end                                      {some CPU time }
end
parameter @dwater.parm              { information for the water D2O}
        nbonds
        atom cdie shift eps=1.0  e14fac=0.4  {nonbonded interaction energy}
        cutnb=7.5 ctonnb=6.0 ctofnb=6.5      {options. Note the reduced }
        nbxmod=5 vswitch                     {nonbonding cutoff to save }
        end                                      {some CPU time }
end
{===>} structure @gen33.psf     end          { read structure file }

coor disposition=reference selection = (all) @xray.pdb end
vector do (harmonic = 0.0) (all)
@xray2.sel
end
{===>} coor @bref_hdg14.pdb                  { read coordinates }

print threshold=0.03 bonds
print threshold=10. angles
print threshold=60.  dihe
print threshold=5.0  impr

constraints fix=(name ALC) end

vector do ( charge=0.0 ) ( resname LYS and
        ( name ce or name nz or name hz* ) )     { Turn off charges on LYS}
vector do ( charge=0.0 ) ( resname GLU and
        ( name cg or name cd or name oe* ) )     { Turn off charges on GLU}
 vector do ( charge=0.0 ) ( resname ASP and
        ( name cb or name cg or name od* ) )     { Turn off charges on ASP}
 vector do ( charge=0.0 ) ( resname ARG and
        ( name cd or name *e or name cz or name nh* or name hh* ) )
 vector do ( charge=0.0 ) ( resname ALC  and ( name ALC ) )
```

Fig. 7.5a: XPLOR positional refinement input

```
flags
        include pele pvdw xref
        ?
end

xrefine
        @xdata.inp
        nreflections=15000
        reflection @allmadnes_loc.toxplor end          { read reflections }
        resolution  10.0 2.4                           { resolution range  }
        reduce
        fwind=0.1=100000
        method=FFT
                fft
                memory=1000000
                end
        tolerance=0.00 lookup=false  { this makes the minimizer happy }
        wa=60000                     {reduced to give X-ray structure more weight}
end

minimize powell                              { Invoke the powell minimizer }
        nstep= 400                           { Do 400 cycles of minimization}
        drop=10.0                            {Initial drop in energy }
end                                          { Minimization will be executed  }
xrefin
        resolution-limits=10.0 2.4     { Resolution for maps }
        update-fcalc                   { Compute FCALCS }
        do scale ( FCALC = FOBS )      { FCALCS will be scaled to FOBS  }
end

constraints interaction (not hydrogen) (hydrogen) end

print threshold=0.02 bonds
print threshold=5.0 angles
print threshold=20.  dihe
print threshold=1.0  impr

constraints interaction (not hydrogen) (not hydrogen) end

print threshold=0.03 bonds
print threshold=10. angles
print threshold=60.  dihe
print threshold=2.0  impr

write coordinates output=pos39.pdb  end      { Write final coordinates }

coor disposition=comp selection = (all) @xray.pdb end
vector ident (store9) (all)
coor fit selection = (store9) end
coor rms end
stop
```

Fig. 7.5b: XPLOR positional refinement input

the geometric improvement made to the structure in any given run. This evaluation was done with and without hydrogen atoms so that a comparison to typical X-ray values could be made. Throughout refinement, the model was very tightly restrained to ideal geometry, especially for the hydrogen atoms, as reflected by the lowered wa value in Fig. 7.5. In addition, the neutron model was restrained to the well-ordered atoms ($B < 30Å^2$) of the X-ray structure with a harmonic restraint value of 20 kcal $mole^{-1}$ $Å^2$. The calcium ion has a small scattering length (4.7 fermi), which makes it difficult to locate this atom in a 2.4Å neutron map; in response to this problem, calcium was fixed in the same relative atomic position as found in the 1.65Å resolution X-ray structure solved by Loll and Lattman [9]. The charges associated with arginine, aspartic acid, glutamic acid, lysine, and the calcium ion were turned off because the dielectrics are not properly described in XPLOR due to the lack of explicit solvent. Positional refinement was alternated with temperature factor refinement, manual manipulation of the model using O [5], and the addition of hydrogen and deuterium atoms at exchangeable proton sites, as well as water molecules.

7.4.2 Temperature Factor Refinement

XPLOR optimizes overall B-factors via a conjugate gradient minimization and applies the shift to the temperature factors in the pdb file. For individual, isotropic B-factors, the program again uses Powell conjugate gradient minimization, these individual B-factors are restrained to be similar to those of neighboring atoms. ASIGMA specifies the target standard deviation between B-factors of selected atoms involved in angles, and BSIGMA specifies the target standard deviation between B-factors of selected bonded atoms. In essence, ASIGMA and BSIGMA restrict the deviation in B-factors between neighboring atoms. Fig. 7.6 is an XPLOR input file containing commands for both overall and individual B-factor refinement. Overall B-factor refinement was done using

```
remarks file xtalrefine/bindref_hdg.inp
remarks B-factor refinement/restrained, individual or overall

topology      @tophcsdx_hdg_neu.pro end        { contains H/D info }
topology      @topallhdg.pdtp      end         { information about inhibitor }
parameter     @parhcsdx_hdg_neu.pro            { contains H/D info }
nbonds
atom cdie shift eps=1.0  e14fac=0.4           {nonbonded interaction energy}
      cutnb=7.5 ctonnb=6.0 ctofnb=6.5         {options.  Note the reduced }
      nbxmod=5 vswitch                        { nonbonding cutoff to save }
      end                                     { some CPU time }
end
parameter  @parallhdg.pdtp                    { information for the inhibitor}
atom cdie shift eps=1.0  e14fac=0.4           {nonbonded interaction energy}
      cutnb=7.5 ctonnb=6.0 ctofnb=6.5         {options.  Note the reduced }
      nbxmod=5 vswitch                        { nonbonding cutoff to save }
      end                                     { some CPU time }
end
{===>} structure @gen33.psf         end       { read structure file }

coor disposition=reference selection = (all) @xray.pdb end
vector do (harmonic = 0.0) (all)
@xray2.sel
end

{===>} coor @gen39.pdb                        { read coordinates }

!vector do ( b=24.0 ) (all)                   {sets single overall b-factor}

xrefine
@xdata.inp                                    { crystallographic data}
      nreflections=20000
      reflection @allmadnes_loc.toxplor   end { read reflections }
      resolution 10.0 2.4                     { resolution range  }
      reduce
      fwind=0.1=100000
      method=FFT
            fft
            memory=400000
            end
       tolerance=0.0 lookup=false             { this makes the minimizer happy }
      wa=60000                                { reduced to give X-ray values more weight }
end
```

Fig. 7.6a: XPLOR B-Factor refinement input

```
! xrefin
! update
!optimize overall                    { optimize overall B-factor      }
!       drop=0.01                    { this avoids "over-shooting" of the minimum }
!       nstep=10
!       end
!end

+{ The following two VECTOR statements make gaussian distributions around current }
+{ factors (stored in array B) with width 1.5 for backbone and 2.0 for sidechains  }
+vector do (b=b + gauss(1.0))
+       ( name ca or name n or name c or name hn or name ha)
+vector do (b=b + gauss(1.5))
+       ( not ( name ca or name n or name c or name hn or name ha))

*xrefin
*optimize bfactors                   { Invoke B-factor optimizer      }
*       nstep=40                     { 40 steps of conjugate gradient  }
*       drop=10.0                    { Expected initial drop in energy }
*       { By default, the program will automatically determine the relative }
*       { weight between restraints and crystallographic residual        }
*       {Target sigma for 1-2 B-factor pairs (for backbone and sidechain)}
*bsigma=( name ca or name n or name c or name o or name hn or name ha)=0.50
*bsigma=( not( name ca or name n or name c or name o or name hn or name ha))=0.75
*       { Target sigma for 1-3 (angle) B-factor pairs }
*asigma=( name ca or name n or name c or name o or name hn or name ha)=0.75
*asigma=( not( name ca or name n or name c or name o or name hn or name ha))=1.00
*       end                          { Refinement will be executed}
*end

write coordinates output=bref_hdg15.pdb end          {Write final coordinates}
stop
```

NOTE: All commands preceeded by an "!" are included for overall temperature factor refinement. All commands preceeded by an "*" are included for individual temperature factor refinement. The lines preceeded by an "+" were included only for the first round of individual temperature refinement.

Fig. 7.6b: XPLOR B-Factor refinement input

data to 2.4Å resolution, starting with an identical B-factor of 24.0Å^2 for all atoms. As refinement progressed, individual B-factor refinement with target standard deviations for bonds set to 0.5Å^2 (main chain) and 0.75Å^2 (side chain), and for angles set to 0.75Å^2 and 1.00Å^2.

7.5 Determining Hydrogen and Deuterium at Exchangeable Sites

The most direct method of determining whether an exchangeable hydrogen/deuterium (H/D) site is occupied by hydrogen or deuterium is to calculate a difference map (F_o-F_c) with the exchangeable sites set to zero occupancy. In the resulting map, a hydrogen atom will have a negative peak and a deuterium atom a positive one. However, the ability to distinguish an hydrogen from a deuterium atom is compromised by three factors. The first of these factors is the resolution of the data; each atom is associated with an individual peak that will broaden and overlap with its neighbors as the resolution decreases. If the overall data quality is poor, individual peaks will be further distorted. Second, the amount of disorder inherent to the region of the H/D atom will affect the quality of the difference density; an H/D atom bound to a well-ordered amide nitrogen will be easier to distinguish than one associated with a neighbor having a high temperature factor. Third, atoms left out of the phasing of a Fourier difference synthesis will appear at half their true height in the resulting maps [3].

These problems can be circumvented by enhancing the difference peaks for each exchangeable site. In an "enhanced" difference map calculation, a deuterium atom is placed at every amide H/D site in the model. In the resulting maps, if a deuterium atom occupies the site in the protein, no peak is seen; however, if a hydrogen atom occupies the site, the peak is negative with an enhanced amplitude due to the opposite signs of the scattering lengths (in effect (-3.74 fermi) - 6.67 fermi = -10.41 fermi). Because the protein is soaked in deuterated mother liquor prior to data collection, it is expected that

most exchangeable backbone amide hydrogen atoms will have been replaced with deuterium atoms. Thus, enhanced difference maps, with deuterium at every backbone amide H/D site will contain better phase information than difference maps with zero occupancy at these sites. Fig. 7.7 shows both regular and enhanced difference maps used to identify hydrogen atoms on the main chain. Not only are the difference peaks more sharply defined in the enhanced maps, but H/D sites that are undeterminable in the regular difference map are clearly defined. This procedure has a certain resemblance to the H_2O-D_2O solvent difference maps used in the analysis of the neutron structure of trypsin [10].

Both "regular" and "enhanced" difference maps were used to distinguish between hydrogen and deuterium atoms at backbone amide H/D sites. All exchangeable sites associated with side chains (see Table 7.1) were assumed to be deuterium atoms because side chain H/D atoms exchange much more quickly than amide H/D atoms and will not be protected from exchange during the long period of data collection [11].

7.6 The Inclusion of Water

All water molecules associated with the neutron structure of perdeuterated Snase will be D_2O because the crystal was soaked in a perdeuterated mother liquor prior to and throughout data collection. Because of the increased neutron scattering of deuterium compared to the X-ray scattering of hydrogen, water molecules should be more easily identifiable in neutron than in X-ray difference maps. Water molecules were introduced into the structure where they could by fit into at least 3.5σ difference density (F_o-F_c) with good hydrogen bonding geometry to the protein.

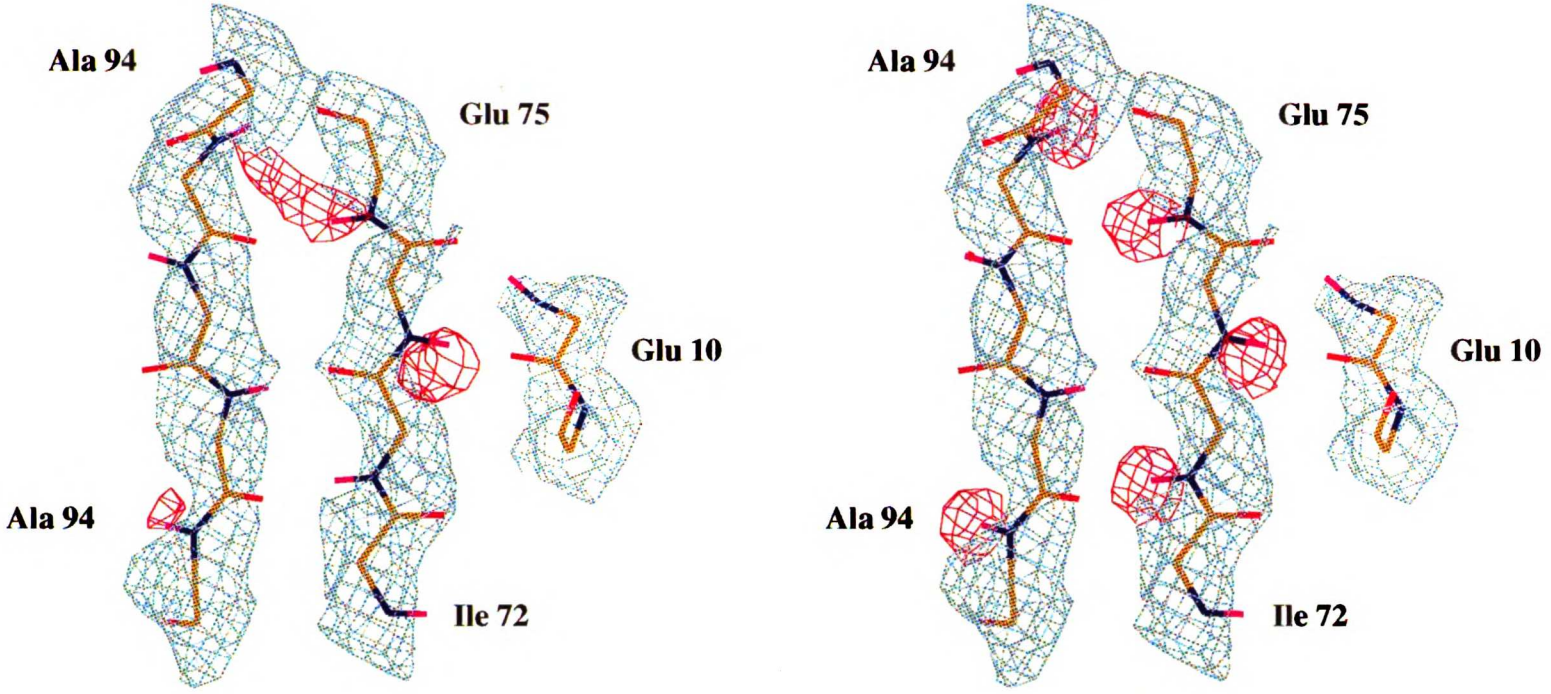


Fig. 7.7: Neutron density maps depicting hydrogen atoms. In the left figure, difference density (shown in red, contoured at -2 σ) indicates hydrogen atoms (shown, but not included, in the phase calculation). In the right figure, difference density (contoured at -3.5 σ) was generated with a deuterium atom at each amide H/D site. In both cases, Fo-Fc density is shown in red and 2Fo-Fc density (contoured at 1 σ) in grey.

7.7 Refinement Results

As described above, the 1.9Å X-ray structure of perdeuterated Snase including inhibitor (pdTp) and calcium, but no water, was used as a starting model for refinement. Temperature factors were included in the starting model. Deuterium atoms were included at exchangeable sites, but given occupancies of zero. Data between 10Å and 2.4Å were used for the refinement. Throughout refinement, only a crystallographic R-value was monitored. No cross validation R-value (R_{free}) was monitored because the starting model was already well refined and represents an accurate model of the protein; a R_{free} requires that about 10% of the data not be used during refinement [4], which would have further compromised the incompleteness of the data set.

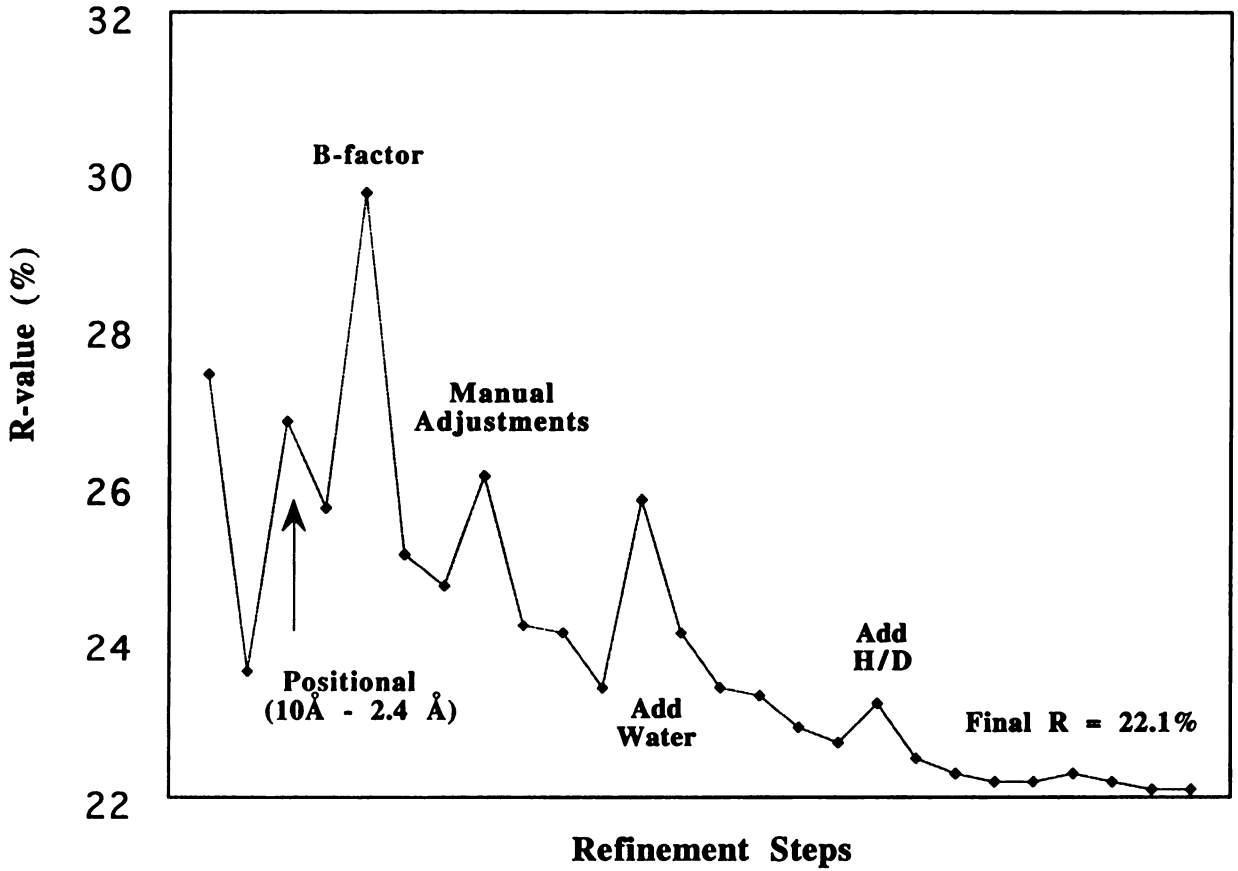
The model was refined to convergence by 25 cycles of XPLOR refinement alternated with manual rebuilding, determination of hydrogen and deuterium occupancy, and the addition of water molecules. A typical round of refinement consisted of conjugate gradient minimization (maximum number of cycles = 400, tolerance = 0.00), restrained individual temperature factor refinement (maximum number of cycles = 40), and more positional refinement (same as first step), at which point manual changes were made to the model and the process repeated.

The first round of positional refinement, using data to 3.0Å, brought the R-value down from 27.4% to 23.6%. All data to 2.4Å were then included in the calculations, bringing the R-value to 25.7%. All temperature factors were given an overall value of 24.0$Å^2$ and a round of overall temperature factor refinement was done. This process was followed by individual temperature factor refinement and more positional refinement, decreasing the R-value to 24.7%. F_O-F_C and $2F_O$-F_C neutron density maps were calculated and the entire model was inspected with respect to these maps. Changes were made to side chains to better fit the density and nine water molecules were added to the structure. The R-value of this adjusted model was 26.1%, and after positional and B-

factor refinement, the R-value was 23.4%. Five more water molecules were added and the exchangeable hydrogen atoms of arginine, asparagine, glutamine, lysine, and tryptophan were built in as deuterium atoms with occupancies of 1.0. Positional and B-factor refinement brought the R-value down to 23.3%.

At this point, 10 exchangeable amide hydrogen/deuterium (H/D) sites were identified as hydrogen and 9 as deuterium based on at least 2σ difference density. The occupancies of these atoms were set to 1.00. Positional and B-factor refinement brought the R-value to 22.7%. Enhanced difference maps with deuterium at every amide H/D site except previously identified as hydrogen were calculated. If no difference density was found at 2σ in either regular or enhanced difference maps, the H/D amide site was characterized as a deuterium atom. 50 deuterium atoms were added to the main chain in this fashion. In addition, two hydrogen atoms and one deuterium atoms were added into 2σ regular difference density, and two more water molecules were included. Positional and B-factor refinement brought the R-value to 22.2%. 16 hydrogen atoms were then built into 4σ enhanced difference density at H/D amide sites and 4 deuterium atoms into a sites with neither regular nor enhanced difference density. The last rounds of positional and B-factor refinement led to a final model having an R-value of 22.1%. Only water molecules with 1σ density in the final 2Fo-Fc maps, good hydrogen bonding geometry, and B-factors less than 50 $Å^2$ were included in the final structure. PROCHECK [13] analysis was performed. Fig. 7.8 shows the progress of refinement as monitored by R-value. Fig. 7.9 shows typical $2F_O$-F_C density. The deuterium atoms of the methyl, methylene, and epsilon amino groups are clearly oulined by the density. Had the protein not been perdeuterated, negative scattering from the non-labile hydrogen atoms would have severly limited the structural information obtained for the side-chains at 2.4Å.

Fig. 7.8: Progress of Refinement as Monitored by R-value



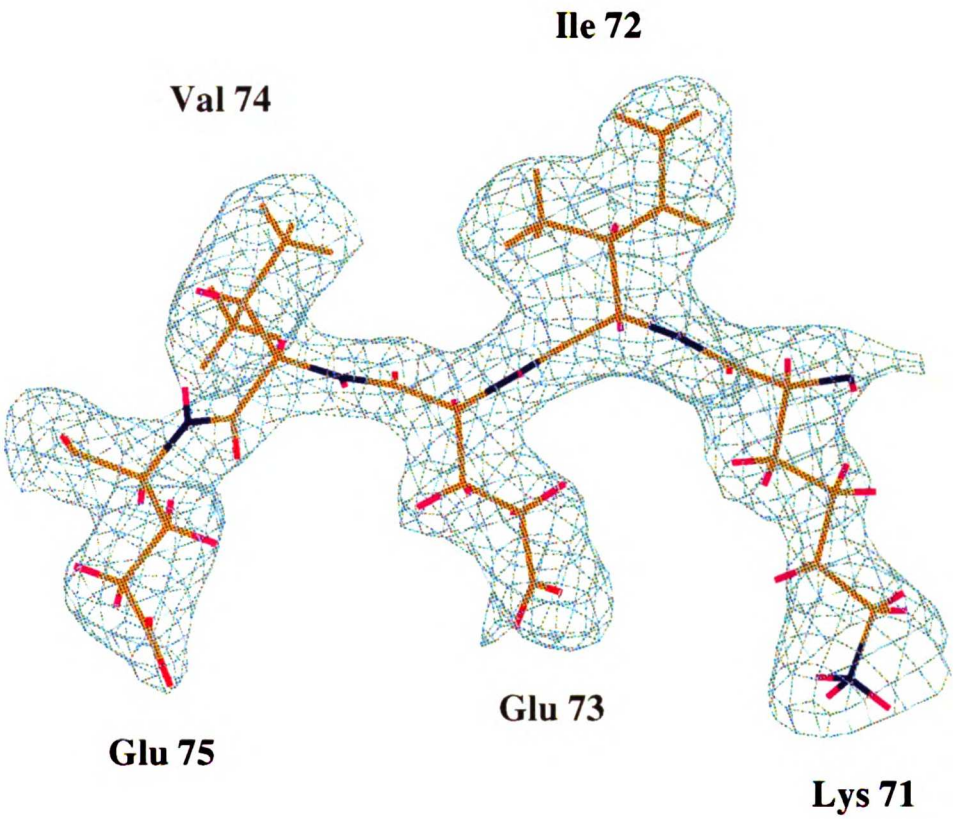


Fig 7.9: Region of the final neutron density map with the model for segment Lys71 - Glu75 superimposed. The density was calculated with coefficients 2Fo-Fc and contoured at 1σ.

7.8 The Final Structure

The structure of perdeuterated *Staphylococcal* nuclease at 2.4Å consists of one protein molecule (residues 7 - 141), one calcium ion, one inhibitor molecule (pdTp), and 13 water molecules. Table 7.2 shows the breakdown of non-hydrogen, hydrogen and deuterium atoms in the model. There are 128 exchangeable backbone proton positions of which 28 (22%) are occupied by hydrogen atoms, 64 (50%) by deuterium atoms, and 36 (28%) could not be determined (Table 7.3). In addition to the amide hydrogen atoms, 11 hydrogen atoms are associated with the inhibitor. All exchangeable H/D atom positions on side chains are occupied by deuterium atoms except for histidine (HD1/HE2), serine (HG), threonine (HG1) and tyrosine (HH). These exceptions were given occupancies of 0.00. As reported by XPLOR, the RMS deviation for the heavy atoms in the final structure for bonds is 0.010Å, for angles is 1.55°, for dihedral angles is 25.9°, and for improper angles is 0.94°. As expected due to tight restraints, the geometry statistics involving hydrogen/deuterium atoms are much better: 0.003Å, 0.66°, 3.4°, and 0.38°, respectively.

Stereochemical analysis of the structure was done using PROCHECK [13]. A Ramachandran plot shows that the majority of the residues (83%) occupy the most favored region of the plot, with most of the other residues (16%) falling into additionally allowed regions (Fig. 7.10). The only residue outside these two regions is Asn138. In the 1.65Å resolution X-ray structure of Snase and inhibitor, Loll and Lattman [9] also found Asn138 to have unfavorable Phi/Psi angles. They rationalize that the favorable interactions allowed by this conformation, mainly a hydrogen bond between the asparagine side chain and the carbonyl oxygen of residue 106, are sufficient to offset the unfavorable steric contribution. Other main chain parameters such as peptide bond planarity and alpha carbon tetrahedral distortion are better than average when compared to other structures at 2.4Å resolution (Fig. 7.11). The quality of the Ramachandran plot,

Table 7.2: Distribution of Atoms in Perdeuterated Snase

Atom Type	Protein	Inhibitor (pdTp)	Water	Calcium
Non-hydrogen	1082	25	13	1
Deuterium	1035	1	26	
Hydrogen	28	11		
Undetermined	36 backbone			
	22 side chain			

Table 7.3: Occupancy of Amide Proton Sites

Residue	Occupancy		Residue	Occupancy
7LEU	N-term		75GLU	H
8HIS	U		76PHE	D
9LYS	D		77ASN	U
10GLU	D		78LYS	U
11PRO	PRO		79GLY	D
12ALA	U		80GLN	D
13THR	D		81ARG	D
14LEU	D		82THR	D
15ILE	D		83ASP	H
16LYS	D		84LYS	D
17ALA	U		85TYR	H
18ILE	U		86GLY	D
19ASP	U		87ARG	H
20GLY	U		88GLY	U
21ASP	D		89LEU	H
22THR	H		90ALA	H
23VAL	U		91TYR	H
24LYS	U		92ILE	U
25LEU	D		93TYR	U
26MET	H		94ALA	H
27TYR	D		95ASP	D
28LYS	D		96GLY	D
29GLY	D		97LYS	U
30GLN	U		98MET	D
31PRO	PRO		99VAL	H
32MET	D		100ASN	H
33THR	U		101GLU	H
34PHE	H		102ALA	U
35ARG	H		103LEU	H
36LEU	H		104VAL	H
37LEU	H		105ARG	H
38LEU	D		106GLN	H
39VAL	H		107GLY	H
40ASP	D		108LEU	D
41THR	D		109ALA	U
42PRO	PRO		110LYS	U
43GLU	U		111VAL	D
44THR	U		112ALA	D
45LYS	D		113TYR	D
46HIS	D		114VAL	D
47PRO	PRO		115TYR	D
48LYS	D		116LYS	D
49LYS	D		117PRO	PRO
50GLY	D		118ASN	D
51VAL	D		119ASN	D
52GLU	D		120THR	D
53LYS	U		121HIS	D
54TYR	D		122GLU	D
55GLY	D		123GLN	D
56PRO	PRO		124HIS	D
57GLU	D		125LEU	U
58ALA	D		126ARG	D
59SER	D		127LYS	U
60ALA	D		128SER	D
61PHE	D		129GLU	U
62THR	U		130ALA	U
63LYS	U		131GLN	D
64LYS	U		132ALA	H
65MET	D		133LYS	U
66VAL	H		134LYS	D
67GLU	D		135GLU	U
68ASN	U		136LYS	D
69ALA	D		137LEU	U
70LYS	D		138ASN	U
71LYS	U		139ILE	D
72ILE	U		140TRP	H
73GLU	H		141SER	D
74VAL	H			

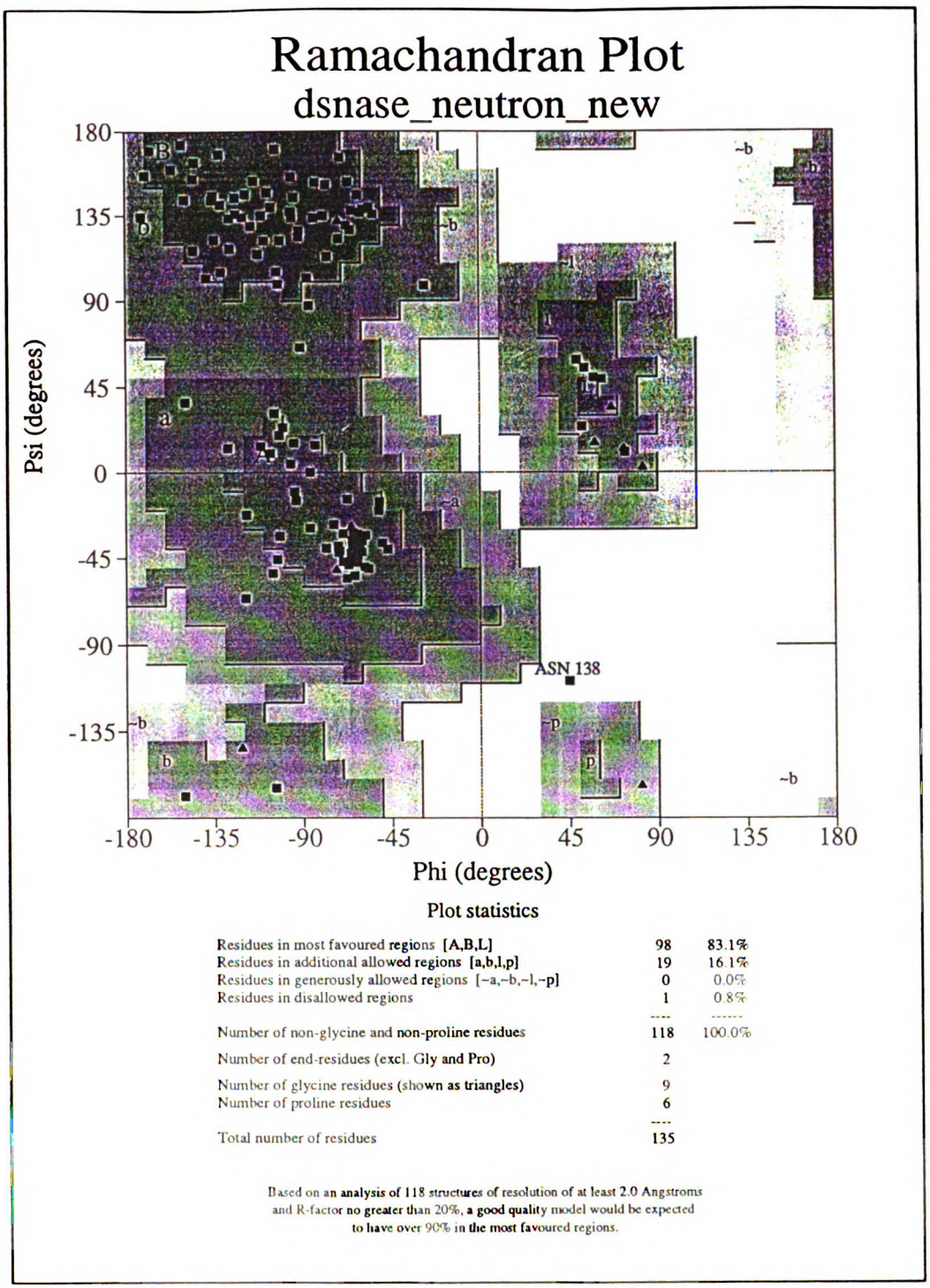


Fig. 7.10: Ramachandran plot of neutron structure of perdeuterated Snase

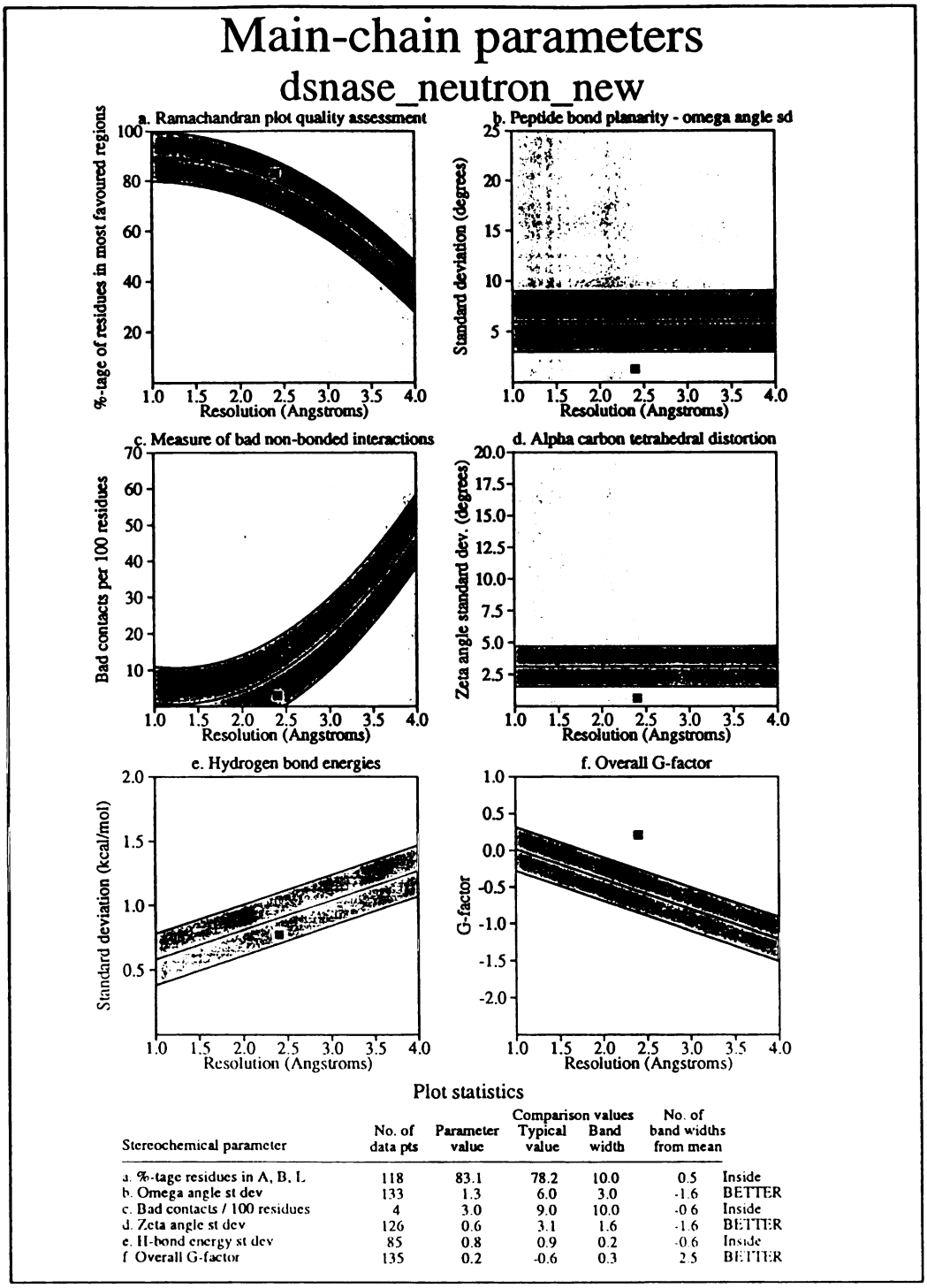
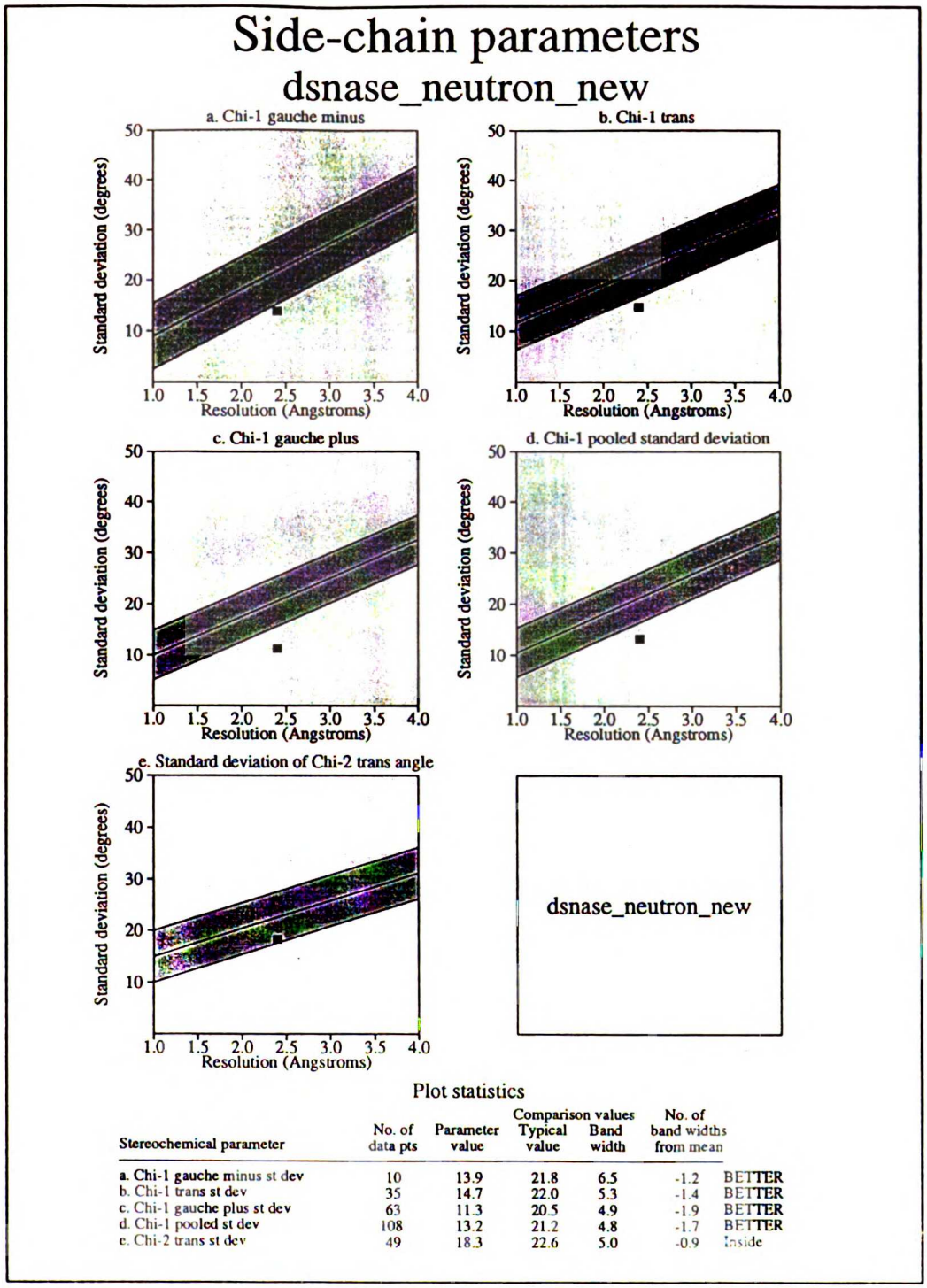


Fig. 7.11: Main-chain parameters

the number of bad non-bonded interactions, and the hydrogen bond energies are well within acceptable values (Fig. 7.11), and the same observation holds for the side chains (Fig. 7.12). The average B-factor for main chain atoms is 23.7Å^2 (RMSD = 0.97Å^2) and for the side chain atoms, 27.20Å^2 (RMSD = 1.94Å^2). The average B-factor for the 13 water molecules is 25.5Å^2, and for the inhibitor, 24.5Å^2. The calcium has a B-factor of 93Å^2. Fig. 7.13a and 7.13b show the B-factor values for individual residues. As expected, the B-factors of the protein are lowest in regions of secondary structure and highest in loop regions, the poorly-ordered loop (residues 44-50) having the highest values (50-80 Å^2).



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	10	13.9	21.8	6.5	-1.2	BETTER
b. Chi-1 trans st dev	35	14.7	22.0	5.3	-1.4	BETTER
c. Chi-1 gauche plus st dev	63	11.3	20.5	4.9	-1.9	BETTER
d. Chi-1 pooled st dev	108	13.2	21.2	4.8	-1.7	BETTER
e. Chi-2 trans st dev	49	18.3	22.6	5.0	-0.9	Inside

Fig. 7.12: Side-chain parameters

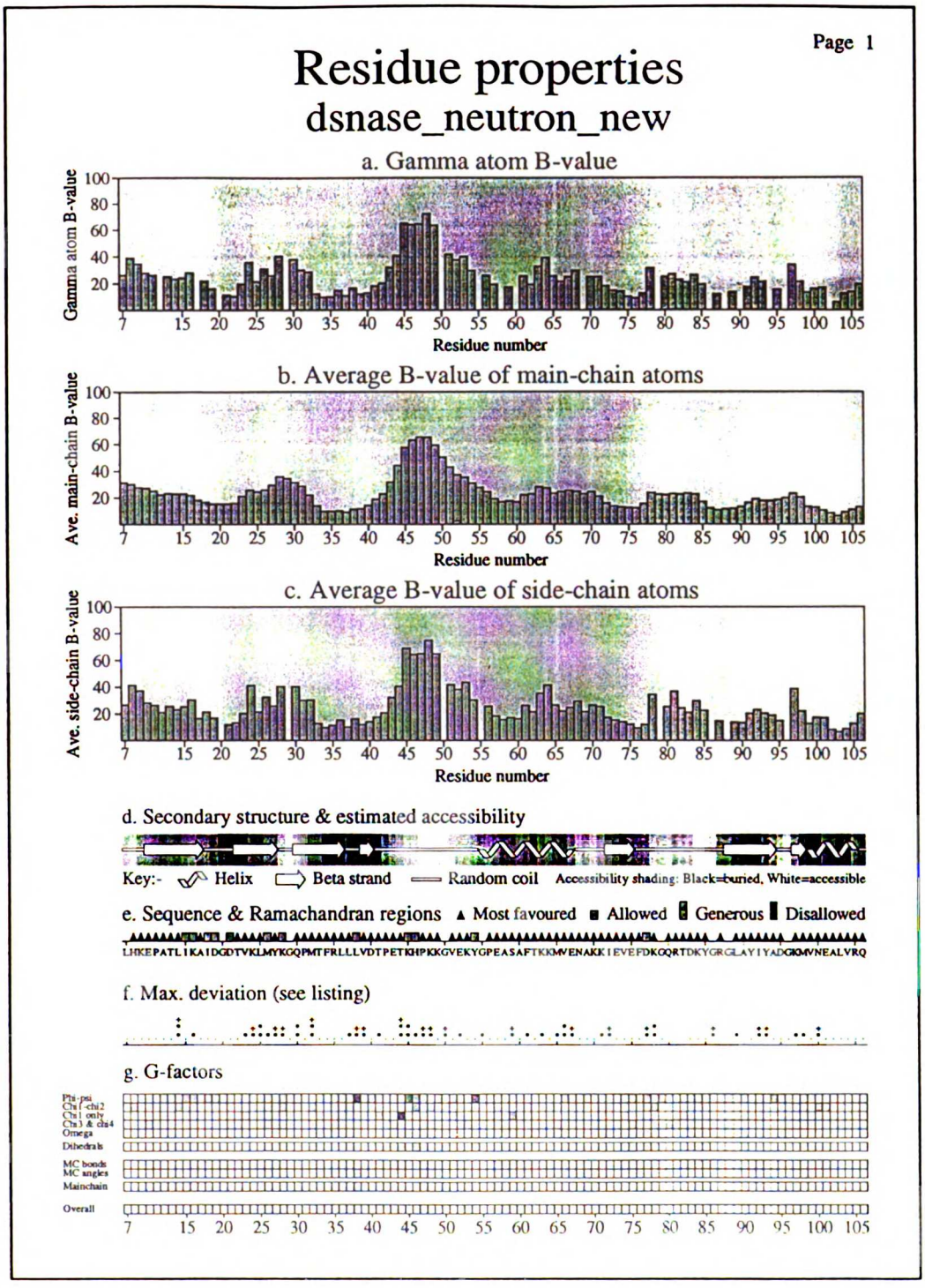


Fig. 7.13a: Residue properties

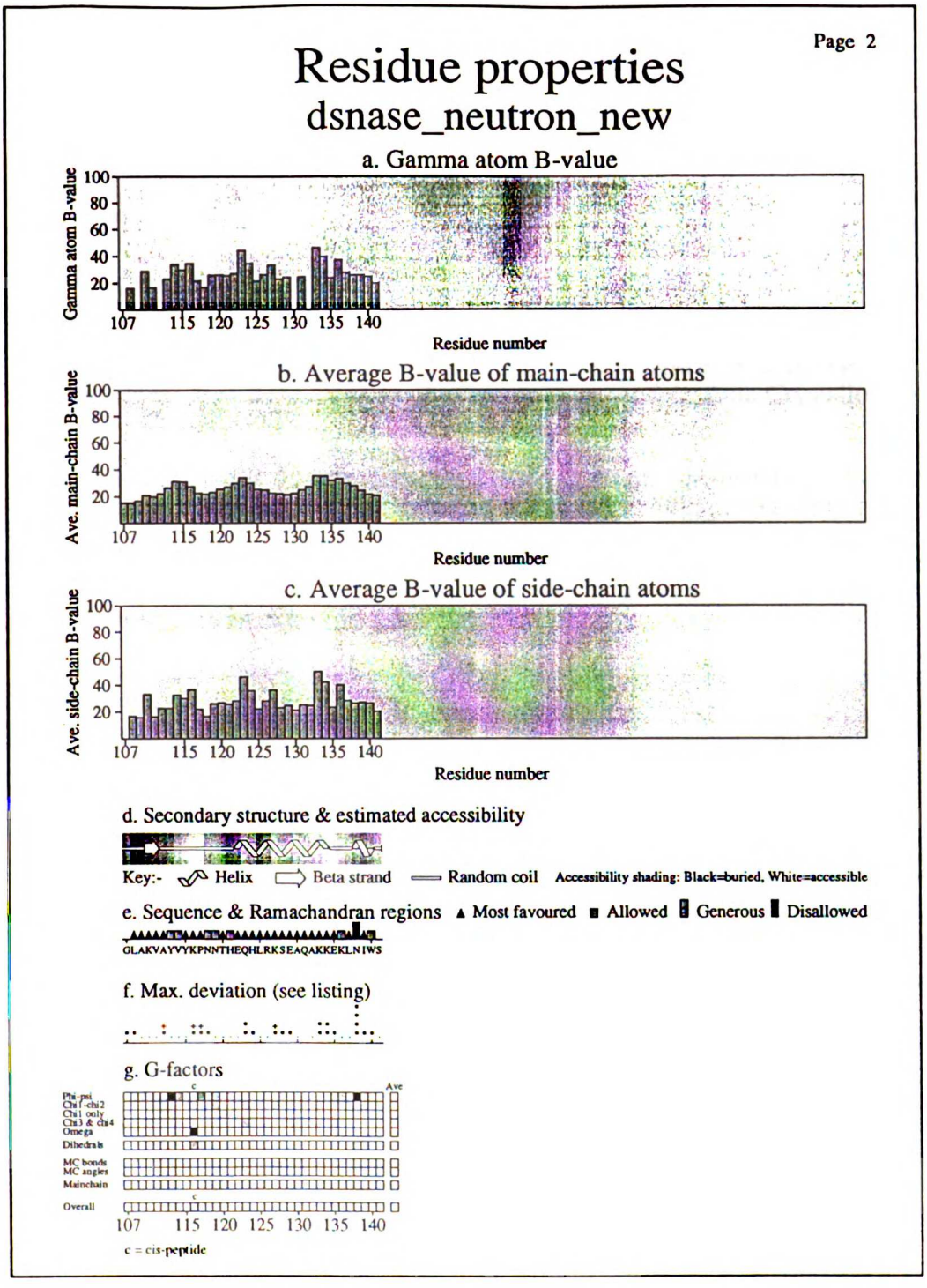


Fig. 7.13b: Residue properties

7.9 References

1. Rhodes, G., *Crystallography Made Crystal Clear: A Guide for Users of Macrocolecular Models*. 1993, San Diego: Academic Press, Inc.

2. Stout, G.H. and L.H. Jensen, *X-ray Structure Determination*. 2nd Edition ed. 1989, New York: John Wiley & Sons.

3. Blundell, T.L. and L.N. Johnson, *Protein Crystallography*. 1976, San Francisco: Academic Press.

4. Brunger, A.T., *X-PLOR Version 3.1 A System for X-ray Crystallography and NMR*. 1992, New Haven: Yale University Press. 382.

5. Jones, T.A., *et al.*, *Improved methods for building protein models in electron density maps and the location of errors in these models*. Acta Crystallogr. Sect. A, 1991. **47**: p. 110-119.

6. Engh, R.A. and R. Huber, *Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement*. Acta Crystallography, 1991. **A47**: p. 392-400.

7. Gamble, T.R., K.R. Clauser, and A.A. Kossiakoff, *The production and X-ray structure determination of perdeuterated Staphylococcal nuclease*. Biophysical Chemistry, 1994. **53**: p. 15-26.

8. McRee, D.E., *Practical Protein Crystallography*. 1993, San Diego: Academic Press, Inc.

9. Loll, P.J. and E.E. Lattman, *The crystal structure of the ternary complex of Staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65Å*. Proteins: Structure, Function, and Genetics, 1989. **5**: p. 183-201.

10. Kossiakoff, A.A., *et al.*, *Analysis of solvent structure in proteins using neutron D2O-H2O solvent maps: Pattern of primary and secondary hydration of trypsin*. Proteins: Structure, Function, and Genetics, 1992. **12**: p. 223-236.

11. Englander, S.W., N.W. Downer, and H. Teitelbaum, *Hydrogen Exchange*. Annual Review in Biochemistry, 1972. **41**: p. 903-924.

12. Collaborative Computational Project, N.4., *The CCP4 suite: Programs for protein crystallography*. Acta Cryst., 1994. **D50**: p. 760-763.

13. Laskowski, R.A., *et al.*, *PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures*. J. Appl. Cryst., 1993. **26**: p. 283-291.

CHAPTER 8. Structure Analysis

In this chapter an anaylsis of the neutron structure of perdeuterated Snase will be described in terms of a comparison between the neutron and X-ray Snase structures, an attempted determination of hydroxyl rotor conformations, and an examination of the hydrogen exchange pattern. Although phase and model bias of the data compromised some of the analysis of the final structure, useful information about the hydrogen and deuterium atoms of the system has been obtained. A comparison of the crystal and solution hydrogen exchanges patterns leads to general conclusions about the effect of crystal packing on protein flexibility and conformation.

8.1 Comparison Between Neutron and X-ray Snase Structure

Due to scattering characteristics of deuterium atoms, neutron maps offer more information for the placement of both the polypeptide backbone and side chains. Especially in the case of a perdeuterated protein, poorly ordered regions of the protein may be more visible. Fig. 8.1 shows the differences in backbone and side-chain atoms between the X-ray [1] and neutron structures of Snase; the root mean square deviation for Cα atoms is 0.17Å, while that for all non-hydrogen atoms, including the inhibitor and calcium, is 0.53Å. As expected, the differences between the main-chain atoms of the two structures are smaller in areas of secondary structure and greater in more flexible loop regions of the protein (Fig. 8.1). A Luzzati plot (Fig. 8.2) shows that the estimated error in the neutron structure's coordinates is 0.30Å; thus the majority of the differences between main-chain atoms of the two structures is due to random errors in the coordinates, and not to significant alterations in the structure.

The residues of greatest variability (>1Å deviation for side chains) are Asp19 (16 $Å^2$), Lys28 (40 Å), Lys48 (73 $Å^2$), and Lys64 (38 $Å^2$). Lys48 is found in a poorly

Fig 8.1: Differences between X-ray and neutron structures of Snase

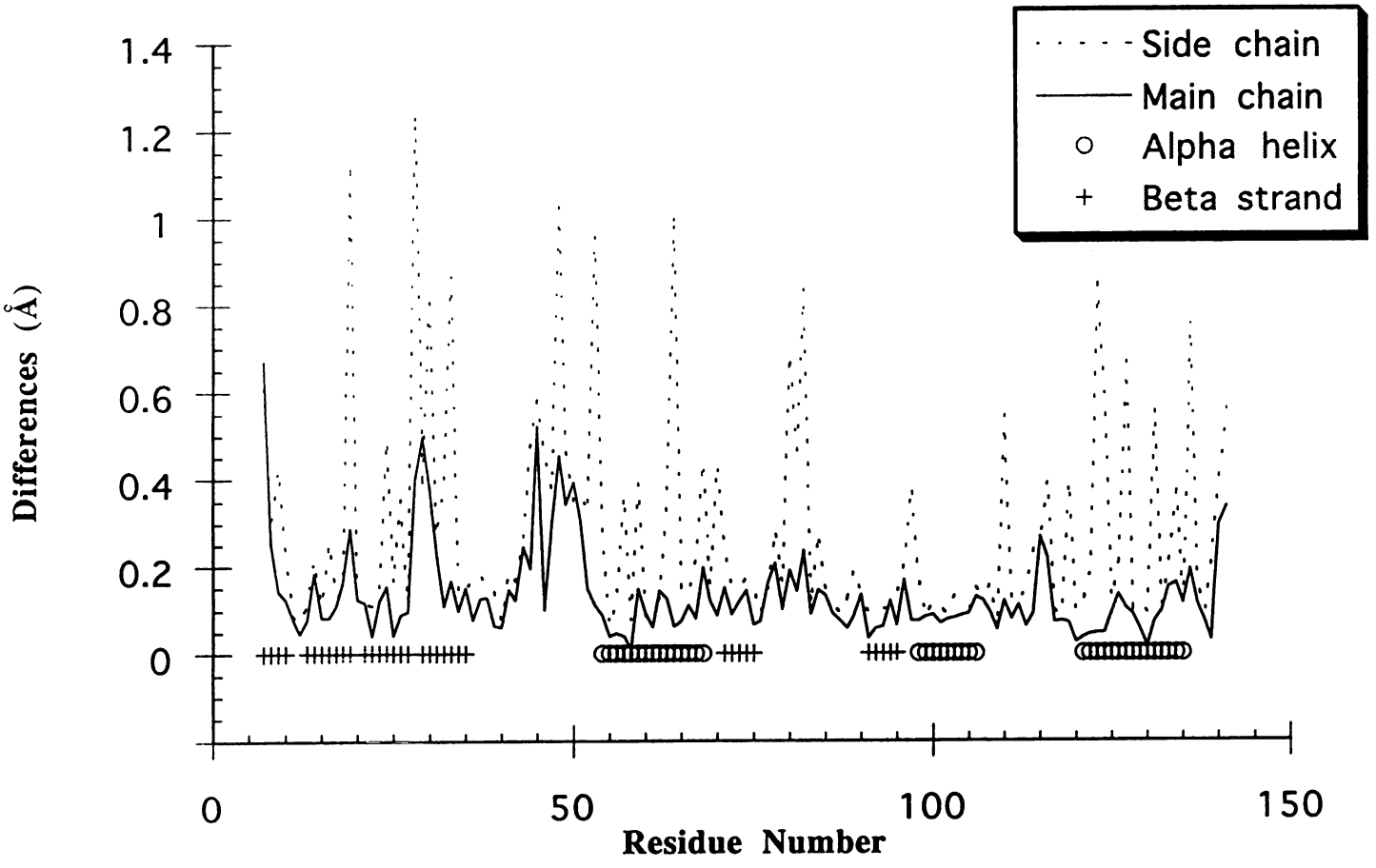
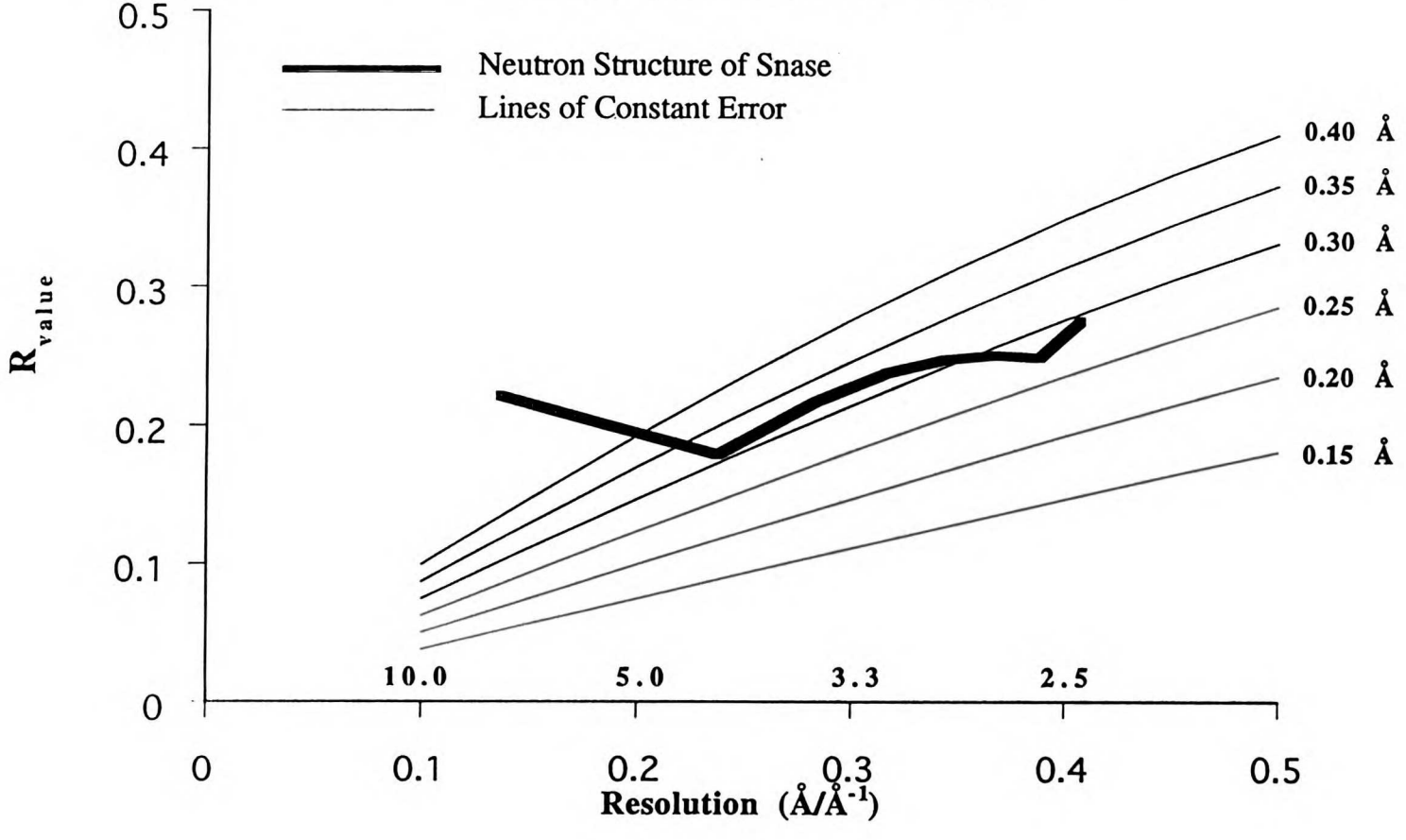


Fig. 8.2: Luzzati plot for neutron structure of Snase
Estimated coordinate error = 0.30Å



ordered loop of the protein and is discussed in section 8.2. Of the remaining residues, $2F_O$-F_C neutron density supports the neutron positions. The added scattering by the deuterium atoms of the side chains has made it possible to re-position these residues. Because the lysines are found on the surface of the protein, and have relatively high B-factor values (average B-factor for all side chains is 27 $Å^2$), the observed differences are not likely to be biologically relevant. As for Asp19, in comparison to the X-ray orientation, the $\chi 1$ angle of the side chain has rotated by 180°, where it can now form a hydrogen bond with the hydroxyl hydrogen of Thr44. In the X-ray structure, the carboxylate group of Asp19 points towards the carboxylate group of Asp21, creating an interaction that is chemically unfavorable. In addition, the X-ray structure contains a water molecule where the neutron structure places the carboxylate group of Asp19. Although the 2Fo-Fc neutron density does not sharply define this altered side chain conformation, the chemical environment supports the change.

8.2 The Poorly Ordered Loop

The loop made up of residues 44-50 is disordered and/or mobile in the both the inhibited and uninhibited structures of Snase [2, 3]. In both the X-ray and neutron structures, this loop has the highest B-factors (neutron = 58$Å^2$//overall = 24$Å^2$, ligated X-ray = 55$Å^2$//overall = 25$Å^2$, unligated X-ray = 38$Å^2$//overall = 19$Å^2$) of the entire protein and density is described as "weak, but interpretable" for the backbone of these residues in the inhibited case [2]. As expected, the RMS deviation for the Cα atoms of these residues is somewhat higher than what is seen for the entire structure (0.36Å vs. 0.17Å), and one of the side chains (Lys48) differs significantly from the inhibited X-ray structure. Because of the poor definition of the loop, this side-chain difference is not believed to be biologically relevant. Although the additional scattering by deuterium

atoms probably helped place this loop, its disorder/mobility continue to make it difficult to determine its conformation accurately.

8.3 The Inhibitor and Active Site

The RMS deviation in the inhibitor molecule between the X-ray and neutron structures is 0.17 $Å^2$. As described in Loll and Lattman [2], the inhibitor (thymidine 3',5'-bisphosphate) fits into a hydrophobic cleft of the protein. In the perdeuterated X-ray structure, the nucleotide base forms no hydrogen bonds directly with the protein; however, its O2, N3, and O4 atoms hydrogen bond to four water molecules (Wat156, Wat157, Wat171, and Wat195); two of these waters (Wat157 and Wat171) are conserved in the neutron structure. The 3'-phosphate, which resides at the top of the cleft and is more exposed than the 5'-phosphate, forms hydrogen bonds with the side chains of Lys84 and Tyr85. The 5'-phosphate maintains all of the interactions seen in the X-ray structure. For example, both of the hydrogen bonds formed between P5' and Arg35, as well as those between P5' and Arg 87 are preserved. The X-ray structure contains two water molecules that interact with the 5'-phosphate, but only one of these (Wat207) is found in the neutron structure, bridging the OH of Y113 to the phosphate O15 and O35 atoms.

The calcium ion was fixed into position during refinement of the neutron structure because of its weak neutron scattering characteristic (4.7 fermi); therefore, it has the identical position as in the superimposed X-ray structure. Loll and Lattman identified seven ligands participating in calcium coordination - one more than expected for octahedral coordination. The protein and inhibitor ligands to the ion (the carboxylate oxygen atoms of Asp21 and Asp40, the carbonyl oxygen of Thr41, and an oxygen of the 5' phosphate) are conserved in the neutron structure. However, the remaining water ligands are not seen in the neutron structure.

The original mechanism proposed for *Staphylococcal* nuclease requires Glu43 to act as a general base and deprotonate the water bridging it to the 5' phosphate of the inhibitor. The resulting hydroxide ion functions as the attacking nucleophile during subsequent hydrolysis of the substrate [2, 4]. Loll and Lattman identify the water (Wat217) that Cotton et al. refer to; however, they also observe a second water (Wat187), which they judge to be an equally likely candidate to act as the attacking nucleophile in the mechanism. Neither of these water molecules are found in the neutron structure. The active site interactions are depicted in Fig. 8.3, while the distances for both the X-ray and neutron structure are listed in Table 8.1. In the inhibited X-ray structure, Glu43 forms hydrogen bonds with the main-chain amide hydrogen atoms of residues Lys45 and His 46, which led to the idea that Glu43 may help stabilize the flexible loop [5]. In the neutron structure, these hydrogen bonds are longer than observed in the X-ray structure (3.28Å/3.29Å compared to 2.89Å/3.02Å), which weakens the argument for such a stabilizing influence.

8.4 Water Structure

Because the resolution of the data does not allow for distinction between the oxygen and deuterium atoms, water orientation analysis is not possible. 11 of the 82 X-ray water molecules found by Loll and Lattman are conserved in the neutron structure, while all 13 waters found in the neutron structure are conserved in the perdeuterated X-ray structure. Under the influence of crystallographic data and the XPLOR force field, the deuterium atoms of the water molecules are oriented with respect to their local environment. All water molecules form at least one, and in most cases multiple, hydrogen bond with the protein (Table 8.2).

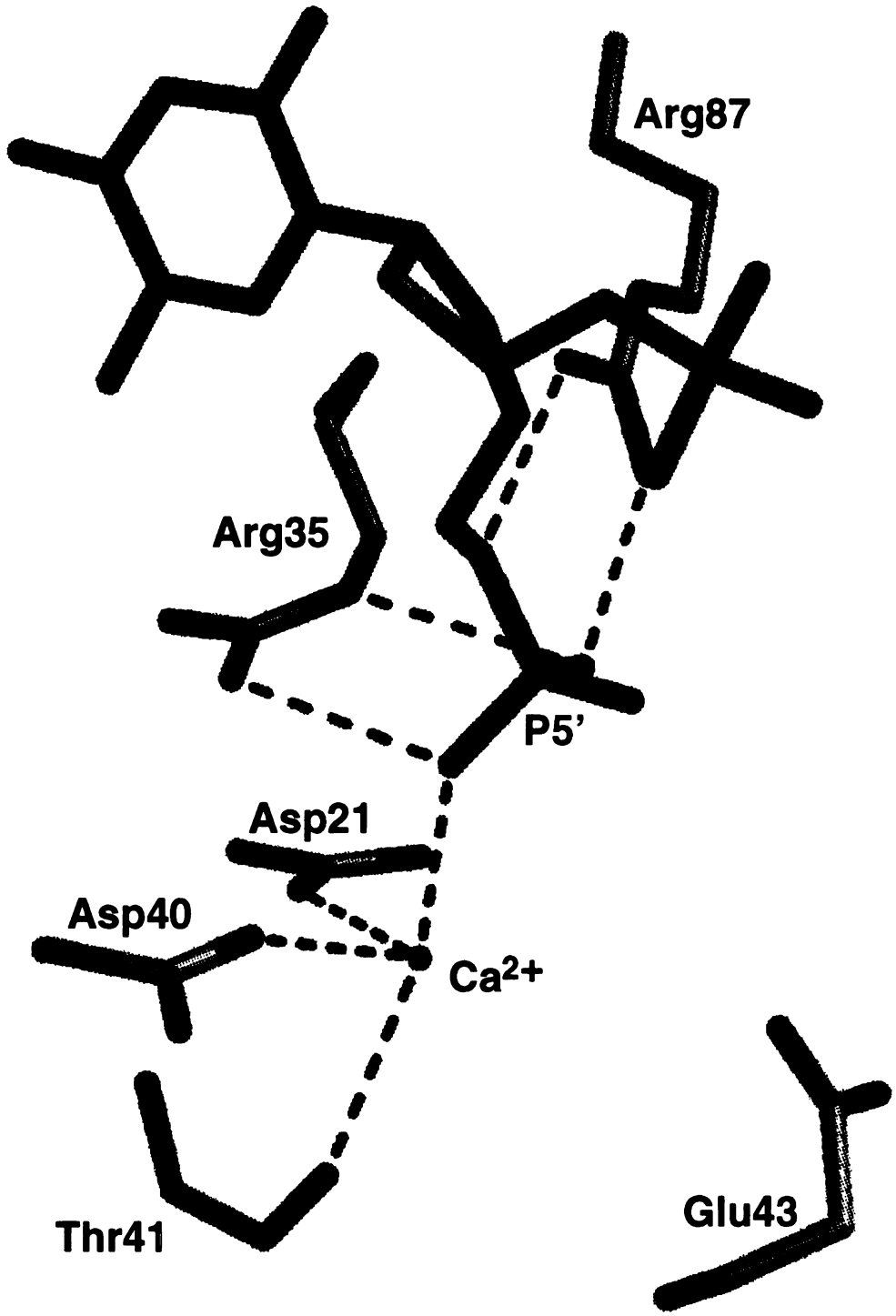


Fig. 8.3: Active site interactions observed in the neutron structure of Snase

Table 8.1: Active Site Interactions

			Neutron Structure	X-ray Structure
	Atom 1	Atom 2	Distance (Å)	Distance (Å)
Inhibitor				
	143 O2	195 OW	-	2.8
	143 N3	157 OW	3.2	3.0
	143 O4	156 OW	-	3.0
	143 O4	173 OW	2.9	3.0
	143 O13	84 LYS NZ	2.8	2.5
	143 O23	85 TYR OH	2.7	2.8
	143 O25	35 ARG NE	3.2	3.2
	143 O15	35 ARG NH2	3.0	2.8
	143 O15	40 ASP OD1	3.1	3.2
	143 O4'	87 ARG NH2	3.1	3.0
	143 O5'	87 ARG NH2	3.3	3.2
	143 O25	87 ARG NH1	3.1	2.9
	143 O25	168 OW	-	2.6
	143 O15	169 OW	2.8	2.8
	143 P5'	183 OW	-	4.6
Calcium				
	142 Ca++	40 ASP OD1	2.6	2.6
	142 Ca++	41 THR O	2.8	2.8
	142 Ca++	21 ASP OD1	2.5	2.6
	142 Ca++	143 O15	2.6	2.6
	142 Ca++	168 OW	-	2.6
	142 Ca++	176 OW	-	2.4
	142 Ca++	183 OW	-	2.4
Glu 43				
	43 GLU OE2	45 LYS N	3.3	2.9
	43 GLU OE2	46 HIS N	3.2	3.0
	43 GLU OE1	183 OW	-	2.7

Table 8.2: Water Hydrogen Bonds

Water	Donor (D)	Acceptor (A)	D-A distance (Å)
202	96 GLY O		2.6
204	109 ALA O		2.7
204	104 VAL O		2.8
204		140 TRP NE1	3.0
207	143 PTP O35		3.1
207	143 PTP O15		2.8
207	113 TYR OH		2.6
207		49 LYS NZ	3.2
209		29 GLY N	3.2
209		28 LYS N	3.7
228		77 ASN N	3.3
228	89 LEU O		2.9
228	80 GLN O		3.0
238		9 LYS NZ	3.1
238	7 LEU O		2.9
239		38 LEU N	3.0
239	113 TYR O		2.9
239	143 PTP O4		2.9
295	77 ASN O		2.7
295	79 GLY O		2.9
299		80 GLN NE2	3.1
310	98 MET O		2.6
320	73 GLU OE1		2.8
321		76 PHE N	3.1
321	8 HIS O		2.6
333		143 PTP N3	3.2

8.5 Hydroxyl Rotor Analysis

Neutron difference density (F_O-F_C) can be used to determine experimentally the conformations of serine, threonine and tyrosine hydroxyl rotors. The relative importance of steric and electrostatic factors in the interior environment of a protein can be assessed by comparing the experimentally determined position of the hydroxyl hydrogen atom with its most favorable steric conformations and participation in local hydrogen bonds. Most hydroxyl rotors will assume low-energy conformations (serine and threonine = -60°, 60°, 180°; tyrosine = 0° and 180°) [6]; however, only experimental determination can assess whether a hydroxyl will form a hydrogen bond despite a sterically unfavorable conformation.

In order to analyze rotor conformation, the hydrogen atom position must first be determined. Because the Snase crystal was soaked in deuterated mother liquor before data collection, all hydroxyl rotors will contain deuterium atoms. These deuterium atoms were left out of the protein for the entire refinement process so that the final model could be used to calculate difference density in an attempt to locate these atoms. For each hydroxyl rotor, the difference density was sampled in 10° intervals around its rotor axis (defined by the C-O bond) at the stereochemically appropriate positions expected for a bonded deuterium atom (bond length = 0.96Å, bond angle (C-O-D) = 109.5°). The molecular display program INSIGHT II [7] was used to perform these measurements. The observed densities were plotted as a function of their rotation angle to give a density profile for each hydroxyl of the protein. The plots for serine, threonine and tyrosine rotors are shown in Figs. 8.4 to 8.6.

It is expected that the position of the deuterium atoms will correspond to the highest positive difference density of the profile; obviously only peaks significantly above the background level of the map can be reliable indicators of these atoms. Unfortunately, for the majority of the hydroxyl rotors (13/18), the positive difference

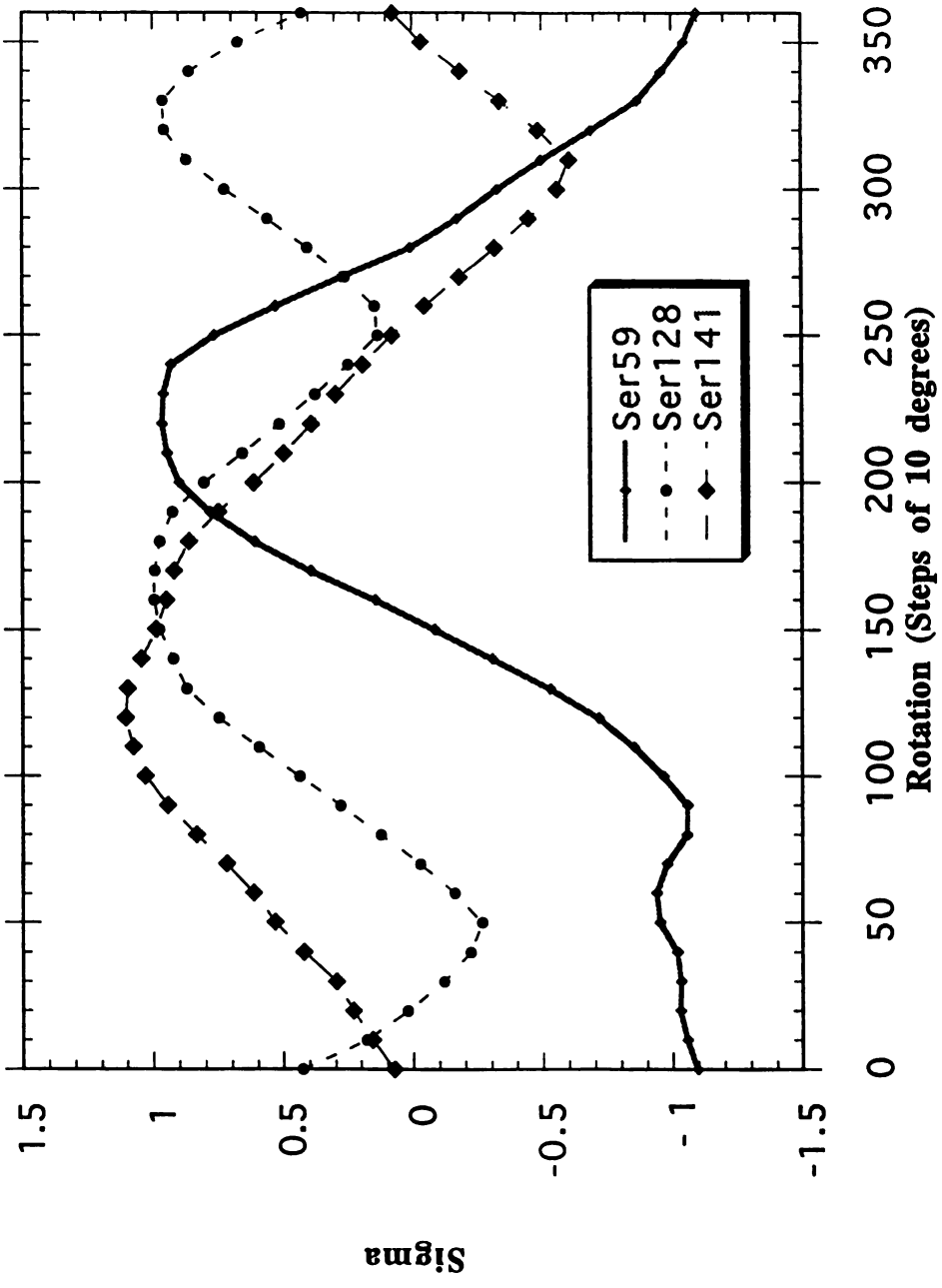
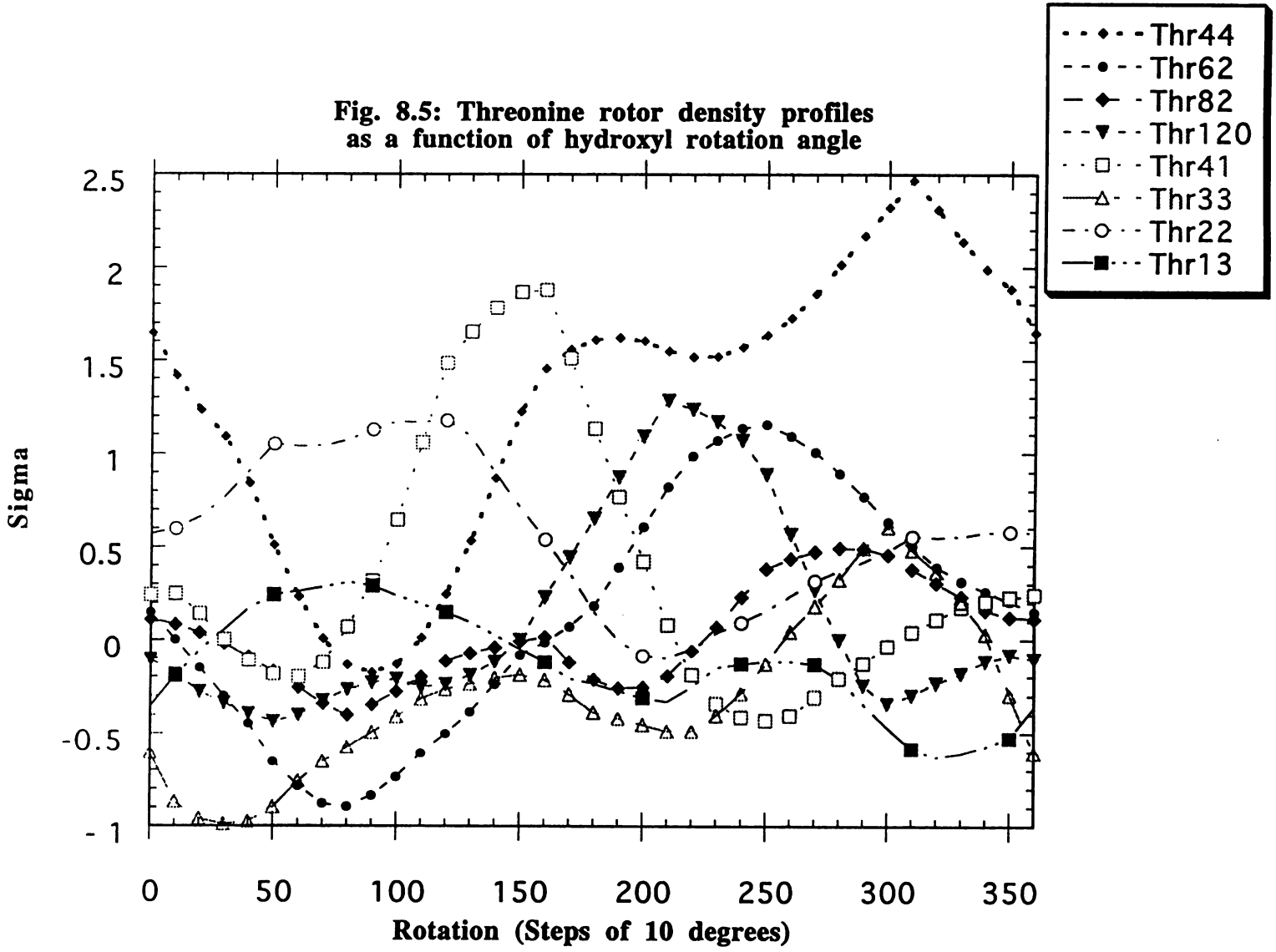


Fig. 8.4: Serine rotor density profiles as a function of hydroxyl rotation angle

Fig. 8.5: Threonine rotor density profiles as a function of hydroxyl rotation angle



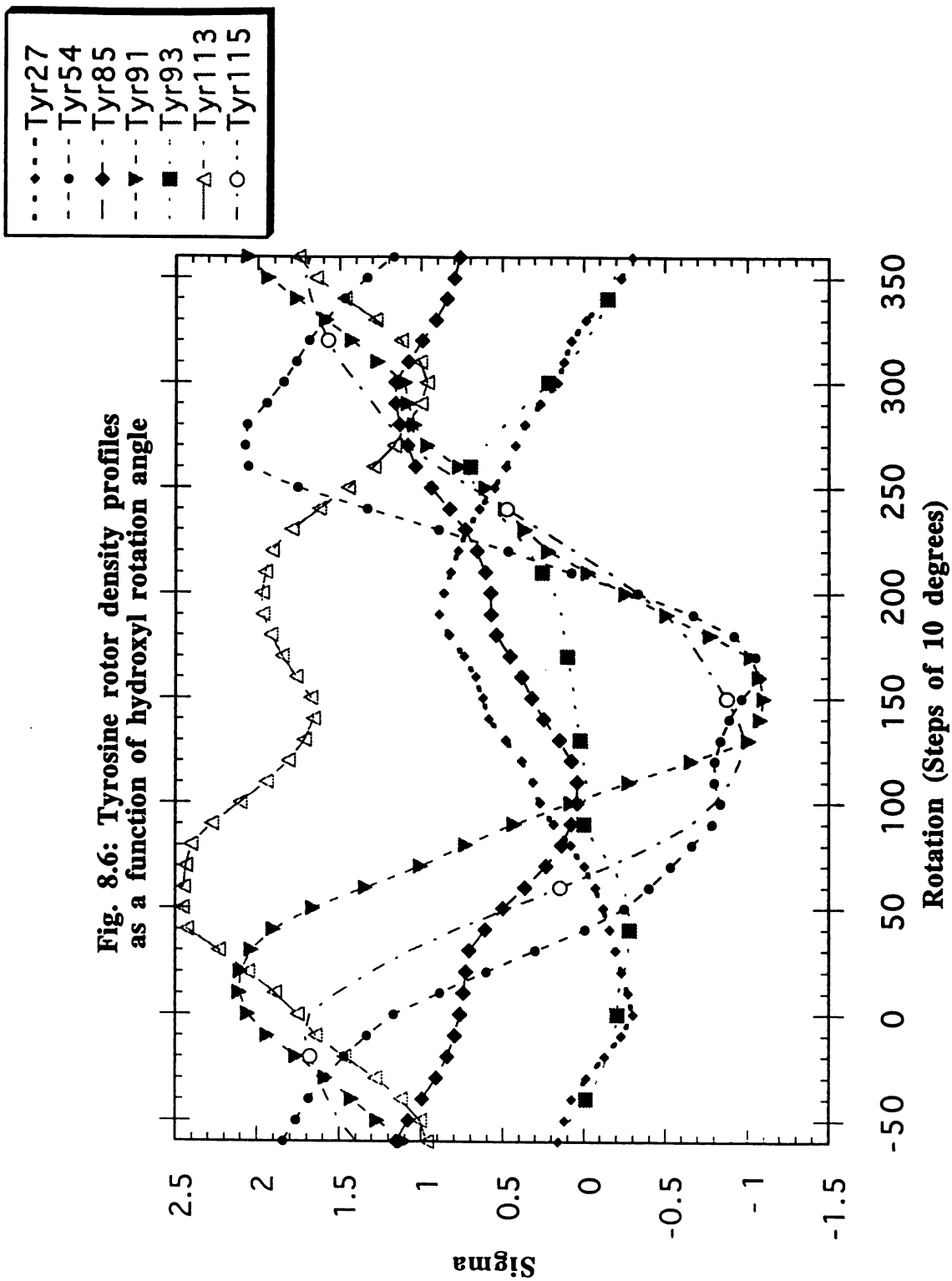


Fig. 8.6: Tyrosine rotor density profiles as a function of hydroxyl rotation angle

peaks fall within the noise of the data (conservatively taken to be 1.5σ), making the positions of their associated deuterium atoms indeterminable. No serine residues have sufficient density to position their hydroxyl rotors; two potential threonine and three potential tyrosine rotor positions, corresponding to the largest positive difference density, are listed in Table 8.3.

As shown in the table, Thr41 and Thr44 both have positive difference density within 30° of the favorable stereochemical positions. Inspection of these rotor positions reveals that hydrogen bonding can occur with neighboring residues. However, of the tyrosine rotors, only one has a minor positive difference peak correlating with a low-energy conformation. The other two tyrosine peaks, which suggest a hydrogen bond despite poor stereochemistry, cannot represent the deuterium atoms of the rotors. For example, if placed in the highest difference peak, the hydroxyl deuterium atom of Tyr54 would be sandwiched between the aromatic ring and the amine group of a neighboring lysine; this is simply not chemically feasible.

Because of the lack of difference density for most rotors and the improbable chemical environment of the positions suggested for the five remaining ones, it is not possible to determine hydroxyl orientations in our structure. This is the result of large amount of phase and model bias caused by weak data (Rmerge = 16.4% on I), incompleteness (84%), and low observation-to-parameter ratio (4625 to 2232 x 4), as well as the relatively low resolution of the maps (2.4Å).

8.6 Hydrogen/Deuterium Exchange

Exchange between main-chain amide hydrogen atoms and the surrounding solvent is possible via both acid and base catalysis, with pH 3 partitioning the two mechanisms [8]. Many factors contribute to the reason why a protein's amide hydrogen atoms exchange at vastly different rates. Hydrogen exchange in a folded protein will depend on

Table 8.3: Hydroxyl Rotor Orientation

Residue	Rotor angle of highest difference peak	B-factor of oxygen (Å2)
TYR 54	270°	28
TYR 91	25°	28
TYR 113	75°, minor peak at 190°	15
THR 41	150°	27
THR44	310°, minor peak at 180°	37
Note: Sterically favorable hydroxyl positions for tyrosine are 0° and 180° and for threonine are 60°,180° and 300° (-60°)		

temperature, pH, and solvent composition, as well as primary, secondary and tertiary structure [8, 9]. In a crystal, it is reasonable to assume that hydrogen exchange rates will be influenced by the packing of protein molecules against their neighbors.

As described in Chapter 7, the hydrogen exchange pattern of the Snase crystal was determined by identifying the isotope occupancy of each main-chain amide H/D site. Fig. 8.7 shows the mapping of the exchange pattern onto the three-dimensional structure of the protein; deuterium atoms are shown in gray, hydrogen atoms in magenta, and undetermined positions in green. Because Snase was purified and crystallized in hydrogenated solvents, and then soaked in deuterated mother liquor, deuterium atoms indicate that exchange has taken place, while the presence of hydrogen atoms denotes protection from exchange. Of the 128 exchangeable amide sites, 64 are occupied by deuterium atoms, 28 by hydrogen atoms, and 36 could not be determined. Of the 28 protected sites, 9 are found in beta strands, 9 in alpha helices, and 3 in type I beta turns. Seven amide protons (Leu36, Leu37, Val39, Arg87, Leu89, Ala90, and Gly107) occur in loop regions, but remain protected from exchange. Inspection of the structure shows that six of these seven protons participate in hydrogen bonds to the protein, the one exception being Val39.

Because of crystal solvent content and the normal fluctuations of the protein, it is expected that most amide H/D atoms will be exposed to solvent and participate in exchange. This expectation is clearly fulfilled in that half (50%) of the main-chain amide H/D sites are occupied by deuterium. On the other hand, protection should occur in more rigid regions of the protein, where amide atoms participate in protein-protein hydrogen bonds. The prediction of a strong correlation between protected amide H/D positions and secondary structure is borne out in these data. According to the regional melting model for hydrogen exchange in proteins, hydrogen bonding networks must be disrupted in order for exchange to occur [10]. The exchange pattern of the beta sheet depicted in Fig. 8.8 shows that complete protection from exchange occurs only at H/D

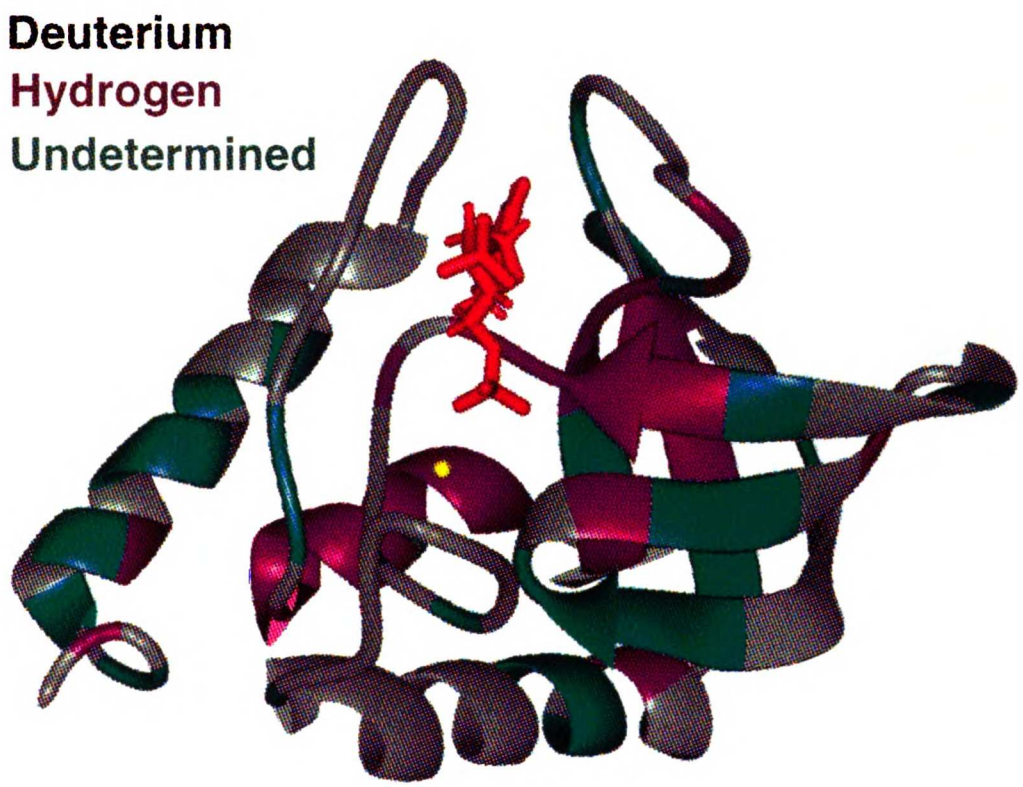


Fig. 8.7: The hydrogen/deuterium exchange pattern for crystallized Snase. Residues with amide deuterium atoms are shown in grey, amide hydrogen atoms in magenta, and undetermined sites in green. The inhibitor is shown in red, and the calcium ion as a small yellow sphere.

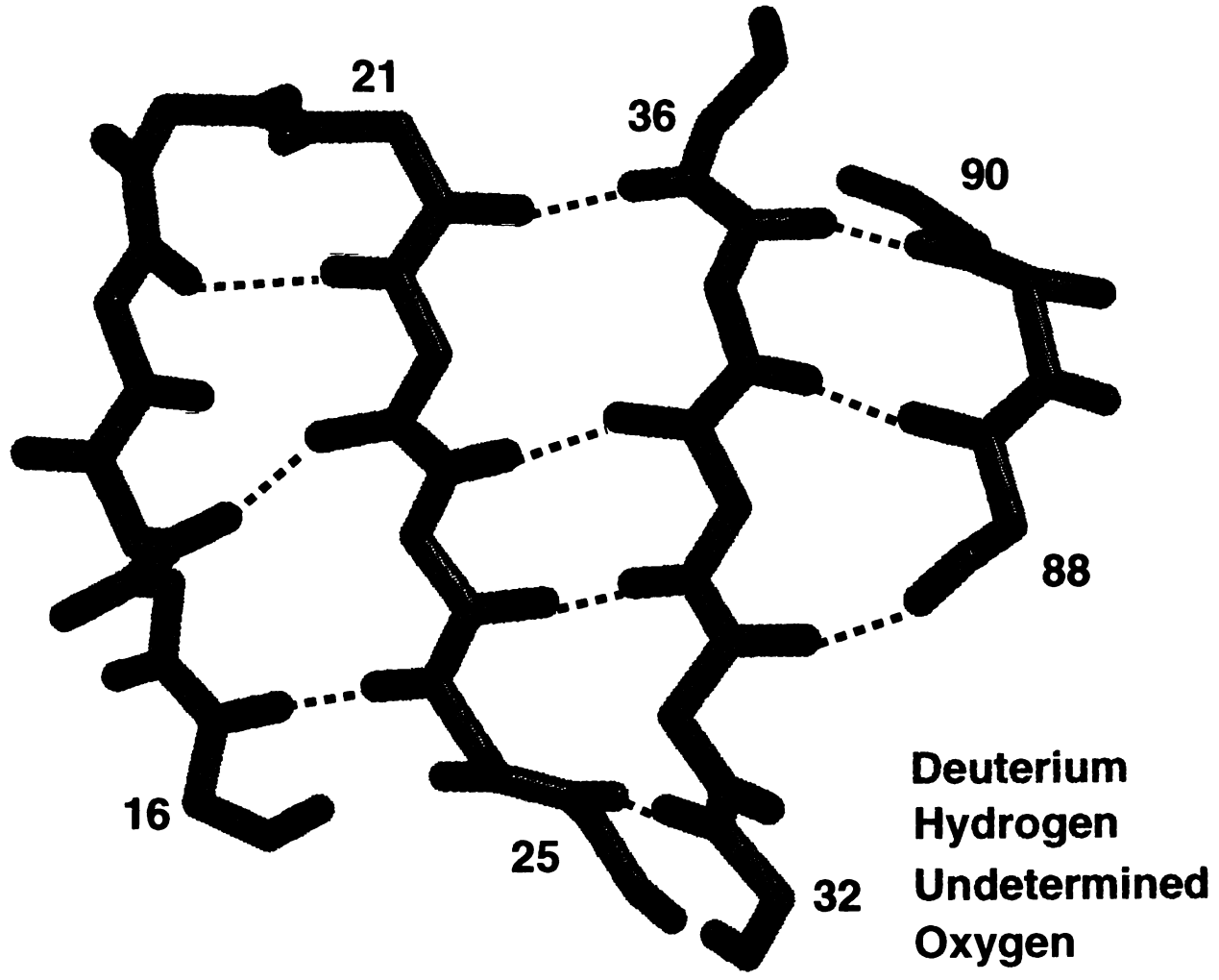


Fig. 8.8: The hydrogen/deuterium exchange pattern in a beta sheet. Backbone and deuterium atoms are shown in grey, hydrogen atoms in magenta, undetermined atoms in green, and oxygen atoms in red.

sites located on the central strands (21-25 and 32-36). The outer strands of the sheet, which have similar hydrogen bonding networks, contain many undetermined and exchanged sites. The exchange pattern observed for this beta sheet suggests that the central strands are more rigid, while the edge strands have more flexibility. Overall, the hydrogen exchange pattern for crystalline Snase agrees with general trends expected from solvent accessibility and secondary structure and can be used to identify regions of greater or lesser flexibility.

The hydrogen exchange pattern for Snase in solution has been determined by NMR in two independent research efforts [11, 12]. In the experiment done by Torchia et al., Snase was allowed to exchange with D_2O for 24 hours at 37°C, pH 7.0 at which point spectra were taken to identify the protected amide protons [12]. Loh et al. determined exchange lifetimes (the inverse of the rate constant) and protection factors for the amide protons, albeit under slightly different conditions (20 minutes of exposure to D_2O at 37°C, pH 5.5) [11].

A comparison of the solution (Torchia) and crystal exchange patterns is shown in Fig. 8.9. The same color scheme is used for both hydrogen exchange patterns; in addition, amide protons protected in the neutron, but not the NMR case, are colored yellow-orange. Most striking is the similarity between the two patterns; 24 of the 45 (53%) protected amides of the solution exchange pattern, including four sites located in loop regions, are still protected from exchange after more than a year of crystal soaking. The lifetime values (Loh) for the protected amide sites common to both the neutron and NMR (Torchia) structure are listed in Table 8.4. The 21 amide H/D sites protected in the NMR but not the neutron structure have predictably exchanged in the time frame of the neutron experiment (> one year); in solution the average lifetime of these sites is 178 hours. Given the differences in pH (neutron = 8.14, Torchia = 7.0, Loh = 5.5), the neutron structure has the potential for the highest rate of exchange due to the base-catalyzed exchange mechanism.

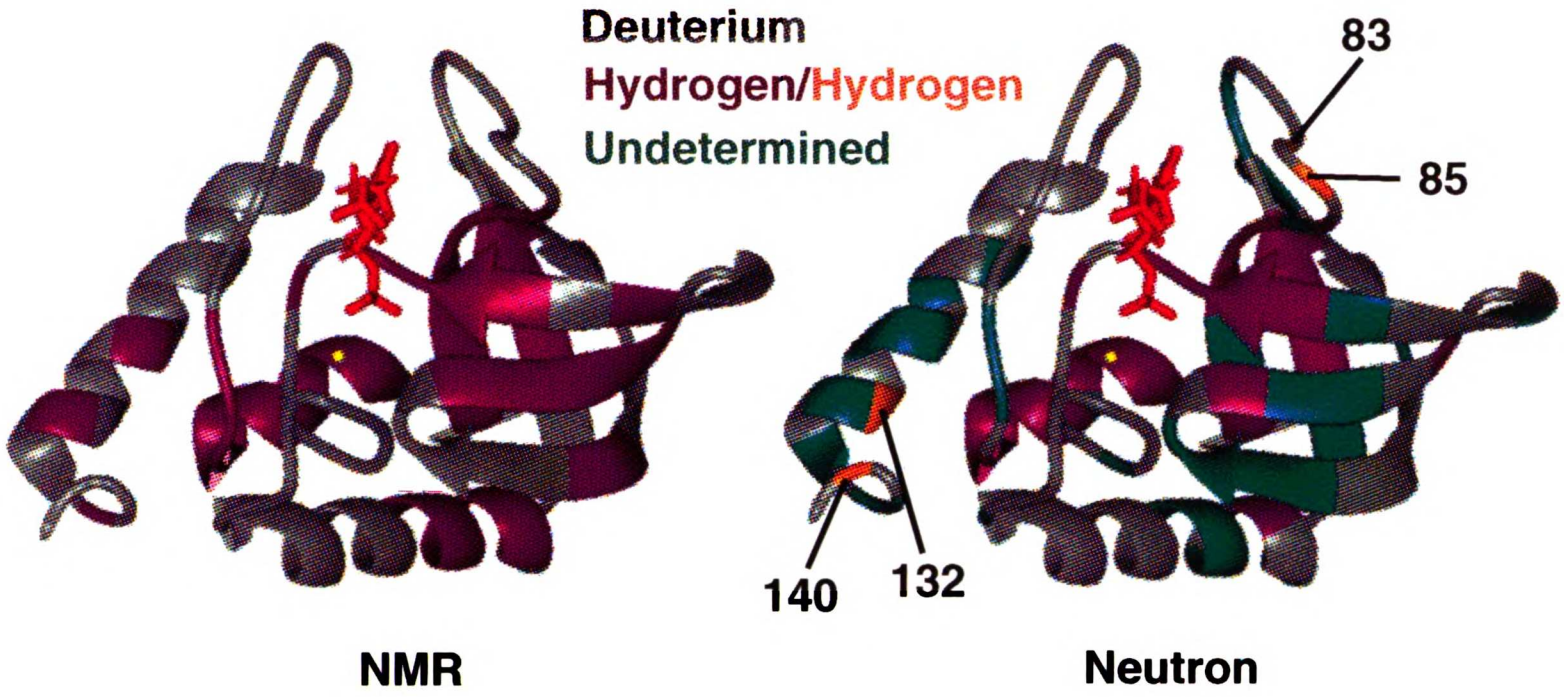


Fig. 8.9:A comparison of solution and crystal hydrogen/deuterium exchange patterns. Residues with amide deuterium atoms are shown in grey, amide hydrogen atoms in magenta, and undetermined sites in green. The inhibitor is shown in red, and the calcium ion as a small yellow sphere. Four amide hydrogens that are protected in the neutron, but not the NMR case, are shown in orange.

Table 8.4: NMR Lifetimes for Protected Neutron Amide Positions

Residue Position	Protected Amides NMR	Protected Amides Neutron	NMR lifetime (h)
22	+	+	
26b	+	+	679 ± 230
34b	+	+	410 ± 95
35b	+	+	353 ± 97
36b	+	+	
37b	+	+	279 ± 100
39	+	+	690 ± 240
66a	+	+	406 ± 87
73b	+	+	337 ± 87
74b	+	+	263 ± 42
75b	+	+	265 ± 87
87s	+	+	
89om	+	+	554 ± 390
90b	+	+	535 ± 90
91b	+	+	763 ± 290
94om	+	+	437 ± 91
99om	+	+	388 ± 71
100	+	+	244 ± 24
101	+	+	280 ± 44
103a	+	+	298 ± 69
104a	+	+	630 ± 170
105a	+	+	264 ± 44
106a	+	+	209 ± 22
107om	+	+	126 ± 15
Key: b = beta sheet h-bond			
a = helical h-bond			
om = other main chain - main chain h-bond			
s = main chain - side chain h-bond			

There appears to be a correlation between the undetermined amide positions of the neutron pattern and the protected sites of the NMR data. 39% (14 out of 36) of the undetermined sites match protected sites of the NMR data; it is likely that these 14 positions represent sites that are partially protected from exchange (i.e. some molecules in the crystal have exchanged at that site, others have not). It is not possible to determine whether the remaining undetermined amide sites (22 out of 36) represent partial exchange or are the result of phasing error and model bias contained in the data. Table 8.5 shows a comparison of the NMR and neutron exchange pattern on a residue by residue basis. Hydrogen bond participation of these backbone amide protons is also included.

Four amide protons (Asp83, Tyr85, Ala132, and Trp140) are protected in the neutron, but not the NMR, structure; these sites provide evidence for local changes in protein flexibility due to direct interaction between symmetry related molecules in the crystal lattice. Asp83 and Tyr85 are found in a loop region of the protein involved in crystal contacts. Main-chain atoms of both residues participate in hydrogen bonds, as well as the side-chain of Asp83 (Fig. 8.10). Although these bonds extend across the loop, exchange is complete in solution. In the crystal, however, these bonds are reinforced by crystal packing with two symmetry related molecules (Fig. 8.11). Ala132 is found in an alpha helix in which protection occurs in solution, but at different sites (Glu129 and Lys133). Because the hydrogen bonding potential of these amide protons is satisfied by the secondary structure, no alternative hydrogen bonding with symmetry related molecules is feasible. However, a symmetry related molecule is in van der Waals contact with the protein in this region as shown in Fig. 8.12. Clearly the flexibility of the helix has been affected by this crystal contact, perhaps preventing unraveling at its C-terminal end, such that different amide protons are protected in crystal and solution environments. Finally, Trp140 is found in a type I beta turn and forms a hydrogen bond with Leu137. Given that this residue is well-ordered (B ~ 20 $Å^2$), the protection of this H/D site is unlikely to be artifactual. Therefore, it is possible that the crystal contact affecting the

Table 8.5: Hydrogen Exchange Data: Neutron vs. NMR

Residue Position	Protected Amides NMR	Protected Amides Neutron	Undetermined Neutron	H-bond
8			+	
12b			+	+
15b	+			+
16b	+			+
17			+	
18b			+	+
19b			+	+
20			+	
22	+	+		
23b	+		+	+
24b	+		+	+
25b	+			+
26b	+	+		+
30b			+	+
32b	+			+
33			+	
34b	+	+		+
35b	+	+		+
36b	+	+		+
37b	+	+		+
39	+	+		
43a			+	+
44			+	+
53			+	
62a	+		+	+
63a	+		+	+
64a	+		+	+
65a	+			+
66a	+	+		+
67a	+			+
68a			+	+
71s			+	+
72			+	
73b	+	+		+
74b	+	+		+
75b	+	+		+
77s			+	+
78s			+	+
83om		+		+
85s		+		+
87s	+	+		+
88b	+		+	+
89om	+	+		+
90b	+	+		+
91b	+	+		+
92s	+		+	+
93b	+		+	+
94om	+	+		+
97om			+	+
99om	+	+		+
100	+	+		
101	+	+		
102a	+		+	+
103a	+	+		+
104a	+	+		+
105a	+	+		+
106a	+	+		+
107om	+	+		+
108om	+			+
109	+		+	+
110om	+		+	+
125a			+	+
127a			+	+
129a	+		+	+
130a			+	+
132a		+		+
133a	+		+	+
135a			+	+
137om	+		+	+
138			+	
140om		+		+
Key: b = beta sheet h-bond				
a = helical h-bond				
om = other main chain - main chain h-bond				
s = main chain - side chain h-bond				

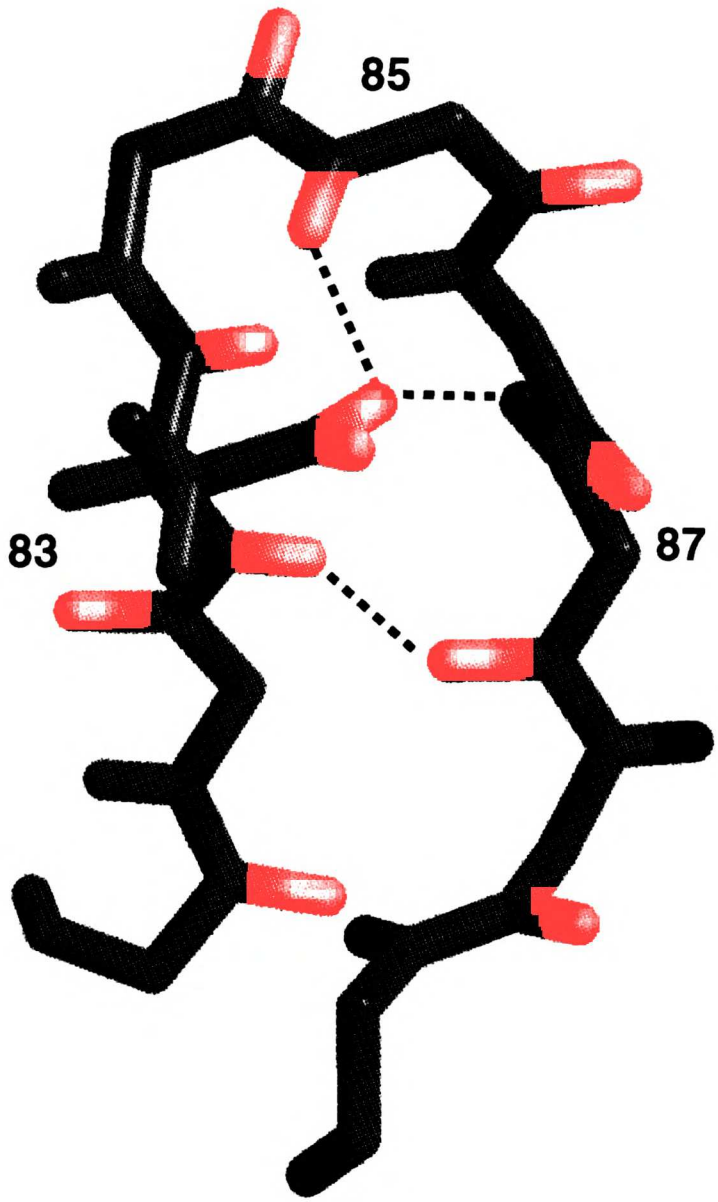


Fig. 8.10: Loop containing Asp83 and Tyr85. Backbone and deuterium atoms shown in grey, hydrogen atoms in magenta, undetermined amide postions in green, and oxygen atoms in red. Two amide hydrogen atoms protected in the neutron but not the NMR structure are shown in orange.

Fig. 8.11: Crystal contacts surrounding Asp83. Main protein molecule shown in gray, Asp83 shown in yellow˙ orange, and symmetry related molecules shown in cyan and blue.

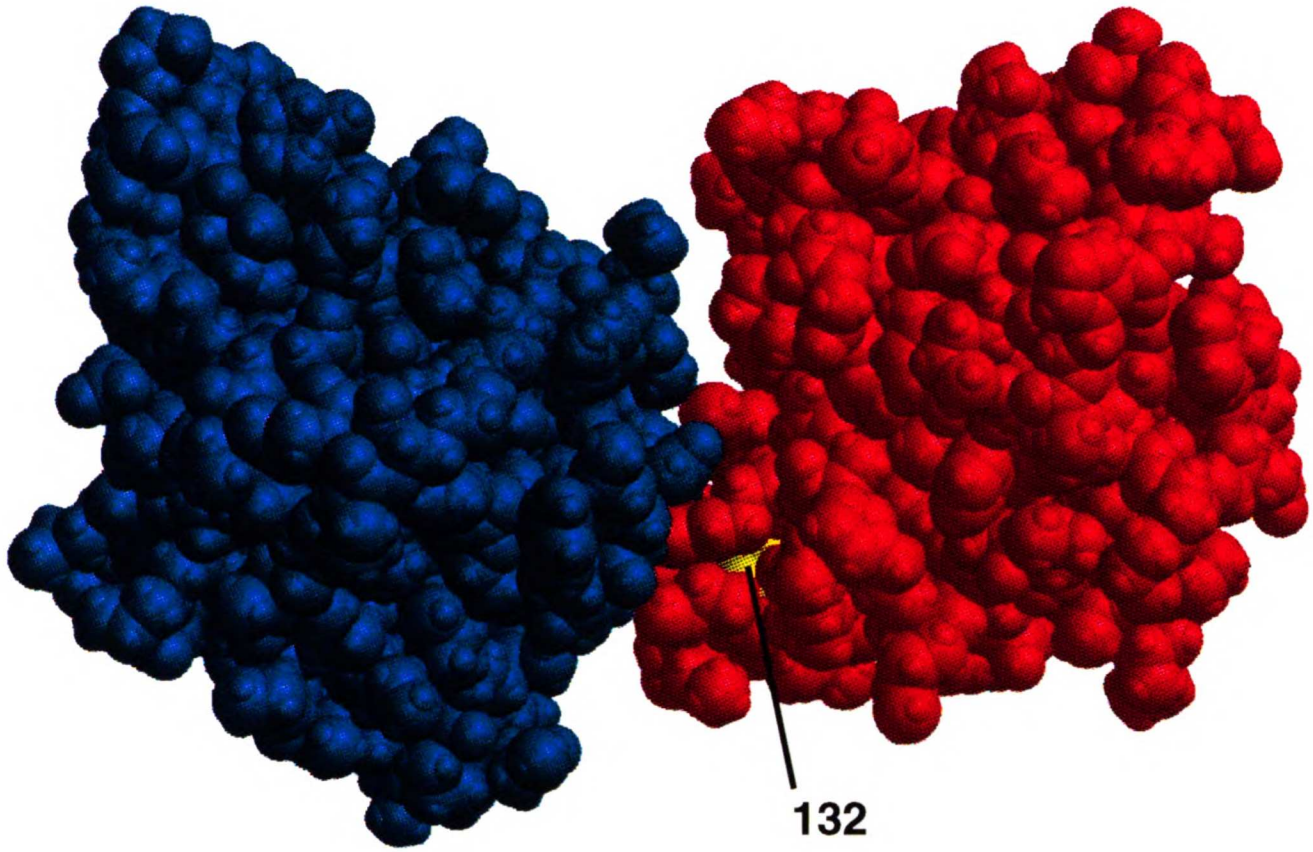


Fig. 8.12: Crystal contact involving the C–terminal helix. Main protein molecule shown in red, symmetry–related molecule in cyan. The amide hydrogen of Ala132 is protected from exchange in the neutron structure but not in the NMR structure.

exchange of Ala132 is also doing the same for the amide hydrogen atom of Trp140.

In addition to the observed local effects of crystal packing, these contacts also influence the global flexibility of the protein. An exchange lifetime is the time at which 64% of the protein population will have exchanged at a particular amide H/D site [11]. It is customary to assume that after 7 lifetimes full exchange has occurred. On average, the protected amide protons found in the neutron structure have lifetimes (401 hours, averaged from Table 8.4) greater than the average lifetime for all amide H/D sites (190 hours). But even the amide proton with the longest lifetime (residue Tyr91, lifetime ≈ 32 days) will have completely exchanged after one year (the length of neutron data collection) in solution. This observed change in exchange rates indicates that the dynamic properties of the protein have been altered in the crystal. This is not to suggest that the time averaged conformation of the crystallized protein differs significantly from the protein in solution. Rather, the parallelism of the exchange patterns gives evidence for agreement between crystal and solution structures, as well as for similar dynamic properties, albeit on different time scales.

Based on the correlation between hydrogen exchange protection and secondary structure in both neutron and NMR experiments, hydrogen exchange patterns can identify flexible and rigid regions of a protein. By comparing the hydrogen exchange patterns of solution and crystal structures of Snase, it is seen that both local and global changes in exchange rates are induced by the crystal environment. The global consequences of crystal packing on hydrogen exchange can be likened to the effect of inhibitor binding on the exchange pattern of unligated Snase [11]. Just like symmetry related molecules, the inhibitor only directly interacts with a small region of the protein, yet a decrease in flexibility is propagated throughout the entire molecule. Furthermore, at the site of a crystal contact, the protein is locally restricted; this is clearly demonstrated by the additional protection of six sites in the neutron structure. Hydrogen bonding across crystal contact interfaces is not necessary in order for enhanced protection to occur, rather

existing intra-molecular hydrogen bonds are reinforced. In conclusion, although overall protein structure is not altered by the crystal environment, the time scale of a protein's dynamic properties clearly are.

8.7 References

1. Gamble, T.R., K.R. Clauser, and A.A. Kossiakoff, *The production and X-ray structure determination of perdeuterated Staphylococcal nuclease.* Biophysical Chemistry, 1994. **53**: p. 15-26.

2. Loll, P.J. and E.E. Lattman, *The crystal structure of the ternary complex of Staphylococcal nuclease, Ca2+, and the inhibitor pdTp, refined at 1.65Å.* Proteins: Structure, Function, and Genetics, 1989. **5**: p. 183-201.

3. Hynes, T.R. and R.O. Fox, *The crystal structure of Staphylococcal nuclease refined at 1.7Å resolution.* Proteins: structure, function, and genetics, 1991. **10**: p. 92-105.

4. Cotton, F.A., E.E. Hazen Jr., and M.J. Legg, *Staphylococcal nuclease: Proposed mechanism of action based on structure of enzyme-thymidine 3',5'-bisphosphate-calcium ion complex at 1.5Å resolution.* Proc. Natl. Acad. Sci. U.S.A., 1979. **76**: p. 2551-2555.

5. Judice, J.K., *et al.*, *Probing the mechanism of Staphylococcal nuclease with unnatural amino acids: Kinetic and structural studies.* Science, 1993. **261**: p. 1578 - 1581.

6. Kossiakoff, A.A., J. Shpungin, and M.D. Sintchak, *Hyroxyl hydrogen conformations in trypsin determined by the neutron diffraction solvent difference map method: Relative importance of steric and electrostatic factors in defining hydrogen-bond geometries.* Proc. Natl. Acad. Sci. USA, 1990. **87**: p. 4468-4472.

7. *Insight II.* 1995, Biosym Technologies: San Diego.

8. Englander, S.W., N.W. Downer, and H. Teitelbaum, *Hydrogen Exchange.* Annual Review in Biochemistry, 1972. **41**: p. 903-924.

9. Bai, Y., *et al.*, *Primary structure effects on peptide group hydrogen exchange.* Proteins: Structure, Function, and Genetics, 1993. **17**: p. 75-86.

10. Kossiakoff, A.A., *Protein dynamics investigated by the neutron diffraction-hydrogen exchange technique.* Nature, 1982. **296**: p. 713-721.

11. Loh, S.N., *et al.*, *Hydrogen exchange in unligated and ligated Staphylococcal nuclease.* Biochemistry, 1993. **32**: p. 11022-11028.

12. Torchia, D.A., S.W. Sparks, and A. Bax, *Staphylococcal nuclease: sequential assignments and solution structure.* Biochemistry, 1989. **28**: p. 5509-5524.

CHAPTER 9. Conclusion

Neutron crystallography is a powerful tool for structural biologists. When used in complement with X-ray crystallography and NMR, the technique can be used to study fundamental protein properties such as protein dynamics and enzyme mechanism. But despite the advantages, the difficulties inherent to neutron diffraction have clearly prevented its widespread use, a point graphically illustrated by the small number of neutron structures in comparison to the hundreds determined by X-ray crystallography and NMR. As the ideas and work for this thesis progressed, two goals have been of equal priority: 1) to use protein perdeuteration to overcome some of the obstacles inherent to neutron diffraction and 2) to use neutron diffraction to study both the specific and general protein properties of *Staphylococcal* nuclease.

One important aspect of this work has been to demonstrate that protein perdeuteration is a viable option for neutron crystallography. Snase can be produced in *E. coli* on deuterated medium at high yields equivalent to wild-type expression. The perdeuteration level of the resulting protein is reproducibly high (>95%) as verified by electrospray mass spectrometry. Perdeuteration was shown not to affect the crystallization conditions or the three-dimensional structure of the protein. It is reasonable to conclude that any bacterially expressed protein with known crystallization conditions would be a good candidate for protein perdeuteration, and, once successfully perdeuterated, a good candidate for neutron diffraction.

Because of technical problems, the effect of protein perdeuteration on signal-to-noise in my neutron data could not be properly assessed. The underlying rationale for protein perdeuteration is sound: deuterium scatters neutrons more coherently than hydrogen; therefore, replacing hydrogen with deuterium will decrease incoherent scattering. This fundamental nuclear diffraction characteristic is why soaking crystals in deuterated mother liquor improves the quality of neutron data. Replacing virtually all of

the non-labile hydrogen atoms in a protein crystal should result in a measurable improvement in the signal-to-noise of the data by reducing the background. Unfortunately, no significant difference in background noise was seen between perdeuterated and native crystals. Even in data frames collected in the absence of a diffracting crystal, there is ample evidence of high, variable background levels. I believe that high background was ubiquitously present throughout data collection, obliterating the possibility of interpreting the data as to the effect of protein perdeuteration.

Despite experimental difficulties (crystal misalignment, equipment failure, a fire (Fig. 9.1)), sufficient neutron diffraction data were collected to successfully determine the neutron structure of perdeuterated *Staphylococcal* nuclease to 2.4Å resolution. This was accomplished in part by the adaptation of two computer programs used in X-ray crystallography, MADNES and XPLOR. In particular, MADNES represents a great advance over the previous generation of neutron data processing software. Problems such as small crystal size, poor signal-to-noise, and missing reflections decreased the level of quality in the final data set and led to significant phase bias in the neutron maps. The combination of mid resolution data and phase bias affected structure analysis. Hydroxyl rotor and water orientation could not be defined and new insights concerning the enzyme's mechanism were not possible. On the other hand, the hydrogen/deuterium exchange pattern for crystallized Snase was successfully determined.

The most important biological result of this research has been the analysis of the effect of crystal packing on the hydrogen/deuterium exchange pattern, and thus flexibility, of Snase. Global effects of crystal contacts were observed through the comparison of the solution and crystal H/D patterns. The rate of exchange for amide protons is significantly lengthened in a crystal; however, the similarities between the neutron and NMR exchange patterns suggest that the crystal environment does not alter the three-dimensional structure of the protein. In addition, evidence for local changes was found in a protected loop and in the C-terminal helix. In both cases, hydrogen atoms

involved in intramolecular hydrogen bonds were indirectly protected from exchange by crystal contacts, in contrast to what was seen in solution. Thus, while crystal packing does not affect the overall structure, it does decrease the overall flexibility of the protein and has a major influence on the dynamics of local regions involved in these contacts.

Fig. 9.1: Cover of Newsday, April 1, 1994

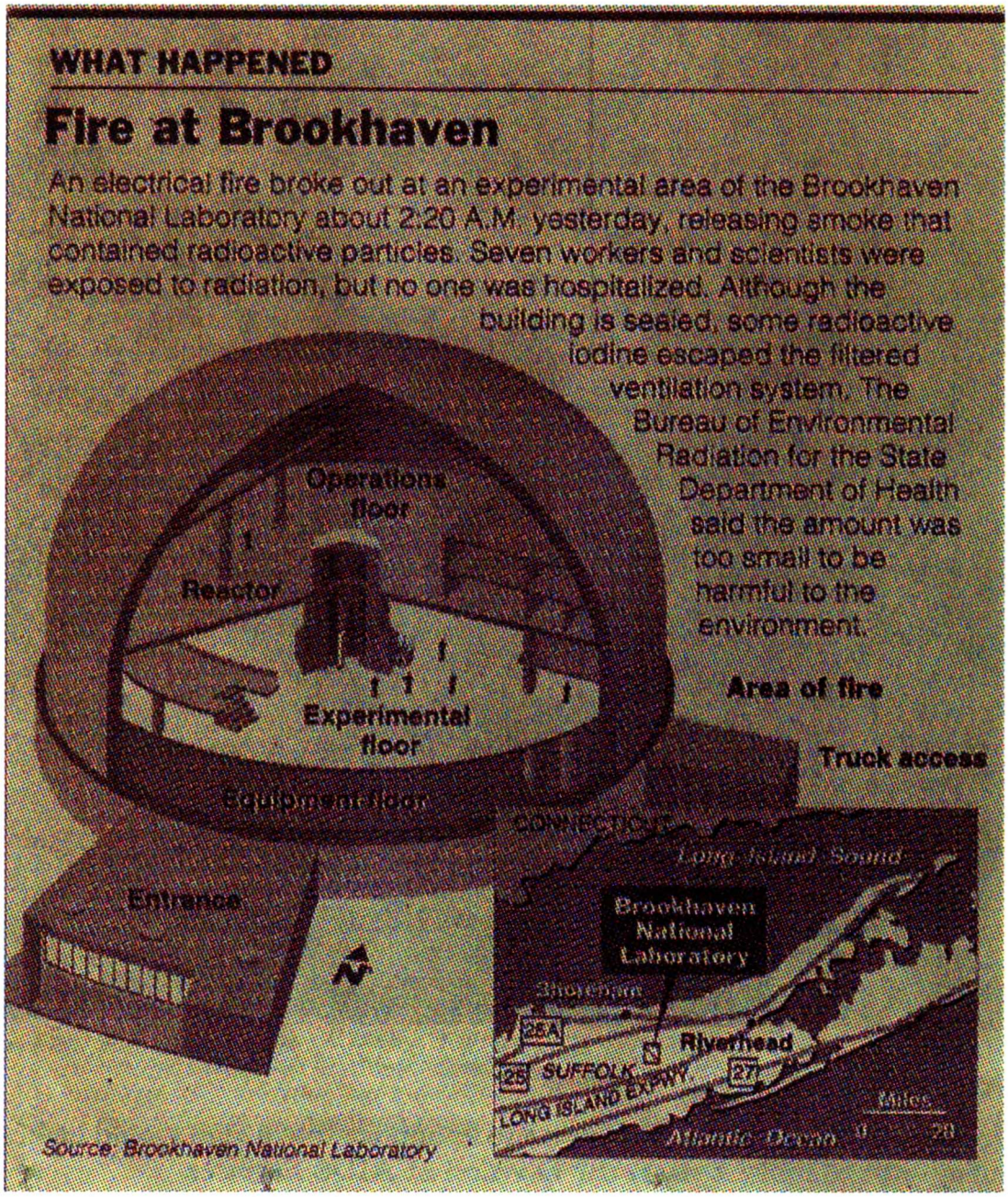


Fig. 9.1b: Fire at Brookhaven

Appendix A

Probing the Mechanism of *Staphylococcal* Nuclease with Unnatural Amino Acids: Kinetic and Structural Studies

Probing the Mechanism of Staphylococcal Nuclease with Unnatural Amino Acids: Kinetic and Structural Studies

J. Kevin Judice, Theresa R. Gamble, Elizabeth C. Murphy, Abraham M. de Vos, Peter G. Schultz*

Staphylococcal nuclease is an enzyme with enormous catalytic power, accelerating phosphodiester bond hydrolysis by a factor of 10^{16} over the spontaneous rate. The mechanistic basis for this rate acceleration was investigated by substitution of the active site residues Glu^{43}, Arg^{35}, and Arg^{87} with unnatural amino acid analogs. Two Glu^{43} mutants, one containing the nitro analog of glutamate and the other containing homoglutamate, retained high catalytic activity at pH 9.9, but were less active than the wild-type enzyme at lower pH values. The x-ray crystal structure of the homoglutamate mutant revealed that the carboxylate side chain of this residue occupies a position and orientation similar to that of Glu^{43} in the wild-type enzyme. The increase in steric bulk is accommodated by a backbone shift and altered torsion angles. The nitro and the homoglutamate mutants display similar pH versus rate profiles, which differ from that of the wild-type enzyme. Taken together, these studies suggest that Glu^{43} may not act as a general base, as previously thought, but may play a more complex structural role during catalysis.

The proposed mechanism of catalysis by staphylococcal nuclease (SNase) is based on high-resolution x-ray crystal structures (*1*, *2*) as well as a series of mechanistic studies (*3*). Glu^{43} is postulated to act as a general base to activate a water molecule for attack on the phosphodiester bond, resulting in a trigonal bipyramidal transition state that is stabilized by interactions with Arg^{87}, Arg^{35}, and the Ca^{2+} ion (Fig. 1). Arg^{87} may also act as a general acid to protonate the 5′-hydroxyl leaving group. Although mutagenesis studies have shown that these residues are important in catalysis, the structural bases for their functions have not yet been clearly defined, largely because substitutions with other naturally occurring amino acids not only reduce catalytic activity, but also alter the protein structure. For example, substitution of Glu^{43} with aspartate leads to a 300-fold reduction in catalytic efficiency (*4*) as well as structural perturbations in the loop adjacent to Glu^{43} (*5*, *6*). Likewise, mutation of either Arg^{87} or Arg^{35} to lysine results in large decreases in catalytic efficiency and changes in protein structure (*7*).

In order to more precisely evaluate the role of Glu^{43} during turnover, we substituted this residue with its nitro analog, S-4-nitro-2-aminobutyric acid (NABA, Fig. 2). Although the nitro group is both isoelectronic and isosteric to the carboxylate group, it is a much poorer base; the pK_a's of the conjugate acids CH_3CO_2H and $CH_2NO_2H^+$ are 4 and −12, respectively (*8*). The geometric requirements for catalysis were probed by comparison of the homologous series homoglutamate (ADPA), glutamate, and aspartate. The Glu^{43}→S-2-amino-5-hydroxypentanoic acid (AHPA) mutant was also generated, in which the carboxylate group of Glu^{43} was substituted with the weakly basic hydroxymethylene group ($pK_a \approx -3$) (*8*). The catalytic roles of Arg^{35} and Arg^{87} were investigated by replacement of these residues with aminoethylhomocysteine (AEHC) and citrulline (CIT). These substitutions were designed to assess the relative contributions of both electrostatic interactions and monodentate as compared with bidentate hydrogen bonding ability to catalysis.

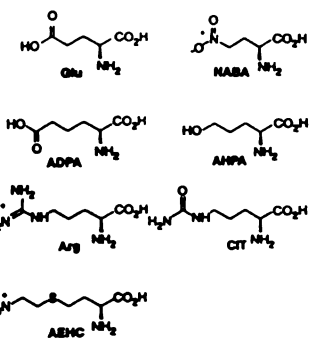


Fig. 2. Amino acids incorporated at the active site of staphylococcal nuclease (*27*).

Incorporation of unnatural amino acids into SNase was accomplished by in vitro suppression of nonsense mutations (TAG) with a chemically aminoacylated suppressor tRNA (*9*, *10*). In vitro expression of SNase from the plasmid pKJSN1, in which the gene for wild-type SNase is under the transcriptional control of the T7 promoter (*11*, *12*), afforded approximately 20 μg of protein per milliliter of reaction mixture with wild-type activity. Addition of suppressor tRNAs aminoacylated (*10*) with the unnat-

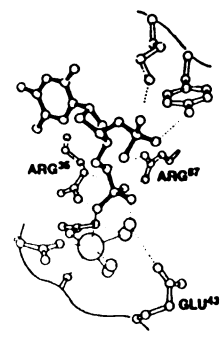


Fig. 1. The active site of staphylococcal nuclease complexed with Ca^{2+} (large sphere) and the inhibitor thymidine 3′,5′-bisphosphate (black) (*2*). Also shown are two water molecules (overlapping small spheres) thought to be candidates for the attacking nucleophile.

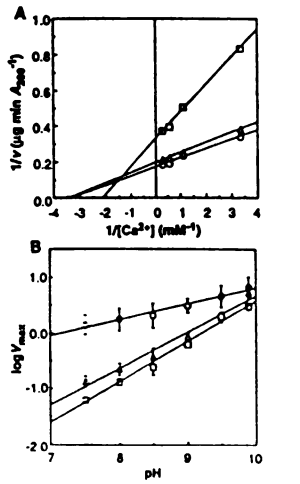


Fig. 3. (A) Lineweaver-Burk plots used to determine apparent V_{max} and $K_m^{Ca^{2+}}$ values of wild-type SNase (○) and the Glu^{43}→NABA (□) and Glu^{43}→ADPA (△) mutants. (B) The pH dependence of the apparent V_{max} for wild type SNase (○) and the Glu^{43}→NABA (□) and Glu^{43}→ADPA (△) mutants.

J. K. Judice, E. C. Murphy, P. G. Schultz, Department of Chemistry, University of California, Berkeley, CA 94720
T. R. Gamble, Graduate Group in Biophysics, University of California, San Francisco, San Francisco, CA 94143, and Department of Protein Engineering, Genentech, Inc., South San Francisco, CA 94080
A. M. de Vos, Department of Protein Engineering, Genentech, Inc., South San Francisco, CA 94080
*To whom correspondence should be addressed.

ural amino acids to in vitro protein synthesis reactions programmed with the Arg^{15}, Arg^{87}, or Glu^{43}→nonsense mutant plasmids (*13*) produced mutant proteins with suppression efficiencies between 15 and 40%. Mutant proteins were initially screened for catalytic activity with the use of a chromogenic plate assay (*14*). In contrast to the previously studied mutants at position 43, both the Glu^{43}→NABA and Glu^{43}→ADPA mutants were found to have high hydrolytic activity. The catalytic efficiency of the Glu^{43}→AHPA mutant was reduced by a factor of at least 10^3 relative to that of the wild-type enzyme and was not further studied.

The Glu^{43}→NABA and Glu^{43}→ADPA proteins were purified to homogeneity by ion exchange chromatography (*15*), and their kinetic properties at pH 9.9 were determined (*16*) by spectrophotometric assay (*17*) (Fig. 3A). The kinetic constants of the Glu^{43}→NABA [V_{max} = 3.0 ± 0.5 A_{260} min^{-1} μg^{-1}, Michaelis constant for DNA (K_m^{DNA}) = 26 ± 8 μg/ml] and Glu^{43}→ADPA (V_{max} = 5.2 ± 0.2, K_m^{DNA} = 10 ± 2) were markedly close to those determined for wild-type SNase (V_{max} = 6.7 ± 0.7, K_m^{DNA} = 8 ± 3). Moreover, the K_m values for Ca^{2+} in all three proteins were similar [$K_m^{Ca^{2+}}$ (wild type) = 320 ± 30 μM, $K_m^{Ca^{2+}}$ (Glu^{43}→ADPA) = 290 ± 30, $K_m^{Ca^{2+}}$ (Glu^{43}→NABA) = 470 ± 40], suggesting that there were no significant changes in the Ca^{2+} binding sites of the mutants. The pH dependence of the V_{max} values of both the nitro and homoglutamate mutants was similar and differs from that of the wild-type enzyme (Fig. 3B); for both mutants, V_{max} was proportional to $log[OH^-]^{0.6}$, whereas for the wild-type enzyme V_{max} was proportional to $log[OH^-]^{0.3}$.

The high catalytic activity of the Glu^{43}→NABA mutant was surprising, in light of the poor basicity of this residue and the low activity of the structurally similar Glu^{43}→Gln mutant (*4*). The possibility of a nitronate anion participating as a general base was considered (pK_a = 9 for $CH_3CH_2NO_2$→$CH_3CHNO_2^-$) (*18*), but the similarity of the pH versus rate profiles

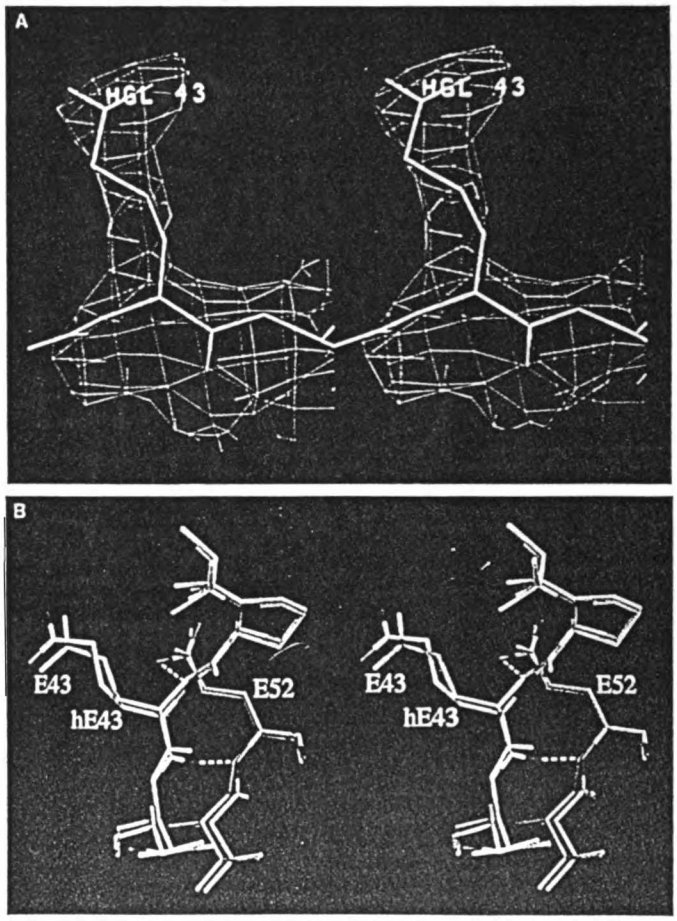


Fig. 4. (**A**) Difference electron density for residues 43 and 44, with the final model superimposed for comparison. Until this stage in the structure determination, loop residues 43 to 50 had been excluded from the model. $ADPA^{43}$ SNase was crystallized under conditions similar to those previously described (*2*), except that the protein concentration was 1.5 mg/ml and the MPD concentration in the reservoir solution was 40% (w/v). Diffraction data to 2.4 Å resolution were collected from five crystals ($P4_1$, a = 48.3 Å and c = 64.0 Å) with the use of a MAR image plate system at the Stanford Synchrotron Radiation Laboratory. The 23,935 measurements of 5960 unique reflections were merged with R = 8.9%, resulting in a data set that is 98% complete between 8 and 2.4 Å. Inspection of electron density maps calculated after rigid-body refinement with the inhibited wild-type structure (*2*) as the starting model revealed that no inhibitor was present in the active site, that the conformations of the active site residues were as seen in apo (*19*) rather than in inhibited SNase, and that density for loop residues 43 to 50 was poor. The model was adjusted accordingly, and cycles of refinement with XPLOR (*28*) and manual adjustment using FRODO (*29*), followed by introduction of water molecules, resulted in a final R factor of 18.7% (8 to 2.4 Å, 5533 reflections with $F > 0$). The final model consists of residues 6 to 45 and 51 to 141, together with 65 water molecules, and has good stereochemistry, as evidenced by rms deviations in bond lengths and angles of 0.008 Å and 1.7°, respectively. The final electron density map shows good density for all residues, with the exception of the termini and the loop residues 44 to 50. Residues 44 and 45 were placed into poor density, and residues 46 to 50 could not be built into the model, although there are patches of density consistent with the apo wild-type structure. The strongest features in the final difference map are four to five rms peaks corresponding to the phosphate moieties of the bound inhibitor in wild-type SNase. It is unclear whether these represent low-occupancy inhibitor or bound phosphate ions; nothing was built into this density. (**B**) Comparison of $ADPA^{43}$ SNase (residue labeled hE43, carbon atoms in white, nitrogen atoms in blue, and oxygen atoms in red) to the apo wild-type structure (all atoms in yellow) (*19*) in the region of residue 43. The torsion angles for the side chain of $ADPA^{43}$ are −88°, 111°, 80°, and −151°. The hydrogen bonds shown in this figure are found between O of 43 and N of 52 (2.68 Å) and between N of 43 and $O_{\epsilon}1$ of 52 (2.77 Å).

of the nitro and homoglutamate mutants makes this unlikely. Thus, it appears that a basic residue at position 43 is not required for high catalytic activity; only a bidendate hydrogen bond acceptor (a role not fulfilled by either glutamine or AHPA) is necessary.

The homoglutamate mutant also retains wild-type catalytic activity at pH 9.9, in contrast to the Glu^{43}→Asp mutant, in which V_{max} is reduced 300-fold. Both substitutions conserve charge but alter side chain geometry and steric interactions. In order to understand the mechanism by which the enzyme accommodates the additional methylene group in the homoglutamate side chain, we determined the x-ray crystal structure of this mutant to 2.4 Å resolution (Fig. 4). The structure of $ADPA^{43}$ SNase is very similar to that of wild-type apo, with a root-mean-square (rms) difference between backbone atoms of 0.2 Å. The largest differences (0.4 to 0.8 Å) occur in residues 42 to 51, 84 to 86, and 113 to 115. Differences in residues 42 to 51 and 113 to 115 have been observed between wild-type apo and inhibited structures (*19*), implying inherent flexibility in these regions. The extra methylene group of the homoglutamate side chain is accommodated by a 0.4 Å shift in the backbone of residue 43, accompanied by an adjustment of its side chain torsion angles (Fig. 4B). The net effect of these differences is that the homoglutamate carboxylate group moves to within 0.5 Å of the Glu^{43} carboxylate group in the apo wild-type structure, while maintaining a similar orientation with respect to the active site. This suggests that whatever role the side chain of Glu^{43} plays during catalysis can be fulfilled by the homoglutamate side chain. Furthermore, two hydrogen bonds between $ADPA^{43}$ and Glu^{52} are maintained (Fig. 4B). The Glu^{43} and Glu^{52} residues shift down slightly, but maintain the hydrogen bond between the side chain of Glu^{52} and the main chain amide of Glu^{43}, an interaction that may be important in preorganizing the adjacent Ω-loop (residues 43 to 52) for function during binding or catalysis.

Although both the nitro and homoglutamate mutants retain high activity at pH 9.9 (where the enzyme is typically assayed), they are less active than the wild-type enzyme at lower pH values (Fig. 3B). The difference in pH behavior may reflect different rate-determining steps in the mutant and wild-type enzymes. Recent studies (*20*) have shown a viscosity dependence on V_{max} for the wild-type enzyme above pH 7.3, suggesting that product release is rate limiting under these conditions. In contrast, a chemical step may be rate limiting in the two mutants described here (*21*). The pH dependence of V_{max} observed for the NABA and ADPA mutants is qualitatively similar to those reported for the Glu^{43}→Asp mutant (although the absolute activity of this mutant is reduced 300-fold) and a mutant in which residues 44 to 49 have been deleted (*20*). This similarity suggests that differences in loop geometry may be responsible for the observed decreases in rate. The hydrogen bonds between the carboxylate group of residue 43 and the main chain amides of residues 45 and 46 in the inhibited wild-type structure (*2*) may be important for fixing the loop into optimal position for catalysis. In the Glu^{43}→Asp mutant, these hydrogen bonds were absent, and the structure of the loop was substantially altered (*5*). The nitro analog of glutamate has only weak hydrogen-bonding properties (*22*), and the position of the homoglutamate carboxylate group differs from that of the wild type by 0.5 Å. Therefore, in each case the necessary hydrogen bonds may not be formed effectively.

Substitution of either Arg^{35} or Arg^{87} with the charged, monodentate hydrogen-bonding analog, AEHC, led to at least a 10^3-fold loss in catalytic activity at pH 7.5. The ammonium ion of AEHC (in contrast to that of Lys) should be able to occupy very nearly the same position as the terminal guanidinium-NH_2 group of Arg in the enzyme active site. Although the pK_a of AEHC is predicted to be one unit lower than that of lysine (*23*), it should be present predominantly in the charged form under the assay conditions. In the absence of any structural perturbations, these results, and the fact that the neutral isosteric CIT mutants also had 10^3-fold decreased activity, suggest that neither Arg^{35} nor Arg^{87} functions via simple electrostatic stabilization of the transition state or by protonation of the 5′-hydroxyl group of the substrate. It appears that both charge and bidentate hydrogen-bonding ability are critical features of these residues, consistent with a previous mechanistic proposal that both guanidinium groups shift to bidentate interactions in the transition state (*24*).

REFERENCES AND NOTES

1. F. A. Cotton, E. E. Hazen, J. J. Legg, *Proc. Natl Acad. Sci. U.S.A.* **76**, 2551 (1979).
2. P. J. Loll and E. E. Lattman, *Proteins* **5**, 183 (1989).
3. S. Mehdi and J. A. Gerlt, *J. Am. Chem. Soc.* **104**, 3223 (1982). B. M. Dunn, C. DiBello, C. B. Anfinsen, *J. Biol. Chem.* **248**, 4769 (1973). E. H. Serpersu, D. Shortle, A. S. Mildvan, *Biochemistry* **26**, 1289 (1987). D. J. Weber, A. K. Meeker, A. S. Mildvan, *ibid.* **30**, 6103 (1991). D. J. Weber, E. H. Serpersu, D. Shortle, A. S. Mildvan, *ibid.* **29**, 8632 (1990). E. H. Serpersu, D. Shortle, A. S. Mildvan, *ibid.* **25**, 68 (1986).
4. D. W. Hibler *et al.*, *Biochemistry* **26**, 6278 (1987).
5. P. J. Loll and E. E. Lattman, *ibid.* **29**, 6866 (1990).
6. J. A. Wilde *et al.*, *ibid.* **27**, 4127 (1988).
7. T. Poorman-Allen, M. Del Acqua, J. A. Gerlt, S. M. Stanczyk, P. H. Bolton, *ibid.* **29**, 3677 (1990).
8. E. M. Arnett, *Prog. Phys. Org. Chem.* **1**, 223 (1963).
9. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **244**, 182 (1989).
10. J. A. Ellman, D. Mendel, C. J. Noren, S. J. Anthony-Cahill, P. G. Schultz, *Methods Enzymol.* **202**, 301 (1991).
11. The polymerase chain reaction was used to delete the *ompA* leader sequence and to introduce an Nde I restriction site at the 5′-terminus of the SNase structural gene encoded in the plasmid pONF1 [M. Takahara *et al.*, *J. Biol. Chem.* **260**, 2670 (1985)] and a Xho I site at the 3′ end of the structural gene. The Nde I–Xho I fragment was then subcloned into pSAL127 (*12*), placing the SNase gene under the transcriptional control of the T7 promoter in the resulting vector, pKJSN1.
12. S. A. Lesley, Promega Corp., 2800 Woods Hollow Rd., Madison, WI 53711, unpublished results.
13. An Eco RI digest of pKJSN1 provided a 589–base pair fragment containing the gene for SNase and the T7 promoter region. This fragment was subcloned into M13mp18, and mutagenesis was done according to the Eckstein protocol [J. R. Sayers, W. Schmidt, F. Eckstein, *Nucleic Acids Res.* **16**, 791 (1988)]. Each mutant gene was sequenced in its entirety. The mutant SNase genes along with the T7 promoter were then cloned into the unique Eco RI site in pUC18 to provide pUCSN.E43am, pUCSN.R87am, and pUCSN.R35am.
14. A chromogenic plate assay [D. Shortle, *Gene* **22**, 181 (1983)] was used to screen for catalytically active mutants. In our experiments, 2 μl of the cleared supernatants from a 30-μl in vitro protein synthesis reaction were spotted on a plate containing 1% agar, 1% NaCl, 50 mM sodium glycinate (pH 9.9), 20 mM $CaCl_2$, single-stranded calf thymus DNA (100 μg/ml), and toluidine blue O (0.2 mg/ml). After incubation for 2 hours at 37°C, catalytically active SNase was identified by the presence of a pink halo against the blue background. When the suppressor tRNA was omitted from or supplied in unacylated form to the in vitro reaction, no full-length protein was produced (data not shown), and no catalytic activity could be observed. In experiments where mock in vitro reactions (containing all of the components except the expression vector) were spiked with known amounts of purified SNase, amounts of wild-type SNase as low as 5 pg/μl produced a detectable halo. Upper limits on readthrough or mutant catalytic activities were then calculated on the basis of an expression level of 20 μg/ml. From this analysis, we estimate that as little as 0.025% readthrough of the TAG codon that resulted in a protein of wild-type activity could have been detected. These controls demonstrate that the in vitro system does not contain endogenous suppressor tRNAs capable of reading through the amber stop codon and that the aminoacyl tRNA synthetases present in the *Escherichia coli* S-30 extract do not aminoacylate the suppressor $tRNA_{CUA}$ with any of the 20 natural amino acids.
15. A 5-ml in vitro protein synthesis reaction was treated with 100 μl of polyethylenimine hydrochloride (pH 7.5) and centrifuged. The cleared supernatants were loaded onto a carboxymethyl-cellulose cation exchange column with 50 mM tris-HCl (pH 7.5) and eluted with a 0 to 2 M NaCl gradient in that buffer. Fractions containing nuclease activity were lyophilized, dialyzed against 50 mM sodium acetate (pH 4.9), 1 mM EDTA, then loaded onto a Mono-S column (Pharmacia) and eluted with a 0 to 1 M NaCl gradient. Fractions containing nuclease activity were pooled and determined to be more than 95% homogeneous by silver-stained SDS–polyacrylamide gel electrophoresis (PAGE) analysis.
16. The assay mixture contained 40 mM sodium glycinate (pH 9.9), single-stranded calf thymus DNA (75 μg/ml), and the concentration of Ca^{2+} was varied between 0.2 and 3.6 mM in a volume of 1.0 ml at 25°C. The assay mixture was determined to contain [illegible] amounts of DNA by the [illegible] acid–precipitable ratios of 2%

and 90 μg per milliliter of DNA. We used similar plots to determine the apparent V_{max} and K_m^{DNA} values by varying the concentration of DNA from 7 to 50 μg/ml in 40 mM sodium glycinate (pH 9.9) with 10 mM Ca^{2+} (data not shown). Reactions were initiated by the addition of 20 ng of protein, and we determined the initial velocities by monitoring the increase in absorbance at 260 nm (A_{260}). The numbers reported represent the standard deviations of three data collections obtained for each protein on the same day. Protein concentrations were initially determined by SDS-PAGE analysis of ^{35}S-methionine–labeled samples of wild-type and mutant SNase. The wild-type band was excised and quantitated by scintillation counting; the remaining wild-type sample was assayed for catalytic activity, and the known specific activity was used to determine the concentration (in micrograms per milliliter) of wild-type SNase present. Quantitation of the gel slices containing mutant proteins by scintillation counting was followed by catalytic assay to provide the specific activity of each mutant in units of ΔA_{260} min^{-1} μg^{-1} SNase. For the pH versus rate profiles, assays were done as described above, at 10 mM Ca^{2+} and at DNA concentrations ranging from 6 to 80 μg/ml. Between pH 7.5 and 8.5, 40 mM tris-HCl was used as the buffer; from pH 8.5 to 9.9, assay mixtures were buffered with 40 mM sodium glycinate. Duplicate data points were gathered for each concentration of DNA at each pH; points on the pH versus rate plot represent the average value of two (NABA and ADPA) or three (wild type) of such sets of data. The lines were fit by linear least squares analysis; no weighting was performed.
17. P. Cuatrecasas, S. Fuchs, C. B. Anfinsen, *J. Biol. Chem.* 242, 1541 (1967).
18. C. D. Slater, *J. Org. Chem.* 46, 2173 (1981).
19. T. R. Hynes and R. O. Fox, *Proteins* 10, 92 (1991).
20. S. P. Hale, L. B. Poole, J. A. Gerlt, *Biochemistry*, in press.
21. Both mutants (Glu^{43}→NABA and Glu^{43}→ADPA) and the wild-type enzyme exhibited a significant kinetic isotope effect in D_2O, with V_{max}^{H}/V_{max}^{D} = 2 to 3. The solvent isotope effects observed here suggest that chemical steps contribute to the rate-determining step in both the wild type and the mutants; chemistry and loop movement may be coupled, as was recently observed for triosephosphate isomerase (*25*). Kinetic isotope effects were determined as described [R. L. Schowen, in *Isotope Effects on Enzyme Catalyzed Reactions*, W. W. Cleland, M. H. O'Leary, D. B. Northrop, Eds. (University Park Press, Baltimore, MD, 1977), pp. 64–99]. Concentrated buffer stocks (400 mM sodium glycinate) were prepared by addition of solid glycine to H_2O or D_2O and adjustment of the pH with 6 N NaOH (in D_2O, pD = pH + 0.4); the final H_2O content in the D_2O buffers was less than 10% (v/v) after this procedure. Assay buffers were then prepared by dilution of stock solutions into H_2O or D_2O. Values of V_{max} were obtained from Lineweaver-Burk plots of two sets of kinetic data for each protein: for pH or pD values from 9.4 to 10.4, assay buffers contained 40 mM NaGly, 10 mM Ca^{2+}, and 10, 20, 40, or 80 μg of DNA per milliliter; for pD values above 10.4, buffers contained 40 mM NaGly, 10 mM Ca^{2+}, and 80, 100, or 120 μg of DNA per milliliter. The kinetic isotope effects were estimated from the differences in maximal values of V_{max}, which were observed at pH 9.9 in H_2O and pD 10.4 (Glu^{43}→NABA) or 10.6 (wild type and Glu^{43}→ADPA) in D_2O. By monitoring the intrinsic fluorescence of Trp^{140} [D. Shortle and A. M. Meeker, *Proteins* 1, 81 (1986)] as a function of pH or pD, we have determined the midpoint for denaturation of wild type SNase to be at pH 10.4 in H_2O and pD 11.2 in D_2O. As such, we do not believe that denaturation effects complicate our analyses.
22. M. D. Joesten and L. J. Schaad, *Hydrogen Bonding* (Dekker, New York, 1974), pp. 309–335.
23. F. Hermann and K. Lemke, *Hoppe Seylers Z. Physiol. Chem.* 349, 390 (1968).
24. D. J. Weber *et al.*, *Proteins* 13, 275 (1992).
25. N. S. Sampson and J. R. Knowles, *Biochemistry* 31, 8488 (1992).
26. D. J. Such, thesis, University of California, Berkeley (1993).
27. CIT and ADPA (L-α-aminoadipic acid) (both from Sigma) were protected for chemical aminoacylation reactions as described (*10*). AEHC was obtained from the reaction of *N*-NVOC (nitroveratryloxycarbonyl)-1-bromo-2-aminoethane with the sodium salt of *N*-NVOC-L-homocysteine. NABA and AHPA were first synthesized by D. J. Such (*26*).
28. A. T. Brunger, *XPLOR Manual* (Yale Univ. Press, New Haven, CT, version 3.1, 1992).
29. A. T. Jones, *J. Appl. Crystallogr.* 11, 268 (1978).
30. Plasmid pONF1 was provided by J. A. Gerlt, University of Maryland. We thank D. Mendel for assistance in developing the purification protocol for SNase, A. Kossiakoff for helpful discussions, and J. Gerlt for comments on this manuscript and for sharing data before publication. We are grateful for financial support of this work from the National Institutes of Health (grant ROI GM-49220). J. K. J. was supported by an NIH Postdoctoral Fellowship (GM 14012-02S1) and T. R. G. by an NIH Biotechnology Training grant (T32GM-08388). P. G. S. is a W. M. Keck Foundation Investigator.

16 February 1993; accepted 15 July 1993

Appendix B

The Production and X-ray Determination of Perdeuterated *Staphylococcal* Nuclease

ELSEVIER

Biophysical Chemistry 53 (1994) 15–26

Biophysical Chemistry

The production and X-ray structure determination of perdeuterated *Staphylococcal* nuclease

Theresa R. Gamble [a,*], Karl R. Clauser [b], Anthony A. Kossiakoff [a]

[a] *Graduate Group in Biophysics, University of California, San Francisco, San Francisco, CA 94143-0448, USA*

[b] *Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143-0446, USA*

Received 22 October 1993; revised 15 December 1993; accepted 20 December 1993

Abstract

Staphylococcal Nuclease (SNase) has been chosen as a model protein system to evaluate the improvement in neutron diffraction data quality using fully perdeuterated protein. Large quantities of the protein were expressed in *Escherichia coli* grown in medium containing deuterated amino acids and deuterated water (D_2O) and then purified. The mean perdeuteration level of the non-exchangable sites in the protein was found to be 96% by electrospray ionization mass spectrometry. The perdeuterated enzyme was crystallized and its X-ray structure determined. Crystals of perdeuterated SNase have been grown to 1.5 mm^3. Crystallization conditions, space group and cell parameters were found to be the same for both native and perdeuterated forms of the protein. Comparison of these two forms of SNase revealed no significant structural differences between them at the atomic resolution of 1.9 Å. Data collection using crystals of the perdeuterated protein is scheduled at the Brookhaven High Flux Beam Reactor.

Keywords: Protein perdeuteration; Neutron diffraction; *Staphylococcal* nuclease; Electrospray ionization mass spectrometry

1. Introduction

Neutrons have been successfully employed to study a broad range of biological problems inaccessible by other experimental techniques. When one assesses the literature it is clear that the neutron scattering applications to solution and semi-ordered systems are disproportionately represented when compared to crystallographic studies. This is even more striking considering the fact that neutron diffraction through its ability to locate hydrogen and deuterium atoms can provide a wealth of structural information that far surpasses that of its X-ray counterpart. The paucity of neutron diffraction structures stems from two principal experimental difficulties: (1) the inherently low fluxes that can be generated at the current reactor facilities; and (2) the incoherent scattering effects produced by hydrogens in the sample. The resulting background levels are so high that the dynamic range of the measured intensities is small, which greatly compromises the quality of the data.

To a degree, the flux problems are addressed by the development of new instrumentation and especially the use of sensitive two-dimensional detectors. Traditionally the magnitude of the background due to the hydrogen atoms has been addressed by exchanging all the waters of crystallization for D_2O and using extremely large crystals. However, there are practical limitations to reducing the backgrounds; even if all the waters of crystallization are exchanged, the nonexchangeable

* Corresponding author.

SSDI 0301-4622(94)00072-R

hydrogens, which make up about 50% of the total atoms in the structure, remain. In addition, because of the buildup of surface deformations inherent in crystal growth, rarely can neutron sized crystals (> 3 mm^3) be grown. Limiting neutron diffraction analyses to only those few systems where large crystals can be obtained, and conversely not being able to study systems where the placement of hydrogens could answer fundamental biological issues, presents a situation that is unacceptable.

In the absence of some new advance that can help circumvent these problems neutron diffraction will never reach its full potential. An approach being undertaken in this lab is to use a technique long employed by the neutron low angle scattering community, protein perdeuteration. The rationale is that the incoherent background scatter can be effectively eliminated by replacing biosynthetically all nonlabile hydrogens by deuteriums, which themselves do not have a significant incoherent term. We note, however, that the elimination of the incoherent scattering is very sensitive to the absolute degree to which the protein and the rest of the crystal (mother liquor, inhibitors, etc.) is perdeuterated. For instance, a fully perdeuterated system can offer as much as a 40-fold reduction in sample derived background, however, a 5% contamination of hydrogen into the crystal system reduces the improvement to only 12-fold. Thus the degree to which the backgrounds are actually reduced must wait for experimental verification and will vary between protein crystal systems.

Production of significant amounts of perdeuterated protein is possible because of the availability of robust *E. coli* expression systems that allow for high levels of protein expression in deuterated media. In the following report we describe the production of 96% perdeuterated *Staphylococcal* nuclease (SNase), its crystallization and a comparison to the diffraction data collected from the native material. SNase is a calcium dependent, extracellular enzyme consisting of a single polypeptide chain of 149 residues [1]. It is currently one of the most intensely studied proteins because of the interest in using it as a paradigm for studying structure-function issues in protein folding [2–4]. One particular research area that will be pursued in this study involves a number of unresolved issues concerning its mechanism of action; location of specific hydrogens in the catalytic site could provide crucial insight. Additionally, there is the possibility, given the diffraction quality of the crystals, that we will be able to observe most of the deuterium atoms in the well ordered regions of the protein.

Table 1
Growth media

Rich Media (25 ml) [a]	Algal Media (25 ml)
0.6 g yeast extract	0.25 g algal whole hydrolysate [b] or Celtone powder [c]
0.3 g tryptone	
0.06 g KH_2PO_4	0.06 g KH_2PO_4 [e]
0.3 g K_2HPO_4	0.3 g K_2HPO_4 [e]
0.1 ml glycerol	0.1 ml glycerol [d]
H_2O to 25 ml	D_2O to 25 ml

[a] Scaled down version of SB media (personal communication, Alan Meeker, The Johns Hopkins University School of Medicine, Baltimore, MD).
[b] Can be purchased in two forms: Algal whole hydrolysate-hydrogenated, Algal whole hydrolysate-deuterated (10% H_2O solution, 98 atom % D) (MSD Isotopes, Division of Merck Frost Canada Inc., Montreal, Canada).
[c] Can be purchased in two forms: Celtone-U (hydrogenated) Celtone-D (^{2}H, 97% +) (Martek Corporation, Columbia, MD).
[d] Can be purchased in a deuterated form: Glycerol-d_8 (98.7 atom% D) (MSD Isotopes, Division of Merck Frost Canada Inc., Montreal, Canada).
[e] Made up as stock solution in D_2O when appropriate.

2. Methods

2.1. Protein production

Escherichia coli strain AR120 containing a pAS1 plasmid was used for the overproduction of SNase [1]. Medium containing deuterated amino acids, deuterated glycerol, potassium phosphate and deuterated water (D_2O) was developed based on a recipe for rich medium (see Table 1). Deuterated amino acids and their hydrogen containing counterparts were purchased from two sources (see Table 1) and could be directly substituted for one another. *E. coli* was grown in rich medium for 12–18 h at 37°C to produce native SNase (personal communication, A.K. Meeker). The same bacteria were grown in deuterated medium, first for 24

[1] Obtained from A. Meeker and D. Shortle, Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD.

h at 30°C, and then for another 24 h at 37°C to produce deuterated SNase. For both culture types, 250 ml of medium was inoculated with 1 ml of *E. coli* grown overnight in rich medium. Cell growth was characterized by cell density at OD_{600} using a Uvikon 860 (Research Instruments International, San Diego, CA).

2.2. Protein purification

The purification method used for both native and deuterated SNase was based on a previous SNase protocol [5] that was further developed by Shortle and Meeker (personal communication). Cells were harvested after the appropriate growth time and a cell pellet was collected by centrifugation (3000×*g*, 4°C, 10 min). The pellet was suspended in urea extraction buffer No. 1 (6 M urea/ 25 mM Tris-HCl/5 mM EDTA, pH 8.1) and gently swirled on a rocker box in the cold room for 20 min. The extracted cells were centrifuged at 3000×*g*, 4°C for 10 min. The supernatant was poured off and discarded; the pellet is loose at this stage. The pellet was resuspended in urea extraction buffer No. 2 (same as above plus 200 mM NaCl) and gently rocked for 30 min, 4°C. Cells were centrifuged at 10 000×*g*, 4°C for 15 min and the pellet was discarded. An equal volume of ice-cold ethyl alcohol (200 proof) was added to the supernatant, and incubated undisturbed for 3–5 hours at −20°C. The precipitate that formed at this stage was collected by centrifugation at 10 000×*g*, 4°C for 15 min and discarded. An equal volume of ice-cold ethyl alcohol (200 proof) was again added and the entire mixture was incubated, undisturbed, at −20°C for 30 min. The resulting precipitate was collected by centrifugation at 2500 RCF, 4°C for 10 min. The supernatant was discarded and the pellet was drained briefly by inversion and then resuspended in 10 ml column load buffer (6M urea/1 mM EDTA/ 25 mM Tris-HCl/pH 8.1) by gently shaking the solution on a rocker box at 4°C overnight.

The protein, now contained in the suspension, was loaded onto a 1 ml Pharmacia Fast Flow S-Sepharose column equilibrated with column load buffer and then rinsed with 10 bed volumes of the same. Protein was eluted by slowly adding 3 bed volumes of column load buffer plus 70 mM NaCl, followed by 3 bed volumes of column load buffer plus 200 mM NaCl. The protein was precipitated by adding 1/20th volume of 1 M Tris-HCl, pH 7.0, and 3 volumes of ice-cold ethanol (200 proof) and incubated at −20°C for 20 min. The precipitate was collected by centrifugation at 2000 RCF, 4°C for 10 min. At this point the protein is a thick band on an 18% Tris-glycine, SDS PAGE gel (Novex, San Diego, CA). Reversed-phase (Vydac C18 Column, Alltech Associates Inc., Deerfield, IL) HPLC (Waters, division of Millipore, Milford, MA) (buffer A: H_2O, 0.1% trifluoroacetic acid, buffer B: Acetonitrile, 0.08% trifluoroacetic acid) was used to further purify the protein.

2.3. Mass spectrometry

The molecular weights of both native and deuterated SNase were measured by electrospray ionization mass spectrometry [6]. After the protein had been dialysed against the crystallization buffer (10.5 mM potassium phosphate, pH 8.15), the sample was desalted by dialysis, PD-10 Sephadex G-25M (Pharmacia, Piscataway, NJ) or by SEP-PAK® C18 cartridges (Waters, division of Millipore, Milford, MA) and concentrated under vacuum at 37°C (Savant SpeedVac Concentrator, Instruments, Inc. Framingdale, NY). Micromolar solutions (1–10 pmol/μl) of the protein in water/acetonitrile 50:50 (v/v) containing 0.5% acetic acid were directly infused at 1.5 μl/min into a Perkin-Elmer SCIEX API-III triple quadrupole mass spectrometer fitted with an Ionspray articulated nebulizer (SCIEX, Thornhill, Ontario, Canada). The mass spectrometer was operated with the Ionspray needle voltage at 4600 V, the interface plate at 650 V, and the orifice potential at 100 V. The third quadrupole was tuned and calibrated using a 10 pmol/μl solution of horse heart myoglobin (Sigma Chemical Co., St. Louis, MO). Mass spectra were recorded in multiple channel averaging mode, typically combining 5 scans. Data were collected every 0.1 mass units (m/z) with the quadrupole scanning from 600–1600 mass units in 20 s (2 ms dwell time).

2.4. Crystallization

Both native and deuterated SNase were crystallized as a complex with Ca^{2+} and the inhibitor thymidine-3',5'-diphosphate (pdTp) under conditions similar to those previously described by Loll and Lattman [7]. For each batch of protein, hanging drop conditions were optimized by varying the amount of precipitant (methyl pentanediol (MPD)) in the reservoir. The per-

centage of MPD that corresponded closest to single, sharp-edged crystals was used in the reservoir for sitting drop crystallizations. Large crystals were grown by setting up sitting drops ranging from 50 to 200 μl. All crystallization experiments were done at 4°C. Before neutron diffraction data are collected, crystals will be soaked in mother liquor containing 10.5 mM potassium phosphate (pH 8.15), 0.6 mM $CaCl_2$, 1.2 mM citric acid, 0.4 mM pdTp, 60% deuterated MPD (2-methyl-2,4-pentane-d12-diol, 99.0 atom% D, Isotec Inc, Miamisburg, OH), and D_2O.

2.5. X-ray structure determination

X-ray diffraction data were collected to 1.9 Å on both native and deuterated SNase crystals using an Enraf Nonius 'FAST' area detector (Delft, The Netherlands) mounted on a Rigaku RU200 X-ray generator (Toyko, Japan) operated at 55 kV, 90 mA. In each case, data were collected on a single crystal at two separate orientations. The crystal-to-detector distance was set at 50 mm and the swing angle (the angle between the direct X-ray beam and the center of the detector) was positioned at 25° corresponding to a maximum resolution of 1.8 Å. Both crystals were oriented with the 4-fold axis along the rotation axis, but no attempt was made to align them precisely. 180° of data were collected at $\chi = 0°$ in increments of 0.1° exposed for 90 s each. An additional 50° of data were collected in the same manner at $\chi = 90°$. The data were processed using MADNES [8] and PROCOR [9]. All refinement was carried out with the program XPLOR [10]. Both R and free R [11] values were monitored during the refinement process. The starting model for refinement was the Loll and Lattman structure for inhibitor-bound SNase [7]. The water molecules were not included. Rigid body refinement was performed using the same model against both native and deuterated data (10 to 2.5 Å). Positional refinement was done incrementally (10 to 2.5 Å, 10 to 2.0 Å, 10 to 1.9 Å) until all data had been included. At this stage, anisotropic scaling factors were calculated and applied to the data. Positional refinement and restrained individual temperature factor refinement was performed in a step wise fashion, alternating with manual corrections using FRODO [12] against Fo–Fc and 2Fo–Fc maps. Fo–Fc maps were inspected for water molecules, which were then added appropriately.

3. Results and discussion

3.1. Producing perdeuterated snase

Using the same expression system and purification scheme, but different medium and growth conditions as used to produce native SNase, milligram quantities of highly perdeuterated SNase can be generated.

Three steps were taken to develop optimal conditions for cell growth and protein expression in a deuterated environment. First, cell growth and protein production of native SNase was verified in rich medium. The system's lead time was less than 3 h, lag phase was reached within 10 h, the final cell density (OD_{600}) was between 1 and 3, and protein yield ranged from 80 to 100 mg/ℓ. Second, an alternative source of amino acids, one that could be purchased in both hydrogen containing and deuterated forms, was substituted for yeast extract in the rich medium. All parameters for cell growth and protein yield in this altered, hydrogen containing medium were optimized in a small volume (25 ml medium) and then scaled up (250 ml medium). Cell growth experiments indicated a range for the percentage of the amino acid substitute in the medium which enhanced cell growth (0.2 to 0.4 g/25 ml medium); outside this range cell growth was inhibited (data not shown). Inhibited cell growth at low percentages of amino acid substitute is assumed to be the result of cell starvation. At high percentages of amino acid substitute, cell growth may be hindered by lowered pH (medium pH decreased with increasing amounts of amino acid substitute) or by the presence of small amounts of toxin in the amino acid substitute. Grown at 37°C for 18 h on a large scale, cells in the altered

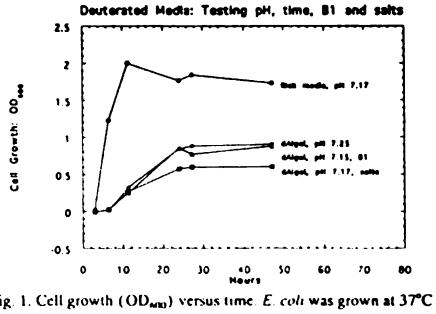


Fig. 1. Cell growth (OD_{600}) versus time. *E. coli* was grown at 37°C in both rich and deuterated media for 47 h.

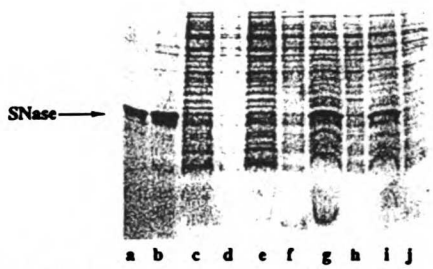


Fig. 2. SNase expression for *E. coli* grown for 24 h at 30°C and 37°C. The lanes shown are as follows: (a) native SNase standard; (b) cells used to inoculate culture, grown at 37°C in rich media; (c) cells grown in rich media, 30°C, 24 h; (d) cells grown in deuterated media, 30°C, 24 h; (g) cells grown in rich media, 37°C, 24 h; (h) cells grown in deuterated media, 37°C, 24 h. Lanes (e), (f), (i), and (j) are identical to (c), (d), (g), and (h) respectively, with the exception that excess ampicillin was added to the medium.

hydrogen containing medium grew to a final cell density which was typically 40% less than cells grown in rich medium; protein yield was decreased by the same amount.

In the final optimization step, all components of the altered, hydrogen containing medium were replaced with deuterated counterparts. Cells grown overnight at 37° in media containing varying amounts of amino acid substitute grew very poorly (final cell density was typically 90% less than cells grown in rich medium); however, the range for optimal growth was comparable to using hydrogen containing amino acids (0.2 to 0.3 g/25 ml medium). In order to improve cell growth, the deuterated medium was adjusted from pH 6 to pH 7 (equivalent to rich medium) and grown at 37°C, for 47 hours; final cell density (OD_{600}) was 50% of cells grown in rich medium (Fig. 1), lead phase was increased by 6 to 11 h, and it took at least 24 h to reach lag phase. The addition of salts ($MgSO_4$ and $FeCl_2$) and vitamins (B1) to the deuterated medium did not alter cell growth significantly (Fig. 1). This improvement in cell growth made by the pH adjustment, however, did not correlate with protein production. Cells grown under these conditions did not express SNase (Fig. 2). In order to address this problem we took advantage of the fact that the pAS1 plasmid used in this expression system contains a temperature sensitive lambda repressor (14); protein production can be suppressed at 30°C and induced at 37°C. Using this inherent characteristic of the expression system, cells were grown up in deuterated medium to lag phase (24 h) at 30°C and then incubated at 37°C for another 24 h. Final cell density (1 to 3 OD) and protein yield (80 mg/l) were comparable to cells grown in rich medium. The effect of deuterium and deuterated nutrients in slowing down cell growth and altering expression protocol is not unique to this project; deuterium isotope effects have been shown to affect both cell lag and log phase, as well as cell division itself [13,14]

3.2. *Protein purification and characterization*

Purification of perdeuterated and native SNase was always done in tandem, the native preparation acting as an internal control for the purification process. As documented in Fig. 3, the perdeuterated protein behaved identically to the native protein at each step in the purification process. Reversed-phase HPLC, phenyl

Fig. 3. Parallel purification of native and deuterated SNase. The lanes shown are as follows: (a) native SNase, cell supernatant after harvest, (c) native SNase, Urea extraction No. 1, pellet, (e) native SNase, Urea extraction No. 1, supernatant, (g) native SNase, Urea extraction No. 2, (i) native SNase, loaded onto S-sepharose column, (k) native SNase, eluted off S-sepherose with column load buffer plus 70 mM NaCl, (m) native SNase, eluted off S-sepherose with column load buffer plus 140 mM NaCl. Lanes (b), (d), (f), (h), (j), (l) and (n) are deuterated SNase under the identical conditions described above.

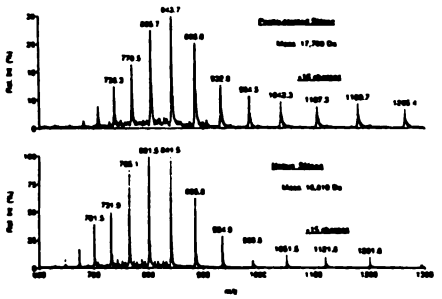


Fig. 4. Electrospray ionization mass spectra of perdeuterated and native SNase. Since the mass spectrometer measures mass-to-charge ratio (m/z), the labeled series of peaks in each spectrum represent a single intact protein species, but with a range of charged states (+14 to +26). The different charged states arise from multiple protons binding to the protein from the acidic sample solution infused into the ion source of the mass spectrometer. Since members of a charged state series must increment in an integral fashion, the intact mass is easily deduced by a mathematical algorithm relating the series members.

superose, and hydroxyapatite [7] were all tried for further protein purification. In terms of final protein yield, reversed-phase HPLC was the most successful final purification step.

In the purified protein, not all hydrogen atoms will be substituted by deuterium. SNase contains a total of 1205 hydrogen sites. These sites can be classified depending on their relative exchangeability in solvent: 921 (76.4%), are non-exchangeable, bound to carbon in amino-acid side chains or the alpha-carbon of the peptide backbone; 284 (23.6%) are exchangeable, bound to amino, carboxyl, hydroxyl groups and to amide groups in the side chains of asparagine and glutamine or the peptide backbone [15]. Since both perdeuterated and native SNase were purified in aqueous hydrogen containing solvents, some of which were denaturing, only non-exhangeable sites of perdeuterated SNase can be expected to remain deuterated.

Electrospray ionization mass spectrometry was used to determine the level of perdeuteration in the protein (see Fig. 4). As shown in Fig. 5a, the measured mass of native SNase is 16 810 daltons, matching the theoretical value. The measured mass of perdeuterated SNase is 17 700 dalton (Fig. 5b), an increase of 890 dalton. If this mass increase is assumed to be a result of deuterium incorporated at non-exchangeable sites only, then an average of 96% of the non-exchangeable

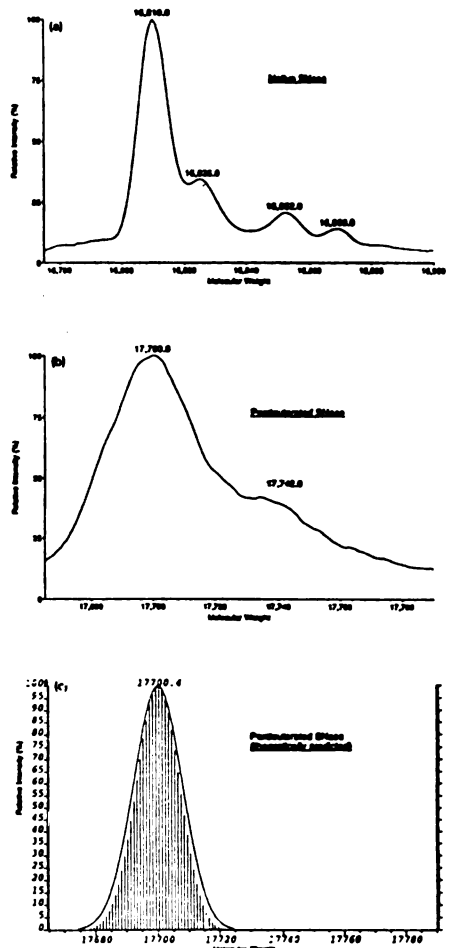


Fig. 5. Electrospray ionization mass spectra. Deconvolutions of the original mass spectra, containing multiply charged states, demonstrating intact species typically present in (a) native SNase and (b) perdeuterated SNase. Peaks corresponding to mass increases of 16 and 42 dalton are consistent with partial methionine oxidation and N-terminal acetylation respectively. A theoretically predicted spectrum for perdeuterated SNase, (c) constructed by presupposing a uniform deuterium relative abundance of 96% at all non-exchangeable sites, and the same instrument resolving power as that used for the experimentally obtained spectra.

sites in perdeuterated SNase have been replaced with deuterium. Typical preparations of perdeuterated SNase yielded >90% deuterium replacement. The upper limit to the level of perdeuteration obtainable is set by the amino acid substitutes, which are generally rated at 97–98% deuterium incorporation (based on initial isotope sources not amino acid products).

The differences in peak width seen between the actual mass spectrum of the perdeuterated SNase and its theoretically predicted spectrum (see Fig. 5c) suggest that the distribution of deuterium among non-exchangeable hydrogen sites is non-uniform. Mass spectral peak width is dependent upon both the resolving power of the instrument and the isotopic distribution of masses present in a sample. The theoretical prediction of the peak width, using the OPUS software package (VG Instruments, Manchester, UK), for perdeuterated SNase is subject to two assumptions: a uniform deuterium relative abundance of 96% at all non-exchangeable sites, and the same instrument resolving power as that used for the experimentally obtained spectra. The peak width of a theoretically predicted spectrum of the native SNase matches the experimentally obtained spectrum (data not shown), confirming the actual instrument resolving power and accounting for naturally occurring isotopic abundances. The actual peak width obtained for perdeuterated SNase is clearly wider than theoretically predicted. In order to create a wider peak without altering the mass, the deuterium distribution must be such that some non-exchangeable sites contain less than 96% deuterium while others contain more. LeMaster and Richards found, when examining isotopically labelled amino acids from protein hydrolysates, that certain non-exchangeable proton sites had reduced isotopic enrichment [16].

3.3. Crystallization

The two main goals in terms of crystallization were to: 1) determine the crystallization conditions for perdeuterated SNase, and 2) to optimize conditions so that large neutron sized crystals could be grown. The same basic protocol resulted in both native and deuterated SNase crystals, although the precise crystallization conditions needed to be optimized independently for each protein preparation. The optimal amount of precipitant (MPD) in the reservoir varied between 20 and 30%. For hanging drops (5–10 μl protein/calcium/inhibitor/precipitant solution), crystals appeared after 4–5 days and grew for up to two weeks. The morphology of the crystals is identical between native and deuterated SNase; the crystals are elongated along one axis and grow rapidly in this direction, but not the other two. (see Fig. 6a and 6b). Crystal shape varied with increasing reservoir precipitant concentration; the edges of the crystals become less angular and more rounded, the overall shape less symmetric and more irregular. Using native SNase, the optimized hanging drop conditions were applied to 50 μl sitting drops in order to grow large crystals. Crystals appeared within two weeks and continued to grow for two weeks, ultimately reaching a size of 0.5 to 1.0 mm^3. Seeding and pulsing techniques were tried to increase the size of these crystals, but both of these methods were unsuccessful because they caused excessive nucleation both in the drop and on the existing crystals. This multitude of nucleation sites is a function of the high concentration of protein necessary to grow these crystals (20 mg/ml); any disturbance to the equilibrium of the drop causes crystal formation. The best results were obtained by using large sitting drops (200 μl); crystals took at least 3 weeks to appear and continued to grow for another month. The largest native crystal was 9.5 mm^3, the typical range of the larger crystals being 1 to 3 mm^3. For the deuterated protein, the largest crystal grown from a 200 μl sitting drop was 1.5 mm^3, the range for the larger crystals being 0.7 to 1.0 mm^3. We do not believe that the observed difference in crystal size between native and deuterated SNase is caused by any property inherent to the protein itself. Crystal size appears to be linked to the number of crystals that form in any given sitting drop which is itself a function of the percentage of MPD in the reservoir and the number of nucleation sites in the drop. Because these parameters vary with each batch of protein and are impossible to control precisely, it is a matter of chance that singular, and thus larger, crystals did not grow from the deuterated protein.

3.4. X-ray structure determination

Native and deuterated SNase crystals grow in the same space group and have identical unit cell dimensions (see Table 2); this is consistent with the parameters reported by of Loll and Lattman for native,

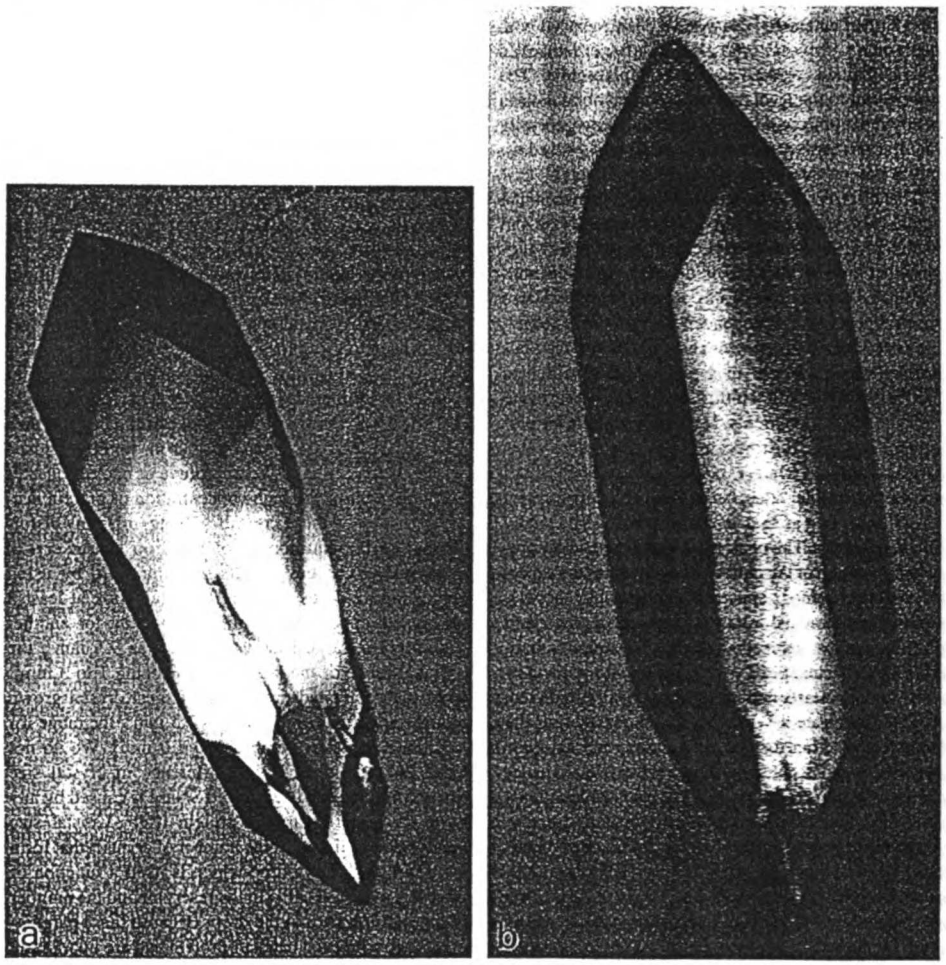


Fig. 6. SNase crystals. Typical morphology of protein crystals grown using (a) native and (b) deuterated SNase. At the time of the photo, the native crystal was 6.25 mm^3, and the deuterated crystal was 1.27 mm^3.

inhibited SNase [4]. The R_{merge} for the two orientations of each crystal, total number of observations, and total number of reflections indicate that each crystal type behaved similarly during data collection and processing. The completeness to 1.9 Å resolution of each data set is better than 95%, with completeness greater than 90% in the highest resolution shell. The two data sets are very similar as evidenced by an R_{scale} between them of 4.05%. No sigma cutoff was applied to either set of data, thus all reflections were used in refinement. Rigid body refinement yielded similar R values using the same starting model for both data sets. Anisotropic scale factors were comparable for the two data sets and lowered the R factor by 2% in each case. From Table

Table 2
Data collection and refinement statistics

Parameter	native SNase	deuterated SNase
space group	$P4_1$	$P4_1$
cell parameters (Å)	$a=b=48.39, c=63.45$	$a=b=48.38, c=63.41$
R_{merge} (%) [a]	5.06	5.55
total No. of observations	36686	44428
total number of reflections	11071	12976
completeness (%)	95.4	96.6
resolution	15–1.9 Å	15–1.9 Å
R_{merge} between 1.97 and 1.9 Å (%)	18.7	16.2
completeness at highest resolution (%) (1.97–1.9 Å)	90.2	93.8
$<I/sig I>$ between 1.97 and 1.9 Å	3.22	3.45
rigid body R value (%)	24.4	24.9
anisotropic scale factors		
native SNase	$B11=-2.53, B22=-2.53, B33=5.05$	
deuterated SNase	$B11=-2.96, B22=-2.96, B33=5.91$	
both cases	$B12=B13=B23=0.00$	
final R value (%)	19.0	19.3
total No. of water molecules	88	89

[a] $R_{merge}=\Sigma_i\Sigma_j|<I_i>-I_{ij}|/\Sigma_i\Sigma_j I_{ij}$, R_{merge} and R_{scale} (see text) are calculated in the same manner.

2 it can be seen that the crystals diffracted better in the direction of the 4-fold rather than perpendicular to it. After refinement with 10 to 1.9 Å data, the model converged at an R value of 24.6% for the native data and 24.3% for the deuterated data, water molecules were added (57 to the native structure, 65 to the deuterated structure). This first round of water addition, together with positional and individual temperature factor refinement improved the R values by 4.3% in each case. In two more rounds, waters were added to both structures giving a total of 88 water molecules for the native structure and 89 water molecules for the deuterated structure. The final R value for the native structure was 19.0%, and for the deuterated structure, 19.3% (see Table 2). The most significant peaks found in a difference map comparing native and deuterated data and using the final native coordinates for the phase information were only 0.11 $e/Å^3$. This difference map is another indicator that perdeuteration did not significantly alter the structure or diffraction characteristics of the protein.

Comparison of the final native and deuterated backbone atoms (Cα,C,N) along with the calcium atom and the inhibitor yields an RMS deviation of 0.08 Å. The largest concerted differences between backbone atoms are found at residues 45 to 51 (0.13 to 0.21 Å). These residues constitute a flexible loop near the active site which previous investigators have found to be partially disordered [4,14]. This flexibility is reflected in temperature factors which are higher than average. The only other differences between the two structures are observed for several of the lysine residues and for Ser141. The largest differences between side chain atoms are also found at these lysine sites, which are located at the surface of the molecule, are exposed to solvent, and are therefore inherently flexible. Ser141 is the last residue of the C terminus to be seen in the electron density; residues 142–149 are presumed to be disordered. All the water molecules except 5 have corresponding mates in the native and deuterated structures; no water with a temperature factor less than 30 $Å^2$ differs from its mate by more than 0.3 Å. Comparison of the two final structures of native and deuterated SNase reveals no conformational alterations caused by the perdeuteration method itself.

4. Conclusion

The purpose of this work has been twofold: (1) to demonstrate that highly perdeuterated protein can be readily produced using established cell lines and protocols and (2) to show that perdeuteration itself does not conformationally alter the protein under study.

Because deuterium isotope effects are known to alter cell division and enzyme function, cell growth and protein expression in a deuterated environment are not expected to be directly comparable to native conditions. The *E. coli* system used in this study was particularly sensitive to the parameters of time and pH; cells in deuterated media took longer to grow than the same cells in hydrogen containing medium and a lowered pH shown to be tolerable to cells in hydrogen containing medium, hindered growth and expression significantly in deuterated medium. Inherent features of the cell line were utilized to overcome observed isotope affects such that cell growth and protein expression were comparable to cells grown in rich medium. Taking into account the ways that deuterium can affect cell behavior, we feel that other cell lines could be optimized to express highly perdeuterated protein.

The expected decrease in background in neutron diffraction data depends directly on the level of protein perdeuteration. It was therefore essential to find a technique that could quantitatively assess the incorporation of deuterium at the hydrogen atom sites. Electrospray ionization mass spectrometry proved to be an excellent candidate for this analysis – the technique requires no special treatment to the protein other than the removal of salts, is accurate to within one or two daltons for proteins of this size, and is becoming more accessible to the scientific community at large. The mass spectrometry data demonstrated that incorporation of deuterium at the non-exchangable hydrogen atom sites can consistently reach 96%, a percentage which approaches the theoretical limit imposed by the deuterated nutrients contained in the medium.

Most importantly, we have shown that perdeuteration itself does not alter the structural features of the protein. The native and deuterated versions of SNase behaved similarly throughout the purification procedure and crystallized under the same experimental conditions. X-ray diffraction was used to compare the diffraction characteristics and structure of the native and deuterated protein, first by direct comparison of the diffraction data, then by analysis of difference Fourier maps made between the two data sets, and lastly by independent refinement and comparison of the two final structures. All of these comparisons confirm that there are no significant structural differences between native and perdeuterated SNase.

Until neutron diffraction data have been collected and analyzed, we cannot quantitatively assess the extent to which protein perdeuteration can enhance the technique. However, there is good reason for optimism given the level of deuterium incorporation, crystal size and quality, and the unaltered state of the perdeuterated protein demonstrated by this work. We hope that protein perdeuteration will enable a much broader range of structural problems to be studied by neutron diffraction than has been possible in the past.

Acknowledgements

We wish to thank Alan Meeker and David Shortle for the pAS1 plasmid and *E. coli* cell line, as well as helpful discussion concerning protein expression and purification, Brad Snedecor for discussion of protein expression in deuterated medium, John Stults and Genentech Inc. for providing a visiting scientist position and access to instrumentation (for KRC) and Bart de Vos for encouragement and for comments on the manuscript. The work done by K.R.C. was supported by Biomedical Research Technology Program of the National Center for Research Resources grants (NIH NCRR BRTP RR01614 and NIH NCRR BRTP RR04112). We are grateful for financial support of this work for TRG by an NIH Biotechnology Training grant (T32GM-08388) and for AAK by an NIH grant (GM33571).

References

[1] P.W. Tucker, E.E. Hazen Jr. and F.A. Cotton, Mol. Cell. Biochem. 27 (1978) 67.
[2] A. Tanaka, J. Flanagan and J.M. Sturtevant, Protein Sci. 2 (1993) 567.
[3] R.A. Kautz and R.O. Fox, Protein Sci. 2 (1993) 851.
[4] T. Nakano, L.C. Antonino, R.O. Fox and A.L. Fink, Biochemistry 32 (1993) 2534.
[5] D. Shortle and A.K. Meeker, Biochemistry 28 (1989) 936.
[6] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong and C.M. Whitehouse, Science 246 (1989) [illegible].
[7] P.J. Loll and E.E. Lattman, Proteins 5 (1989) 183.
[8] A. Messerschmidt and J.W. Pflugrath, J. Appl. Cryst. 20 (1987) 306.
[9] W. Kabsch, J. Appl. Cryst. 21 (1988) [illegible].
[10] A.T. Brunger, XPLOR Manual Version 3.1 (Yale Univ. Press, New Haven, 1992).

[11] A.T. Brunger, Nature 355 (1992) 472.
[12] A.T. Jones, J. Appl. Cryst. 11 (1978) 268.
[13] A.R. Shatzman and M. Rosenberg, Meth. Enzymol. 152 (1987) 661.
[14] E. Flaumenhaft, S. Bose, H.L. Crespi and J.J. Katz, Intern. Rev. Cytol. 18 (1965) 313.
[15] S.W. Englander, N.W. Downer and H. Teitelbaum, Ann. Rev. Biochem. 41 (1972) 903.
[16] D.M. LeMaster and F.M. Richards, Anal. Biochem. 122 (1982) 238.