INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Graduate School Form 9
(Revised 7.94)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis Acceptance

This is to certify that the thesis prepared

By Jiyan Chen

Entitled Part One: Structural Refinement Methodology by using Three-Dimensional NMR Data. Part Two: Chemical Nuclease Design, Synthesis, Activity and Mechanistic Study.

Complies with University regulations and meets the standards of the Graduate School for originality and quality

For the degree of Doctor of Philosophy

Signed by the final examining committee:

Donald E. Bergstrom, Co, chair

Mark Lee

Merritt Cole, Co, Chair

David R McMillin

Approved by: R E Wild
Department Head

4/25/97
Date

This thesis ☐ is ☑ is not to be regarded as confidential. Donald E. Bergstrom
Major Professor

Format Approved by:

Chair, Final Examining Committee

or Rita F. Biederstedt
Thesis Format Adviser

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PART ONE: STRUCTURAL REFINEMENT METHODOLOGY BY USING THREE-DIMENSIONAL NMR DATA
PART TWO: CHEMICAL NUCLEASE DESIGN, SYNTHESIS, ACTIVITY AND MECHANISTIC STUDY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jiyan Chen

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 1997

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 9808427

UMI Microform 9808427
Copyright 1997, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized copying under Title 17, United States Code.

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGMENTS

First of all I should thank God for giving me this great opportunity of doing my Ph.D. research work here at Purdue, one of the best universities in the world. Secondly, I want to thank my parents not only for giving me my physical body, but more importantly for their raising, educating, encouraging, supporting and understanding. My parents are the greatest parents on the earth. Also, I would like to thank my wife, Ya Li, for all sorts of support she has been giving me over the years. My thankfulness should also be extended to all my three wonderful older sisters for their loving and caring.

Academically I am deeply indebted to my two Ph.D. advisors, Professor Donald Bergstrom and Professor David Gorenstein. I want to thank them from the very bottom of my heart for their kindness, generosity, encouragement, patience, guidance, support and personal help. The deepest impression they have given me is their enthusiasm about science. I also want to greatly thank Professor Merritt Andrus, Professor David McMillin and Professor Mark Lipton for being my committee members and giving me the encouragement and advice which I needed. Many people in the Chemistry Department and Department of Medicinal Chemistry and Molecular Pharmacology have generously given me unforgettable help.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

I have enjoyed working with many of my fellow students. Particularly I want to thank all the group members in Dr. Gorenstein's group not only for the enjoyment I had with them in those two years, but far more importantly for their great help during my excruciating days in the summer of 1994. My appreciation to them is far beyond what my words can express. Thank you, Frank, David, Thivi, Eric, Maria, Jian, Elliott, Mark, Jill, Kumar, Dr. Luxon and Dr. Zhang, thank you so much! In Dr. Bergstrom's group, my gratitude goes to Norman for teaching me the ^{33}P labeling technique and sharing his valuable expertise on DNA chemical cleavage reaction with me. My gratitude also goes to Dr. Peiming Zhang and Dr. Jufang Shi for all the help they gave me. I also had unforgettable enjoyment with other research family members in the past three years.

Without the help of the people mentioned, my achievement would have been impossible. They have turned my dream into reality; they have helped me build my beautiful Purdue memory; they have helped me build my bright tomorrow. I want to thank you all again. May God bless us all!

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

ABSTRACT xi

PUBLICATIONS xiii

PART ONE: STRUCTURAL REFINEMENT METHODOLOGY BY USING THREE-DIMENSIONAL NMR DATA 1

CHAPTER 1. INTRODUCTION TO NUCLEIC ACID STRUCTURE REFINEMENT BY USING MULTIDIMENSIONAL NMR DATA 2

1.1 Basics of multidimensional NMR 2
1.1.1 General introduction 2
1.1.2 Fourier Transform NMR spectroscopy 3
1.1.3 Two-dimensional NMR 6
1.1.4 Basic 2D NMR experiments 8
1.1.5 Multidimensional NMR and 3D NOE-NOE 13
1.2 Basics of nucleic acid structure 15
1.2.1 Basic structure of DNA 15
1.2.1 NMR parameters of DNA 18
1.3 Iterative relaxation matrix/MD simulation methodology in structure refinement 25
1.3.1 NOSEY data and interproton distance calculation 26
1.3.2 Molecular dynamics simulation 29
1.3.3 Iterative hybrid matrix/MD refinement 33

CHAPTER 2. STRUCTURE REFINEMENT VIA DECONVOLUTED 3D NOE-NOE DATA: THE HYBRID-HYBRID MATRIX METHOD 35

2.1 Introduction 35

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Page

2.2 Theory and method 40
2.3 Results and discussion 46
2.4 Conclusion 49

PART TWO: CHEMICAL NUCLEASE DESIGN, SYNTHESIS, ACTIVITY AND MECHANISTIC STUDY 52

CHAPTER 3 INTRODUCTION TO ARTIFICIAL NUCLEASES 53

3.1 General Introduction 53
3.2 Artificial nucleases composed of oligonucleotide and transition-metal-complex conjugates 56

CHAPTER 4 CHEMICAL NUCLEASE DESIGN, SYNTHESIS, ACTIVITY AND MECHANISTIC STUDY 74

4.1 Design of major-groove-positioned redox-based chemical nucleases 75
4.2 Syntheses 81
4.2.1 5-[2-(4′-Methyl-2,2′-dipyrid-4-yl-carboxamido)ethylthio]-2′-deoxyuridine phosphoramidite 81
4.2.2 3′-Deuterio-2′-deoxyadenosine phosphoramidite 84
4.3 Synthesis of bipyridine-tethered 2′-deoxyuridine and 3′-deuterio-2′-deoxyadenosine modified oligonucleotides 89
4.4 Nuclease activity assay and mechanistic study 91
4.4.1 Iron mediated single-stranded DNA cleavage 91
4.4.2 Other transition metal mediated single-stranded DNA cleavage 119
4.4.3 Probing the active oxidative species 120
4.5 Discussion and conclusion 122
4.6 Experimental 126

CHAPTER 5 POTENTIAL APPLICATIONS OF A BIPYRIDINE-TETHERED CHEMICAL NUCLEASE AS A TOOL IN MOLECULAR BIOLOGY 145

5.1 Introduction 145
5.2 Results and discussion 145
5.3 Experiments 153

LIST OF REFERENCES 155

VITA 166

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

Table		Page
1.1	Structural parameters for A-form, B-form and Z-form DNA	20
1.2	Approximate ^{1}H chemical shift ranges in single stranded and duplex DNA and RNA fragments	22
4.1	MALDI-TOF results of bipyridine and 2′-deoxy-3′-deuterioadenosine modified oligonucleotides	90

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Figure		Page
1.1	FT NMR	5
1.2	Pulse sequences of selected 2D experiments	9
1.3	The pulse sequences of 3D NOE-NOE	14
1.4	Structure of a polydeoxyribonucleotide	16
1.5	Watson-Crick purine-pyrimidine base pairing	17
1.6	A-form, B-form and Z-form DNA	19
1.7	DNA local motions: twist, roll, tilt	21
1.8	NOE connectivity along a DNA duplex strand	24
1.9	Flow chart of MORASS methodology	28
1.10	Pseudo-harmonic potential function used in distance restrained structure refinement	32
2.1	Flow chart of structure refinement by hybrid-hybrid matrix method of 3D NOE-NOE data analysis through the MORASS/MD protocol	39
2.2	Scheme of linear table for 3D NOE-NOE volume data	44
2.3	The %rms error in the volumes of the deconvoluted 2D volumes plotted against the amount of random noise added to the simulated 3D data set	44
2.4	Plot of representative 3D experimental NOE-NOE volumes for the GG mismatch 12-mer d(CGGGAATTCGCG)$_2$ duplex versus the 3D simulated volumes	48

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure Page

2.5 Representative experimental 2D NOE volumes for the 12-mer duplex (**PAT1**) plotted against the simulated and deconvoluted 3D volumes ... 48

3.1 General scheme for artificial nuclease design ... 55

3.2 Typical transition metal ion complexes which function as redox-active DNA cleaving agents ... 58

3.3 Oligonucleotide-EDTA conjugate as an artificial nuclease ... 59

3.4 Oligonucleotide-phenanthroline conjugate as an artificial nuclease ... 64

3.5 Minor-groove-positioned oligonucleotide-phenanthroline conjugate as an artificial nuclease ... 66

3.6 Oligonucleotide-porphyrin conjugate as an artificial nuclease ... 69

3.7 DNA fragmentation patterns via H3′ hydrogen abstraction in the major groove ... 71

3.8 DNA fragmentation patterns via minor groove hydrogen abstraction in a DNA duplex ... 72

3.9 DNA fragmentation via H2′ hydrogen abstraction in a DNA duplex ... 73

4.1 Design of a major groove positioned bipyridine tethered oligonucleotide ... 77

4.2 Bipyridine covalently positioned in the major groove of a DNA duplex ... 79

4.3 Stereo views of the overlay of a few dynamic motion frames of the bipyridine tethered oligonucleotide duplex (left: major-groove view; right: minor-groove view) ... 80

4.4 Synthesis of C5 bipyridine tethered 2′-deoxyuridine phosphoramidite (**7**) ... 82

4.5 Synthetic scheme for D3′ labeling of 2′-deoxyadenosine phosphoramidite (**23**) ... 86

4.6 Oligonucleotide **O2** mediated cleavage of substrate **O1** in the presence of reductants ... 92

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure Page

4.7 Oligonucleotide **O2** mediated cleavage of substrate **O1** in the presence of hydrogen peroxide 94

4.8 Phosphor Imaging™ quantitation of substrate oligonucleotide **O1** cleavage distribution 95

4.9 Self-cleavage assay of bipyridine modified oligonucleotide **O2** 96

4.10 **O2** Nuclease activity assay after first round cleavage reaction 98

4.11 Assay of sequence, bipyridine position and targeting oligonucleotide length effect on bipyridine modified oligonucleotide nuclease activity 99

4.12 Nuclease activity comparison in different sequences 100

4.13 Effect of bipyridine position on nuclease activity 102

4.14 Nuclease activity assay for structurally perturbed duplexes 103

4.15 Cleavage yield quantitation for non-perfectly matched DNA duplexes 104

4.16 Fenton chemistry 106

4.17 Fe(II) and Fe(III) mediated DNA cleavage 107

4.18 Quantitation of Fe(II) & Fe(III) mediated DNA cleavage 108

4.19 Effect of time on the cleavage yield 109

4.20 Salt concentration effect on DNA cleavage 111

4.21 Temperature effect on nuclease activity of perfectly matched duplex 113

4.22 Temperature effect on cleavage efficiency of 3′-end T bulge substrate **(3′T bulge)** 114

4.23 Temperature effect on cleavage efficiency of 5′-end T bulge substrate **(5′T bulge)** 115

4.24 Deuterium labeling on substrate 117

4.25 Cleavage profile comparison between H3′-A21 **O1** and D3′-A21 **O1** 118

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure Page

4.26 Cleavage profiles in the presence of DMSO 121

5.1 T4 kinase activity towards an oligonucleotide 3′ phosphate 147

5.2 Oligonucleotide-bipyridine conjugate targeting pBluescript™ plasmid substrate 149

5.3 Oligonucleotide-bipyridine conjugate targeting a 448 bp Pvu II restriction fragment from pBluescript™ plasmid 150

5.4 Scheme for using bipyridine tethered oligonucleotides as chemical nucleases for engineering plasmid sequences 152

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

Chen, Jiyan. Ph. D., Purdue University, May 1997. Part One: Structural Refinement Methodology by Using Three Dimensional NMR Data Part Two: Chemical Nuclease Design, Synthesis, Activity and Mechanistic Study. Major Professor: Donald E. Bergstrom.

PART ONE

A hybrid-hybrid matrix method is described that quantitatively analyzes 3D NOE-NOE NMR data. Experimental 3D data are deconvoluted into a 2D hybrid NOESY spectrum. The deconvoluted 2D hybrid NOESY spectrum can then be merged with experimental 2D NOESY data along with additional simulated 2D data as necessary to create a hybrid-hybrid 2D NOE volume matrix. This hybrid-hybrid volume matrix is then used with the complete relaxation program, MORASS, to calculate a rate matrix, and the resulting distances taken from the off-diagonal cross-relaxation rates can then be utilized in a distance geometry or restrained molecular-dynamics refinement of the structure.

PART TWO

Bipyridine was covalently linked from its 4-position to C5 of 2′-deoxyuridine through the tether, -$SCH_2CH_2NH(CO)$-, to give 5-[2-(4′-methyl-2,2′-dipyrid-4-yl-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

carboxamido)ethylthio]-2′-deoxyuridine. This nucleoside was incorporated into oligonucleotides of different sequences via standard phosphoramidite chemistry.

In the presence of 0.01 mM Fe(II) and 0.4 mM H_2O_2, a complementary substrate DNA can be cleaved by the bipyridine tethered targeting oligonucleotide. When the bipyridine is positioned close to the 3′-end of the targeting strand, the opposing strand can be cleaved consistently in 80% yield over a three nucleotide span regardless of DNA duplex sequences. No targeting strand self-fragmentation is observed during the cleavage reaction. The cleavage yield is very sensitive to salt concentration, temperature and certain structure modifications. The cleavage can not be quenched by DMSO at 0.24 mM concentration. Fe(III) shows substantial nuclease activity in the presence of H_2O_2 and the absence of reductants. On the other hand, Fe(II) does not show nuclease activity in the presence of reductants as co-reactants. It was postulated that a fundamental step in the cleavage reaction includes abstraction of a 3′ hydrogen atom from deoxyribose on the opposing sequence. However, no isotope effect was observed in the cleavage reaction when C3′ was labeled with ^2H.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PUBLICATIONS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PUBLICATIONS

A Hybrid-Hybrid Matrix Method for 3D NOE-NOE Data Analysis, Zhang, Qun; Chen, Jiyan; Gozansky, Elliott K.; Zhu, Frank; Jackson, Patricia L; and Gorenstein, David G.; *J. Magn. Reson.*, Series B, 1995, 106, 164-170

Sequence-Specific Oligodeoxyribonucleotide Cleavage by A Major-Groove-Positioned Metal-Binding Ligand Tethered to C-5 of Deoxyuridine, Bergstrom, Donald E.; Chen, Jiyan; *Bioorg. Med. Chem. Lett.*, 1996, 6(18), 2211-2214

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

JOURNAL OF MAGNETIC RESONANCE, Series B 106, 164–169 (1995)

A Hybrid–Hybrid Matrix Method for 3D NOE–NOE Data Analysis

QUN ZHANG, JIYAN CHEN, ELLIOTT K. GOZANSKY, FRANK ZHU, PATRICIA L. JACKSON, AND DAVID G. GORENSTEIN*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

Received February 18, 1994; revised May 10, 1994

A hybrid–hybrid matrix method is described that quantitatively analyzes 3D NOE–NOE NMR data. Experimental 3D data are merged with simulated 3D data to create a hybrid 3D NOE–NOE spectrum. This is then deconvoluted into a 2D hybrid NOESY spectrum. The deconvoluted, 2D hybrid NOESY spectrum can then be merged with other 2D NOESY experimental data along with additional simulated 2D data as necessary to create a hybrid–hybrid 2D NOE volume matrix. This hybrid–hybrid volume matrix is then used with the complete relaxation program, MORASS, to calculate a rate matrix, and the resulting distances taken from the off-diagonal cross-relaxation rates can then be utilized in a distance geometry or restrained molecular-dynamics refinement of the structure. This process is repeated until a satisfactory agreement between the calculated and observed 3D volumes is obtained. This hybrid–hybrid matrix method retains computational efficiency and utilizes the resolution of the 3D data set while retaining any information content of the available 2D data. The initial tests of the deconvolution algorithm give high correlation results even with the introduction of random error into the 3D data set. Our results suggest that the hybrid–hybrid matrix method for analysis of 3D NOE–NOE spectra may provide a viable tool in the refinement of large molecules. © 1995 Academic Press, Inc

As described in the accompanying paper (*1*), 3D NOE–NOE spectroscopy has shown great promise for structural refinement of large biomolecules. Several different approaches have been developed to refine structures from 3D NOE–NOE spectra (*2–5*). Three-dimensional volumes (A_{ijk}) may be treated by the "two-spin approximation" method,

$$A_{ijk} \propto r_{ij}^{-6} r_{jk}^{-6}, \qquad [1]$$

where the volume depends upon the inverse sixth power of the distances between the two sets of spins i, j and j, k. Alternatively, a slightly more accurate method utilizes various terms of the Taylor series expansion of the exponential rate matrix (*2, 6, 7*). However, as demonstrated in the accompanying paper, at realistic mixing times that allow reasonably accurate measurement of 3D NOE–NOE volumes, both the two-spin and the Taylor series expansion approximation methods can lead to considerable systematic errors.

Distance constraints derived from either approximation method may be used in distance geometry (*8–10*) or restrained molecular-dynamics refinement (*11–13*). Alternatively, a direct, NOE gradient refinement method based upon the 3D NOE–NOE (*5*) volumes has been used. However, gradient methods are generally very computationally demanding. Thus in the original Yip and Case (*14*) paper, the gradient required calculating the product of five matrices, which for larger molecules is computationally prohibitive. More recent NOE gradient calculations based upon the first two terms of the expansion of the exponential rate matrix greatly reduce the time required (*5, 15*). Kaptein and coworkers (*5*) have shown that their method scales with n^3, where n is the number of spins. These direct gradient refinement methods utilizing matrix calculations of the volumes thus hold great potential for refinement of larger molecules from 3D NOE–NOE spectra.

The relaxation-matrix approach (*16–24*) avoids the two-spin approximation by employing a matrix eigenvalue/eigenvector solution to the Bloch equations. Importantly, cross-relaxation rates evaluated by a matrix method include effects from multiple relaxation paths, and therefore distances and structures derived from them are more precise and accurate (*16, 19, 25–27*). However, while the 3D volumes may be more accurately calculated from the complete relaxation-matrix method (*1*), accurate refinement methods from a 3D relaxation-matrix method have yet to be fully developed for large biomolecules.

For two-dimensional NMR, accurate distances can be directly obtained from the complete relaxation method by diagonalizing the 2D volume matrix which represents a 2D NOESY spectrum (*19, 28*). Unfortunately the eigenvalue/eigenvector solutions are very sensitive to the accuracy and completeness of the NOESY volume matrix (*19*). A hybrid matrix solution to this problem was originally proposed by Kaptein and co-workers (*29, 30*) and implemented in several programs [IRMA (*29, 30*), MARDIGRAS (*28*), and MO-

* To whom correspondence should be addressed, at Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX 77555.

1064-1866/95 $6.00
Copyright © 1995 by Academic Press, Inc
All rights of reproduction in any form reserved

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RASS (*13, 31*), for example]. This hybrid matrix approach combines the information from the experimental NOESY volumes, A_{ij}^{exp}, and calculated volumes, A_{ij}^{calc}, derived from an initial structure. The hybrid volume matrix, A^{hyb}, contains the well-resolved and measurable cross peaks from the experimental NOESY spectrum, while overlapping or weak cross peaks and diagonals are from the calculated spectrum. This hybrid volume matrix is then used to calculate a rate matrix, and the resulting distances taken from the cross-relaxation rates are then utilized in a distance geometry or restrained molecular-dynamics refinement of the structure. This process is repeated until a satisfactory agreement between the calculated and observed cross-peak volumes is obtained. In various structural refinements, three to six iterations are typically required to achieve convergence with a family of structures consistent with the NOE data. In this paper we show that a similar "hybrid-hybrid" relaxation-matrix method may also be extended to 3D NOE-NOE spectra. It represents a more computationally efficient method than the current gradient methods, avoids the assumptions of the various approximation methods, and potentially provides a method for the refinement of larger biomolecules.

THEORY AND METHODS

In order to mathematically represent the 3D NOE-NOE data, various mathematical expressions have been suggested (*1, 3-5, 32*). For example, two such expressions are

$$A_n(t_m^{(2)}) = \exp\{-t_m^{(2)}R\}A_n(t_m^{(1)}), \qquad [2]$$

where R is the relaxation rate matrix, $t_m^{(1)}$ and $t_m^{(2)}$ are the two mixing times, and $A_n(t_m^{(2)})$ represents a 2D plane cut though the 3D NOE-NOE spectrum (*3*), or

$$A_{ijk}^{3D} \propto A_{ij}^{(1)} A_{jk}^{(2)}, \qquad [3]$$

where A_{ijk}^{3D} is a single 3D volume and $A_{ij}^{(*)}$ are the 2D volumes during the two mixing times (*) (*32*).

Deconvolution of 2D NOE Volumes from 3D NOE-NOE Volumes

A 3D experiment such as a NOE-NOE spectrum inherently contains no more information than the respectively combined 2D experiments. The main advantage of a 3D NOE-NOE spectrum over a 2D NOESY spectrum then is the added spectral resolution provided by the third frequency dimension. For molecules > 10,000 Da, spectral overlap often prevents adequate measurement of enough NOEs from the 2D spectrum, severely limiting the precision and accuracy of any refined structure. This of course has been a major impetus in the continued development of new 3D and 4D experiments. Both 2D and 3D NOE spectra are a function of the 2D relaxation matrix, which is a function of the distances between pairs of spins.

A 3D volume matrix can be deconvoluted into two 2D matrices according to Eq. [3]. Theoretically, a 2D NOE matrix can be obtained from any single plane in the 3D matrix; however, each plane in the 3D matrix is invariably incomplete due to numerical truncation. Clearly a 3D volume A_{ijk}^{3D} will not be experimentally observable if either spin pair i, j or j, k is > 5 Å. However, since each plane will be incomplete in a region different from any other plane, the full data set is still recoverable. This is done by deconvoluting each of the incomplete planes in the 3D matrix into separate incomplete 2D planes representing a part of the full 2D NOE volume matrix using Eq. [2] or [3]. Then the 2D planes are linked together to form a more complete 2D NOE volume matrix. When more than one 2D plane contains a value for any one term, these values are averaged together to give a single element $\langle A_{ij} \rangle$. Errors are minimized in this treatment by averaging many computed A_{ij} values derived from many 3D volumes (A_{ijk}, $k = 1, n$).

Our algorithm requires us to calculate the 2D volumes, A_{ij}, from the 3D volumes and the identical set of 2D volumes A_{jk} (for equal mixing times, we remove the superscripts in Eq. [3]):

$$A_{ij} = A_{ijk}^{3D}/A_{jk}. \qquad [4]$$

In principle, of course, we gain nothing. If there is enough spectral resolution in the 3D spectrum, A_{jk} values can be obtained from the cross-diagonal volumes A_{jjk}, A_{jkk}, or A_{jkj}, or 2D volumes independently measured from a well-resolved 2D NOESY spectrum. Unfortunately, often this will not be possible for larger biomolecules.

Once again we can fashion a hybrid matrix solution to this problem and this has been implemented in a 3D version of MORASS [*m*ultiple *O*verhauser *r*elaxation *a*nalysis and *s*imulation (*19, 31*)], as shown in Fig. 1. An initial model structure and the measured r_{ij} values are used to calculate the rate matrix. We then simulate 2D NOESY and 3D NOE-NOE spectra. The experimental and simulated 3D NOE-NOE data are scaled and then merged to create a hybrid 3D data set, vol3d.hyb. Because of the tremendous number of 3D elements (n^3) in larger biomolecules, the disk file containing only the nonzero elements (about 1-2% of the total elements) would still require approximately 140-280 Mbytes for 600 spins.

The 3D hybrid data are then deconvoluted into a 2D matrix, vol2d.dec, with elements

$$\langle A_{ij} \rangle = \sum_{k=1}^{n} A_{ijk}^{exp}/A_{jk}^{n}, \qquad [5]$$

where nonzero A_{jk}^{n} values may be obtained from * =

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

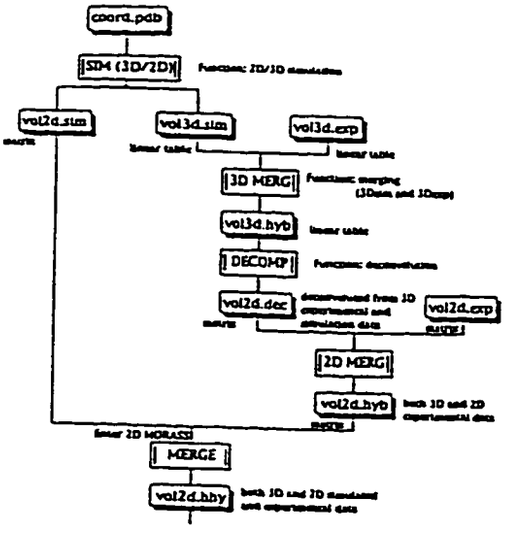


FIG. 1. Schematic description of the hybrid–hybrid matrix method for 3D NOE–NOE structural refinement. The 3D hybrid volume matrix, vol3d.hyb, is generated by scaling and merging the simulated and experimental 3D data. Vol3d.hyb is then deconvoluted into a 2D NOESY volume matrix, vol2d.dec, and scaled and merged with any additional 2D, NOESY (or heteronuclear 3D) experimental data vol2d.exp. The resulting volume matrix vol2d.hyb is then further merged as necessary with the simulated 2D NOESY data, vol2d.sim, to give a hybrid–hybrid 2D volume matrix vol2d.hhy. The relaxation-rate matrix is then calculated by the standard complete relaxation-matrix methodology (cf. MORASS 2D). Distances derived from the cross-relaxation rates are then utilized in restrained molecular dynamics, a new structure is calculated, and both new 3D and 2D NOE spectra are simulated. If they do not agree with experimental spectra, a new iterative cycle is begun until final convergence and agreement between the simulated and experimental 2D and 3D NOE spectra are reached.

exp(experimental) or * = simul (simulated) 2D or 3D spectra. Additional experimental or simulated 2D NOESY volumes can then be merged into the vol2d.dec matrix to give a complete 2D hybrid–hybrid volume matrix, vol2d.hhyb. The rate matrix can then be calculated from the hybrid–hybrid volume matrix using the MORASS (or other) relaxation-matrix approach. The resulting distances are taken from the cross-relaxation rates and the distances are then utilized in a distance geometry or restrained molecular-dynamics refinement of the structure. The entire process is repeated until a satisfactory agreement between the calculated and observed 3D cross-peak volumes is obtained. It remains to be seen whether the 3D to 2D deconvolution portion of the method can be separately optimized independent of the structural refinement stage.

Because two independent 3D data sets are merged, the hybrid–hybrid matrix approach will rely heavily on careful scaling. One way to scale the experimental and theoretical volume matrices is to match the volumes of several "markers" (*29, 30, 33*). Some of the back-transfer volumes (A_{jkj}) in the 3D spectrum will hopefully be well resolved and correspond to spin pairs of fixed distance, much as proton pairs one would use as reference volumes in the 2D hybrid matrix or two-spin methods. Alternatively, all experimental volumes may be used for scaling. The ratio of the sum of these volumes gives the appropriate scale factor, S,

$$S = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} A_{ijk}^{simul}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} A_{ijk}^{exp}}, \qquad [6]$$

where the summation is taken over all experimentally integrated volumes. Importantly, since each refined structure from each cycle is a new molecular structure, S must be reevaluated at every iteration.

EXPERIMENTAL

Synthesis

The GG mismatch 12-mer d(CGGGAATTCGCG) used in this study was synthesized using the phosphite triester solid-phase synthesis methodology as previously described (*34*). The oligonucleotide was purified by HPLC. The NMR sample contains 35 mg of DNA in 0.55 ml D_2O, 0.1 *M* NaCl, 0.1 *M* phosphate buffer, pH 7.0 (uncorrected meter reading).

NMR Data Acquisition

All spectra of the 12-mer d(CGGGAATTCGCG)$_2$ duplex were acquired on a Varian 4000 Series VXR 600 MHz instrument. The 2D NOESY spectra were acquired at room temperature and 10°C. The 10°C NOESY spectrum was acquired with an acquisition time of 0.165 s, a spectral width of 6199 Hz, a 90° pulse width of 18.5 μs, and a 3 s recycle delay. The residual HOD peak was also saturated using hard saturation. Eight transients were acquired per FID for 512 t_1 increments. The spectra were processed using 90°-shifted sine-bell apodization in each dimension. The 3D NOE–NOE spectrum was acquired using the pulse sequence described in Ref. (*35*). The 3D NOE–NOE spectrum at 20°C was acquired with 1024 points in the first dimension, 64 points per plane, and 128 planes. The spectrum was acquired with a spectral width of 5999 Hz. A 1 s recycle delay was used. Eight transients per FID were collected. A mixing time of 100 ms was used for each of the NOE mixing periods in the 3D spectrum.

NMR Data Processing

The 2D NOESY spectra were processed using the VNMR (Varian) software. The 3D NOE–NOE spectrum at 20°C

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

was processed using FELIX 2.05 or 2.10 (Biosym), and the processed data contain 512 × 128 × 256 real data points. It was processed using 90°-shifted sine bells in the first dimension and Kaiser windows in the second and third dimension. No processing to remove residual water was necessary. Linear prediction of the first point was used in the first dimension, and linear prediction of the first three points was used in the second and third dimensions. A fifth-order polynomial was used for baseline correction in the second and third dimensions.

3D Simulation and Effect of Added Noise

In order to test the hybrid–hybrid, deconvolution methodology, a 3D NOE–NOE data set for the 12-mer GG mismatch duplex was simulated, based on the two identical mixing times of 100 ms and a spectrometer frequency of 500 MHz. The structure of the 12-mer duplex had been previously solved in our laboratory by a MORASS/restrained-MD calculation from the experimental 2D NOESY spectrum (*36*). The coordinates were also used to generate a 2D NOE volume matrix at a mixing time of 200 ms.

In order to test the robustness of the deconvolution algorithm, various levels of random noise were added to the simulated 3D data set. The type of experimental error included in this simulation is most common to multidimensional NMR, that of low signal-to-noise and incorrect volume integration, which is a particular problem in 3D spectra. Elements of the relaxation-rate matrix were calculated from the set of proton Cartesian coordinates and a correlation time of 3.5 ns. A partial set of 107 spins from the 12-mer duplex was used in the simulations, and only those volumes that could potentially be integrated were included in the simulations (maximum of approximately 50,000). All other 3D matrix elements were set to zero, and only a linear table of the nonzero elements was stored on disk. From the relaxation matrix the 3D volumes were generated for a given t_m using Eq. [3]. This represented the target spectrum, with noise-free data.

Experimental noise was then added to the target, simulated 3D spectrum. Random noise from a Gaussian distribution was added to all peak volumes in A^{3D} to simulate both a constant low-level thermal noise and a peak-integration error. A level of noise from ±5 to ±20% proportional to the individual volume was added to each A_{ijk} element.

RESULTS AND DISCUSSION

Simulations and Effect of Noise

A 3D NOE–NOE spectrum of a 12-mer GG mismatch $d(CGGGAATTCGCG)_2$ duplex was simulated, and 5–20% random noise proportional to the 3D volumes was added. We used the back-transfer volumes, A_{jkj}, to calculate an initial estimate of A_{jk} ($=[A_{jkj}]^{1/2}$). The 3D spectrum was deconvoluted as described under Theory and Methods, and the calculated averaged 2D volumes $\langle A_{ij} \rangle$ compared to the exact simulated values. Figure 2 shows a plot of the % rms deviation of the A_{ij} data after deconvolution as a function of the added random noise. As expected, the resulting rms errors for a single deconvoluted A_{ij} volume derived from the simulated (plus noise) 3D volumes are about one-half of the errors introduced into the 3D volumes (solid line). However, when the planes are linked together and the $\langle A_{ij} \rangle$ values derived from the simulated (plus noise) 3D volumes are compared to the target A_{ij}, the error is dramatically reduced two- to threefold (dashed line), due to the effects of averaging.

Comparison to Experimental 3D and 2D NOE Spectra

Forty-six representative, true 3D peaks ($f_1 \neq f_2 \neq f_3$) from the experimental 3D NOE–NOE spectrum (mixing times of 100 ms) of the 12-mer duplex were integrated using FELIX software (Biosym. Inc.) and compared against the data from the full matrix calculation using Eq. [3] and the NMR-derived solution structure. These were chosen because both an A_{ijk} and a corresponding back-transfer volume A_{jkj} could be reliably integrated. As shown in Fig. 3, a linear regression correlation coefficient of 0.91 was obtained.

The 3D simulated data were also deconvoluted into a 2D NOE volume matrix, as shown in Fig. 1, and compared to the experimental 2D NOESY spectrum. Using 87 representative 2D peak volumes, the correlation coefficient was 0.83 (Fig. 4). These results suggest that the hybrid–hybrid method of 3D NOE–NOE data analysis could provide a viable tool for refinement studies of large biomolecules. Other elements

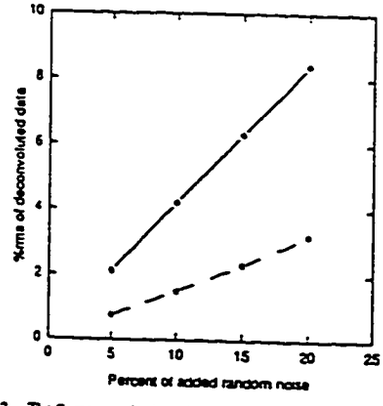


FIG. 2. The %rms error in the volumes of the deconvoluted 2D volumes is plotted against the amount of random noise added to the simulated 3D data set. The perturbed data were deconvoluted and the rms error was calculated for each plane (solid line) and the resulting rms after the planes were averaged together (dashed line). Only a single 2D volume is shown since the remainder of the data is virtually identical.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

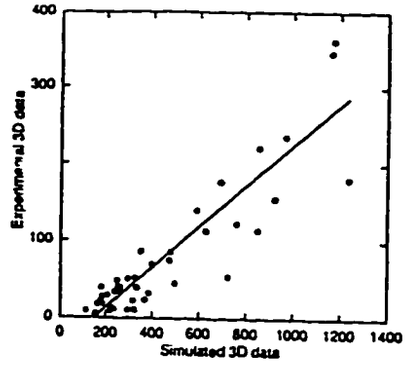


FIG. 3. Plot of representative 3D experimental NOE-NOE volumes for the GG mismatch 12-mer d(CGGGAATTCGCG)$_2$ duplex versus the 3D simulated volumes. The data represent both 2D back transfer (V_{BT}) and true 3D V_{3D} volumes. The solid line is the best fit of the data from the linear regression with a correlation coefficient of 0.91.

of the algorithm in Fig. 1 are currently being tested. A major limitation in all 3D methods is the difficulty of accurately integrating the many thousands of 3D volumes. Development of efficient software to automate the peak picking and integration of 3D data sets will be crucial. A full refinement of the GG 12-mer mismatch using the hybrid-hybrid 3D matrix method is in progress. In principle, we should be able to obtain many more distance constraints for refinement of a molecule from the 3D spectrum. However, this must be balanced against the poorer precision of 3D versus 2D volume integration.

Computational Efficiency

In designing any methodology, an important consideration is calculational efficiency. Here, only the 3D peaks that are required for the deconvolution and scaling are simulated and stored in memory (or on disk if necessary). This data can then be merged with the 3D experimental data, deconvoluted, merged with the experimental 2D data, and then stored as a simple 2D array. As a result, there is no need to store a three-dimensional array for the 3D data. This saves time and disk space. This design partially removes the extra limitations on the number of spins the computer can handle for 3D data, thus making the methodology applicable to larger systems.

CONCLUSIONS

Simulations have shown that the highest accuracy and precision in structural refinement is achieved if *both* increased numbers and accuracy of the constaints are used (*25, 37*). The use of approximate methods has led investigators to apply liberal error bars when using NMR-derived distances during constrained structural refinements. While it is true that many good-quality structures have been obtained from NMR (*38*)—especially with the development of 3D and 4D NMR (*39–42*)—it is obvious that using more and tighter constraints will achieve greater precision (and hopefully accuracy) in the structures obtained by NMR (*25, 43, 44*). This appears to be generally true for both proteins (*37, 44*) and nucleic acids (*25, 43, 45–47*). However, it has been reported that a relaxation-matrix analysis for proteins may actually lead to poorer accuracy (*48*).

Thus, 3D NOE-NOE spectra hold the promise of giving more accurate structures given the vastly increased number of resolvable 3D NOE-NOE volumes (*5*). As we have shown in the accompanying article (*1*), approximation methods may not yield very accurate distances for the longer mixing times required to achieve adequate magnetization transfer and signal-to-noise in these 3D spectra. Calculations based upon the complete relaxation matrix are well within current computational resources, even for larger macromolecules (<600 spins). Competing with the hybrid-hybrid matrix method described here are various direct gradient-NOE refinement methods (*5*). Even with approximations to the gradient calculations, these scale as n^3, whereas the hybrid-hybrid matrix method is still basically an n^2 problem (diagonalization of the $n \times n$ volume matrix). It must be emphasized that NOE gradient refinements should utilize relaxation-matrix methods for calculating 3D volumes in order to achieve the highest accuracy for the refined structure (*1*).

An additional advantage of this methodology is that software written for 2D NOESY data analysis does not need to be significantly modified since the 2D deconvoluted matrix

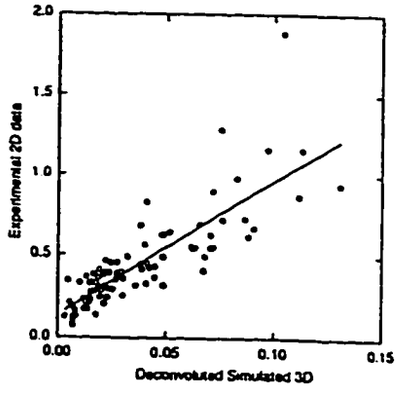


FIG. 4. Representative experimental 2D NOE volumes for the 12-mer duplex are plotted against the simulated and deconvoluted 3D volumes. The solid line is the best fit of the data from the linear regression with a correlation coefficient of 0.83.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

replaces the experimental 2D NOESY volume matrix. The analysis techniques described here are also general enough that they can be expanded to include even higher dimensional experiments (such as a 4D NOE–NOE–NOE spectrum).

ACKNOWLEDGMENTS

Supported by NIH (AI27744), the Purdue University Biochemical Magnetic Resonance Laboratory, which is supported by the NIH designated AIDS Research Center at Purdue (AI27713), and the NSF designated Biological Facilities Center on Biomolecular NMR, Structure, and Design at Purdue (Division of Biological Instrumentation Grants BBS 8614177 and 9214258). The contributions of Florence LeBreton, Vikram Roongta, Bruce Luzon, Michael Hills, and Dean Carlson are much appreciated.

REFERENCES

1. D. G. Donne, E. K. Gozansky, and D. G. Gorenstein, *J. Magn. Reson. B* 106, 156 (1994).
2. J. Habazettl, A. Ross, H. Oschkinat, and T. A. Holak, *J. Magn. Reson.* 97, 511 (1992).
3. H. Kessler, S. Seip, and J. Saulitis, *J. Biomol. NMR* 1, 83 (1991).
4. R. Berstein, A. Ross, C. Cieslar, and T. A. Holak, *J. Magn. Reson. B* 101, 185 (1993).
5. A. M. J. J. Bonvin, R. Boelens, and R. Kaptein, *J Magn Reson.* 95, 626 (1991).
6. J. Habazettl, M. Schleicher, J. Otlewski, and T. A. Holak, *J. Mol. Biol.* 228, 156 (1992).
7. T. A. Holak, J. Habazettl, H. Oschkinat, and J. Otlewski, *J. Am. Chem. Soc.* 113, 3196 (1991).
8. T. F. Havel, I. D. Kuntz, and G. M. Crippen, *Bull. Math. Biol* 45, 665 (1983).
9. W. Braun and N. Go, *J. Mol Biol* 186, 613 (1983).
10. K. Wüthrich, *in* "NMR of Proteins and Nucleic Acids," Wiley, New York, 1986.
11. M. Nilges, A. M. Gronenborn, A. T. Brünger, and G. M. Clore, *Protein Eng.* 2, 27 (1988).
12. E. Zuiderweg, R. Scheek, R. Boelens, R. van Gunsteren, and R. Kaptein, *Biochemie* 67, 707 (1985).
13. D. G. Gorenstein, R. P. Meadows, J. T. Metz, E. P. Nikonowicz, and C. B. Post, *in* "Advances in Biophysical Chemistry" (C. A. Bush, Eds.), pp. 47–124, JAI Press, London, 1990.
14. P. Yip and D. A. Case, *J. Magn. Reson.* 83, 643 (1989).
15. A. M. J. J. Bonvin, R. Boelens, and R. Kaptein, *J. Biomol. NMR* 1, 305 (1991).
16. J. W. Keepers and T. L. James, *J. Magn. Reson.* 57, 404 (1984).
17. A. A. Bothner-By and J. H. Noggle, *J. Am. Chem. Soc.* 101, 5152 (1979).
18. S. Macura and R. R. Ernst, *Mol. Phys.* 41, 95 (1980).
19. C. B. Post, R. P. Meadows, and D. G. Gorenstein, *J. Am. Chem. Soc.* 112, 6796 (1990).
20. E. T. Olejniczak, R. T. Gampe, Jr., and S. W. Fesik, *J. Magn. Reson.* 67, 28 (1986).
21. H. Robinson and A. H. J. Wang, *Biochemistry* 31, 3524 (1992).
22. R. Nibedita, R. A. Kumar, A. Majumdar, and R. V. Hosur, *J. Biomol. NMR* 2, 467 (1992).
23. K. Banks, D. Hare, and B. Reid, *Biochemistry* 28, 6996 (1989).
24. N. B. Ulyanov, A. A. Gorin, V. B. Zhurkin, B. C. Chen, M. H. Sarma, and R. H. Sarma, *Biochemistry* 31, 3918 (1992).
25. K. Kaluarachchi, R. P. Meadows, and D. G. Gorenstein, *Biochemistry* 30, 8785 (1991).
26. R. P. Meadows, C. B. Post, K. Kaluarachchi, and D. G. Gorenstein, *Bull. Magn. Reson.* 13, 22 (1991).
27. S. B. Landy and B. D. N. Rao, *J. Magn. Reson. B* 102, 209 (1993).
28. B. A. Borgias, M. Gochin, D. J. Kerwood, and T. L. James, *Prog. NMR Spectrosc.* 22, 83 (1990).
29. R. Boelens, T. M. G. Koning, and R. Kaptein, *J. Mol Struct.* 173, 299 (1988).
30. R. Boelens, T. M. G. Koning, G. A. van der Marel, J. H. van Boom, and R. Kaptein, *J. Magn. Reson.* 82, 290 (1989).
31. R. Meadows, C. B. Post, and D. G. Gorenstein, MORASS Program, 1989.
32. R. Boelens, G. Vuister, T. Koning, and R. Kaptein, *J Am Chem. Soc.* 111, 8525 (1989).
33. E. P. Nikonowicz, R. P. Meadows, and D. G. Gorenstein, *Biochemistry* 29, 4193 (1990).
34. V. A. Roongta, C. R. Jones, and D. G. Gorenstein, *Biochemistry* 29, 5245 (1990).
35. F. Lebreton, M.S. Thesis, Purdue University, 1992.
36. V. Roongta, Ph.D. Thesis, Purdue University, 1989.
37. P. D. Thomas, V. J. Basus, and T. L. James, *Proc Natl Acad Sci. USA* 88, 1237 (1991).
38. K. Wüthrich, *Science* 243, 45 (1989).
39. S. Grzesiek, H. Dobeš, R. Gentz, G. Garotta, A. M. Labhardt, and A. Bax, *Biochemistry* 31, 8180 (1992).
40. L. E. Kay, G. M. Clore, A. Bax, and A. M. Gronenborn, *Science* 249, 411 (1990).
41. E. R. P. Zuiderweg, A. M. Petros, S. W. Fesik, and E. T. Olejniczak, *J. Am. Chem. Soc.* 113, 370 (1991).
42. G. M. Clore, P. T. Wingfield, and A. M. Gronenborn, *Biochemistry* 30, 2315 (1991).
43. A. Pardi, D. R. Hare, and C. Wang, *Proc. Natl. Acad Sci. USA* 85, 8785 (1988).
44. A. T. Brunger, G. M. Clore, A. M. Gronenborn, R. Saffrich, and M. Nilges, *Science* 261, 328 (1993).
45. L. Nilsson, G. M. Clore, A. M. Gronenborn, A. T. Brunger, and M. Karplus, *J. Mol. Biol.* 188, 455 (1986).
46. A. Gronenborn and G. Clore, *Biochemistry* 28, 5978 (1989).
47. W. J. Metzler, C. Wang, D. B. Kitchen, R. M. Levy, and A. Pardi, *J. Mol. Biol.* 214, 711 (1990).
48. G. M. Clore, M. A. Robien, and A. M. Gronenborn, *J Mol. Biol.* 231, 82 (1993).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pergamon

Bioorganic & Medicinal Chemistry Letters, Vol. 6, No. 18, pp. 2211-2214, 1996
Copyright © 1996 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0960-894X/96 $15.00 + 0.00

PII: S0960-894X(96)00402-7

SEQUENCE-SPECIFIC OLIGODEOXYRIBONUCLEOTIDE CLEAVAGE BY A MAJOR-GROOVE-POSITIONED METAL-BINDING LIGAND TETHERED TO C-5 OF DEOXYURIDINE

Donald E. Bergstrom* and Jiyan Chen†

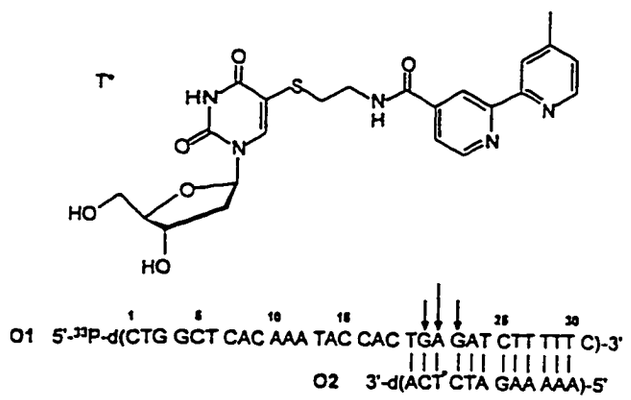
Department of Medicinal Chemistry and Molecular Pharmacology and †Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-1333 and Walther Cancer Institute, Indianapolis, Indiana 46208

Abstract: Highly efficient cleavage of the single-stranded oligonucleotide 5'-d(CTGGCTCACAAATACCAC TGAGATCTTTTTC)-3' (**O1**) at nucleotides G20, A21, and G22 was achieved by an iron-bipyridine complex linked to a targeting oligonucleotide 5'-d(AAA AAG ATC T*CA)-3' (**O2**) through -$SCH_2CH_2NHC(O)$- at C5 of a 2'-deoxyuridine (T*). Copyright © 1996 Elsevier Science Ltd

There is substantial interest in designing chemical nucleases that are able to cleave nucleic acids with high sequence specificity and efficiency. Numerous site-specific chemical nucleases that function via metal-mediated redox or hydrolytic processes have been reported.[1-16] In an effort to increase both the specificity and efficiency of cleavage, we have been studying the effect of linking redox-metal binding ligands at nucleotide sites, which on hybridization, lead to placement of the ligand either in the minor or major groove near abstractable hydrogen atoms on the target oligonucleotide. We recently reported site-specific cleavage of a 31-mer by complementary oligonucleotides containing a phenanthroline ligand linked through N2 of deoxyguanosine. Iron-mediated abstraction of H1' and H4' deoxyribosyl protons from within the minor groove led to cleavage of the target oligonucleotide.[17]

In this communication, we report the high cleavage specificity and efficiency of a single-stranded oligonucleotide (**O1**, Figure 1) mediated by a *major groove directed* iron-bipyridine complex linked to a targeting oligodeoxyribonucleotide (**O2**) through -$SCH_2CH_2NHC(O)$- at C5 of a 2'-deoxyuridine.

Figure 1



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The bipyridine-conjugated oligonucleotide [O2, 5'-d(AAAAAGATCT*CA)-3'] is complementary to nucleotides 19-30 of O1. On hybridization the bipyridine would be expected to occupy the space in the major groove in the vicinity of nucleotides 19-21 (Figure 1). For redox mediated cleavage from within the major groove, the only readily abstractable atom is the deoxyribosyl H3'.[18]

Synthesis of O2 was previously described.[19] The substrate oligonucleotide O1 was synthesized by the standard solid-phase synthesis method with phosphoramidites as building blocks, and then purified on 20% polyacrylamide gel. 5'-End ^{33}P labeling of O1 was accomplished with ^{33}P ATP and polynucleotide kinase from Amersham following the manufactures protocol. In our previous study we determined that the only metal ion that effectively promoted redox cleavage of a complementary sequence was Fe^{2+} in the presence of a reducing agent and oxygen.[17] This has also proven to be true in the present case. Although bipyridine complexed copper ion has been considered as a possible reagent for hydrolytic cleavage of oligoribonucleotides[20] and phenanthroline complexed copper ion is an effective reagent for the redox mediated cleavage of both DNA[2] and RNA,[21] virtually no cleavage was observed when oligonucleotides O1 and O2 were combined in the presence of Cu^{2+} either in the presence or absence of H_2O_2. On the other hand, in the presence of either Fe^{2+} or Fe^{3+} and H_2O_2, O1 was cleaved in high yield. A polyacryamide gel of the products from a cleavage reaction are shown in Figure 2.

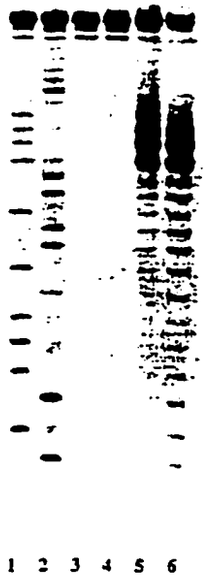


Figure 2. 0.9 pmol of ^{33}P labeled target oligo O1 was combined with 4.5 pmol bipyridine-modified oligonucleotide O2, and evaporated to dryness. The mixed oligonucleotides were redissolved in 8 μL reaction buffer (50 mM TRIS pH 7.4, 50 mM NaCl) and annealed. The cleavage reaction was initiated by adding 1 μL of freshly prepared 0.1 mM ferrous ammonium sulfate and 1 μL of 4 mM hydrogen peroxide. After 30 min another 1 μL of 4 mM hydrogen peroxide was added to the reaction mixture. The reaction system was quenched after 3 hr at 22 °C by the addition of 1 μL 10 mM EDTA and 1 μL calf thymus DNA. Water was removed by evaporation and the oligonucleotides dissolved in 3.5 μL of loading buffer (formamide: 10xTBE, 9:1). The reaction mixture was denatured by heating to 90 °C for 5 min followed by cooling on ice immediately before loading onto 0.4 mm 20% polyacrylamide gel. After electrophoresis at 1800V for 3 hr, the labeled oligonucleotide was transferred to a nylon membrane by 45 min capillary blotting. The membrane was exposed to Kodak X-ray film after dryness to visualize the oligonucleotides. Lane 1: Maxam–Gilbert sequencing (A,G); lane 2: Maxam–Gilbert sequencing (C, T); lane 3: O1; lane 4: O1 subjected to Fe^{2+} and H_2O_2 in the absence of O2; lane 5: Reaction of O1 with Fe^{2+} and H_2O_2 in the presence of O2; lane 6: reaction products of O1 mediated by Fe and treated with 1M piperidine at 90 °C for 30 min.

The amount of cleavage was quantitated on a phosphorimager with Molecular Analyst® (Version 1.30). The cleavage pattern and yield is completely reproducible with the primary sites of cleavage occurring at G20,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A21, and G22. Eighty percent total cleavage was obtained with only five fold excess of O1 (Figure 3). The three-site cleavage specificity of this study is much greater than observed for iron-EDTA mediated cleavage reactions in which hydroxyl radical is generally believed to be the responsible oxidative species. The cleavage yield is sufficiently high, that it will be of interest to see if these results can be extended to much longer sequences and to double stranded DNA. Another remarkable finding of this study is that very little self-cleavage of O2 was observed. A control experiment in which O2 was 5'-end labeled and subjected to Fe^{2+} and hydrogen peroxide showed less than 10% cleavage (Figure 4).

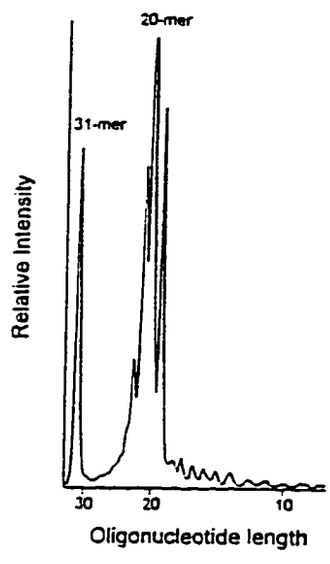


Figure 3: Phosphorimager quantitation of lane 5 (Fig. 2).

Figure 4: Lane 1: 5'-^{33}P-labeled O2; lane 2: O1 (unlabeled) and O2 in the absence of Fe^{2+} and H_2O_2; lane 3: O2 treated with Fe^{2+} and H_2O_2; lane 4: reaction of O2 with Fe^{2+} and H_2O_2 in the presence of O1 (unlabeled) under the same conditions as described above in the caption to Figure 2.

The experimental results suggests that the responsible oxidative species, either metal-oxo or restricted hydroxyl radical, is highly localized around the complementary strand and the metal-ligand complex. Since no external reductant was added, it seems unlikely that a Fenton-type iron cycling process for hydroxyl radical generation is operating. If the bipyridine complexed iron is indeed localized within the major groove then the observed cleavage would be expected to result primarily through abstraction of deoxyribosyl H3'. The 5'-end labeled cleavage products co-migrated on polyacrylamide gel electrophoresis with the products of a Maxam–Gilbert sequencing reaction of O1. This indicates that they are terminated by a 3'-phosphate group. DNA cleavage mediated by photolysis of phenanthrenequinone diimine complexes of rhodium(III) occurs through

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

abstraction of a 3'-hydrogen atom to give oligonucleotides terminated primarily by 3'-phosphate.[22] In that study, lesser amounts of 3'-phosphoglycaldehyde terminated oligonucleotides apparently resulted from the reaction of dioxygen with the intermediate 3'-radical. We were not able to detect the slower moving 3'-phosphoglycaldehyde terminated oligonucleotides.

The results reported here illustrate the potential for achieving high specificity and efficiency by positioning a redox active metal–ligand complex in the major groove of an oligonucleotide duplex. The elucidation of the oxidation mechanism and further improvement of the site specificity and cleaving efficiency is actively being pursued.

Acknowledgment: Technical assistance by and helpful discussions with Dr. Norman Gerry are greatly appreciated. The National Institutes of Health is gratefully acknowledged for support of this research through NIH Grant AI-26029.

References and Notes:

1. Chen, C.-H. B.; Gorin, M. B.; Sigman, D. S. *Proc. Natl. Acad. Sci. U.S.A.* 1993, *90*, 4206.
2. Chen, C.-H. B.; Sigman, D. S. *Proc. Natl. Acad. Sci. U.S.A.* 1986, *83*, 7147.
3. Chu, B. C. F.; Orgel, L. E. *Proc. Natl. Acad. Sci. U.S.A.* 1985, *82*, 963.
4. Dreyer, G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* 1985, *82*, 968.
5. Francois, J.-C.; Saison-Behmoaras, T.; Chassognol, M.; Thuong, N. T.; Helene, C. *J. Biol. Chem.* 1989, *264*, 5891.
6. Frolova, E. I.; Fedorova, O. S.; Knorre, D. G. *Biochimie* 1993, *75*, 5.
7. Boutorin, A. S.; Vlassov, V. V.; Kazakov, S. A.; Kutiavin, I. V.; Podyminogin, M. A. *FEBS Lett.* 1984, *172*, 43.
8. Boidot-Forget, M.; Chassignol, M.; Takasugi, M.; Thuong, N. T.; Helene, C. *Gene* 1988, *72*, 361.
9. Sergeyev, D. S.; Godovikova, T. S.; Zarytova, V. F. *FEBS Lett.* 1991, *280*, 271.
10. Kady, I. O.; Groves, J. T. *Bioorg. Med. Chem. Lett.* 1993, *3*, 1367.
11. Le Doan, T.; Perrouault, L.; Helene, C.; Chassignol, M.; Thuong, N. T. *Biochem.* 1986, *25*, 6736.
12. Groves, J. T.; Kady, I. O. *Inorg. Chem.* 1993, *32*, 3868.
13. Pitie, M.; Casas, C.; Lacey, C. J.; Pratviel, G.; Bernadou, J.; Meunier, B. *Angew. Chem. Int. Ed. Engl.* 1993, *32*, 557.
14. Magda, D.; Wright, M.; Miller, R. A.; Sessler, J. L.; Sansom, P. I. *J. Am. Chem. Soc.* 1995, *117*, 3629.
15. Kuroda, R.; Satoh, H.; Shinomiya, M.; Watanabe, T.; Otsuka, M. *Nucleic Acids Research* 1995, *23*, 1524.
16. Oikawa, S.; Kurasaki, M.; Kojima, Y.; Kawanish, S. *Biochem.* 1995, *34*, 8763.
17. Bergstrom, D. E.; Gerry, N. P. *J. Am. Chem. Soc.* 1994, *116*, 12067.
18. Miaskiewicz, K.; Osman, R. *J. Am. Chem. Soc.* 1994, *116*, 232.
19. Wang, G.; Bergstrom, D. E. *Tetrahedron Lett.* 1993, *34*, 6721.
20. Modak, A. S.; Gard, J. K.; Merriman, M. C.; Winkeler, K. A.; Bashkin, J. K.; Stern, M. K. *J. Am. Chem. Soc.* 1991, *113*, 283.
21. Chen, C.-h. B.; Sigman, D. S. *J. Am. Chem. Soc.* 1988, *110*, 6570.
22. Sitlani, A.; Long, E. C.; M., P. A.; Barton, J. K. *J. Am. Chem. Soc.* 1992, *114*, 2303.

(*Received in USA* 25 *July* 1996; *accepted* 20 *August* 1996)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PART ONE: STRUCTURAL REFINEMENT METHODOLOGY BY USING THREE-DIMENSIONAL NMR DATA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 1. INTRODUCTION TO NUCLEIC ACID STRUCTURE REFINEMENT BY USING MULTIDIMENSIONAL NMR DATA

1.1 Basics of multidimensional NMR

1.1.1 General introduction

Three dimensional structure determination of biomolecules plays a critical role in understanding their biological functionalities and their potential applications. In structural elucidation, X-ray diffraction method has been a very important tool for crystal structures. Since most chemical and biochemical processes take place in solution, solution structures of proteins and nucleic acids are of significant interest to chemists and biochemists. This interest has been a driving force for scientists in different areas to work together to advance techniques for biomolecule solution elucidation.

In the last two decades, Nuclear Magnetic Resonance (NMR) spectroscopy has been developed into a very powerful new technique which can resolve relatively large size biomolecule structures in high spectrum resolution.[1-3] This new application of NMR results from deeper understanding of NMR theory, more powerful modern

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

computer techniques, and more importantly, Fourier Transform NMR which makes multidimensional NMR possible.

How is NMR data used for solution structure refinement? All the refinement methods developed so far use NMR data (NOE intensities and/or coupling constants) as certain restrains to re-shape an initial structure gradually to a "final" structure, when the consistency between the "final" structure and the experimental data satisfies a set of criteria.

There are different approaches and algorithms designed to better use NMR information for structure determination, but new structure refinement methodology is needed when new type of NMR data result from more advanced NMR experiments. In the first part of this thesis, the main focus is the further development of Multispin Overhauser Relaxation AnalySis and Simulation (MORASS) for incorporating 3D NOE-NOE information into nucleic acid structure refinement. The following sections of this chapter should provide the reader with an understanding of the basic properties and processes involved in NMR directed nucleic acid structure refinement and an appreciation for the information available through powerful modern NMR techniques. The fundamental concepts of theory, the experimental capabilities and potential of multi-dimensional NMR techniques is discussed in the following sections.

1.1.2 Fourier Transform NMR spectroscopy

The success of the study of biomolecule solution structures is largely due to the development of Fourier Transform NMR spectroscopy.[3]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

An atomic nucleus with non-zero spin quantum number I has a magnetic momentum and its associated $(2I+1)$ different orientations with their corresponding quantum numbers $\pm I, \pm(I-1), \ldots, 0$. These degenerated energy states have different energy levels when the nucleus is placed in a magnetic field, even though the energy differences are very small. If a low energy sweeping radiation is exerted on this nucleus in a magnetic field, these energy level differences can be mapped out to yield a CW (continuous wave) NMR spectrum.

Modern FT NMR uses a low energy radio frequency to radiate nuclei in a very short period of time (rf pulse), causing many absorbtions to occur simultaneously. The characteristic energy level (frequencies) relaxations in time domain are picked out by Fourier Transformation and the signals in the frequency domain are displayed (Figure 1.1). This technique has revolutionized NMR spectroscopy, and provides an extremely useful tool for chemists and biochemists.

Figure 1.1 shows the simplest vector model of rf pulse NMR. In an external magnetic field, unevenly populated two states lead to a net magnetization (**M**) which is lined up with the external magnetic field (B_0). When a radio frequency pulse is applied to this magnetization (**M**), **M** is flipped to an angle which is controlled by the pulse duration and its energy. The **M** rotates in the XY plane with a characteristic frequency ω_0 given by:

$$\omega_0 = \gamma B_0 \qquad (1.1)$$

where γ is the magnetogyric ratio of the nucleus under study. The motion of **M** is recorded as a time dependent signal by a receiver coil placed in the XY plane. The time

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

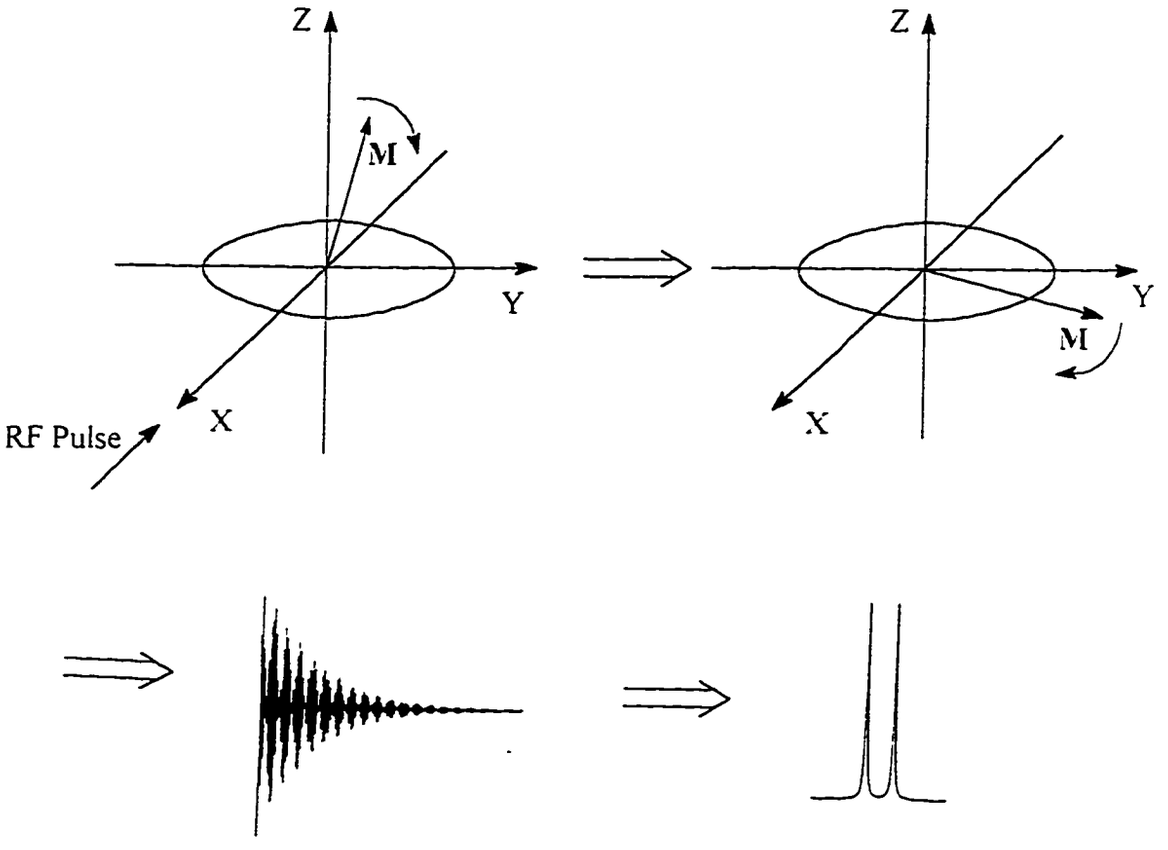


Figure 1.1 FT NMR. Nuclear magnetic signal in time domain is transformed into frequency domain with Fourier Transformation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

domain signal is then transformed into frequency domain signal. **M** will gradually diminish to zero due to two processes: the longitudinal or spin-lattice relaxation (T_1) and transverse or spin-spin relaxation (T_2).

Because the intensity of NMR signal is very weak, multi-scan signal cumulation is necessary, particularly for those nuclei with small magnetogyric ratios and low natural isotope abundance. Theoretically, the signal to noise ratio increases proportionally to square root of the number of accumulated scans, which is only limited by the excited nucleus relaxation rate in FT NMR, since rf pulse rates are very fast (< 1ms). The FT NMR is more efficient in averaging cumulated signals than CW NMR.[4] FT NMR also provides the possibility to excite multi nuclei simultaneously, which is fundamental to many modern NMR experiments discussed later. However, the advantages of FT NMR are not limited to these two merits.

One of the most important factors for NMR spectroscopy is the strength of external magnetic field. Today 750 MHz spectrometers are available, which can provide very high resolution spectra for analysis of large biomolecules.

1.1.3 Two-dimensional NMR

In 1971 Jeener[5] first proposed a two dimensional NMR experiment, which subsequently led to numerous 2D and multi-dimensional NMR experiments. These experiments provide tremendous information for very complicated molecular structure study.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For a normal FT experiment, the FID (free induction decay) obtained after a rf pulse (or pulses) is dependent solely on the time over which the FID is acquired (t_2). However, the frequency signal can also be dependent on the length of a time delay prior to the detection period. This delay time (t_1) can also be varied. The NMR signal obtained at the end of an experiment would then be a function of two time variables, t_1 and t_2. After two Fourier transformations the NMR signals are displayed in a two-frequency-matrix domain. The information coded in this 2D matrix is only determined by the pulse combination. For example, 2D NOESY (two dimensional Nuclear Overhauser Enhancement SpectroscopY) experiment consists four basic consecutive time periods, as shown in Figure1.2.[1] In the first time period referred to as the preparation time, all perturbed nuclei relax back to their equilibrium states; in the second time period referred to as the evolution time, all of the vectors carrying their characteristic frequencies "show up" in the external magnetic field; the third time period started from the second rf pulse provides an opportunity to allow the coupled-spins to carry other spin's frequency information which are correlated to itself, in different 2D experiments they can carry different correlation information; the final time period started from the third rf pulse is the recording process, from which the frequency domain spectrum can be developed after Fourier Transform. The new information obtained in 2D spectrum is from the cross-peaks (off-diagonal peaks). Different 2D experiments are designed to obtain different molecular structure information.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.1.4 Basic 2D NMR experiments

2D NOESY (Two dimensional Nuclear Overhauser Enhancement SpectroscopY) is by far the most important technique in the large molecular structure analysis, because 2D NOE crosspeak intensities are directly related to proton-proton distances, which can be subsequently incorporated into structural refinement.

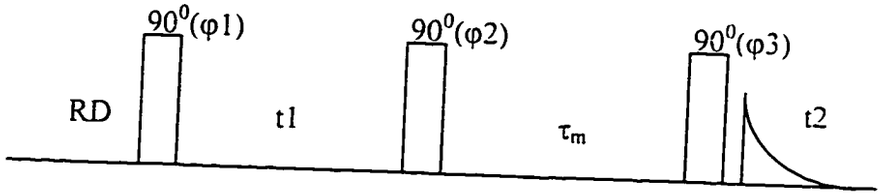
This type of 2D NMR experiments provides protons' interaction through space proximity based on the dipole-dipole coupling, rather than through the chemical bonds as in 2D COSY and TOCSY experiments. When nuclei A and B are close in space, and A has no nearby neighbors but B has several (B′, B″, ...), a decoupling field applied at the resonance frequency of B results in an enhancement of the A signal intensity. However, the situation differs when A is saturated by application of decoupling field. Saturation of A results in the enhancement of signal B, but B's magnetization is transferred to its neighboring nuclei (B′, B″ ...) through dipolar-coupling which subsequently dramatically reduces the signal intensity of nucleus B. Consequently in a steady-state experiment $I_{(A,B)} \neq I_{(B,A)}$. For the symmetrical 2D NOESY, the 2D experiment is *transient*. The intensity of the NOE observed in the transient experiment depends not only on the inter-proton distance, but also the mixing time, therefore the peak intensity of 2D NOESY is a function of mixing time (t_m).

The pulse sequence of 2D NOESY is presented in Figure 1.2 and consists of three 90° pulses, according to the pulse scheme:

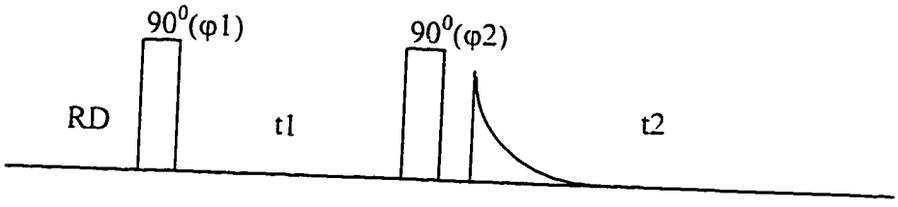
$$\frac{\pi}{2} - t_1 - \frac{\pi}{2} - t_m - \frac{\pi}{2} - t_2$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

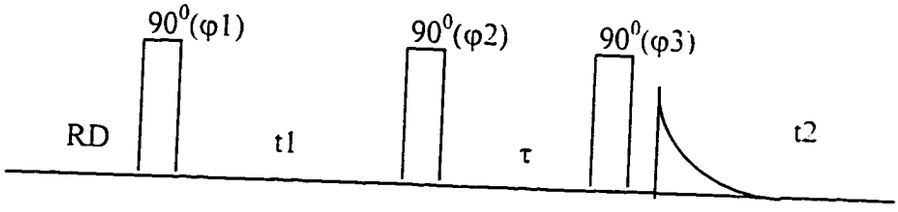
A) NOESY



B) COSY



C) DQCOSY



D) TOCSY

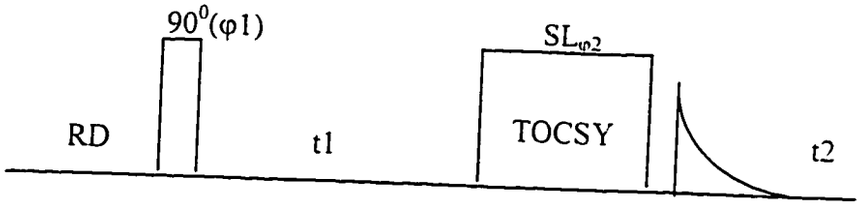


Figure 1.2 Pulse sequences of selected 2D experiments

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where $\pi/2$ represents a selective pulse of 90° on the spins; t_1 is the time allowed for chemical shift evolution; t_m is the mixing time when spins are allowed to "mix" or transfer magnetization; and t_2 is the detection time when the FID of the system is collected. Each repetition of the pulse scheme is delayed by a time factor called the relaxation delay (RD) to allow all magnetization to return to equilibrium before the next pulse scheme is executed.

The length of time chosen for this mixing period (t_m) depends quite dramatically on the properties of the molecule being investigated. Both size and shape of a molecule are important, because both of these effect the correlation time which in turn affects the spin lattice relaxation rate and hence the NOEs observed. When the molecule is small, and moves rapidly in solution, the nuclei have long T_2s and hence long delays are required (as relaxation rates, T_2s, are slow) during which NOEs build up. On the other hand, when the molecule is large, as protein and DNA molecules, or the solution is viscous so that tumbling motions are greatly reduced, and T_2s are reduced so that the NOEs build up in a much shorter time. The correlation time of a molecule (τ_c) can be expressed in the terms of the size of the molecule (α, radius of a spherical particle) and the viscosity coefficient of the solution (η):

$$\tau_c = \frac{4\pi\eta\alpha^3}{3kT} \quad (1.2)$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

where k is the Boltzman constant, and T is the temperature. For small molecules the mixing times (t_m) are of the order 1-5 second and for large molecules the mixing times are of the order of 50-200 millisecond. The 2D NOESY experiment is particularly useful in the study of biomacromolecules where, for a given mixing time, it can provide the complete NOE connectivity in a molecule structure.

COSY (COrrelation SpectroscopY) is designed to provide information about scalar coupled spins. Protons which are coupled through scalar coupling show crosspeaks in the COSY two dimensional spectrum as absorptive antiphase doublets in t_1 and t_2. The pulse scheme for COSY experiment is:

$$\frac{\pi}{2} - t_1 - \frac{\pi}{2} - t_2$$

The execution of the pulse scheme is delayed by a relaxation delay (RD) period following the acquisition of the FID. The first 90° pulse places the spins out of equilibrium, and different spins are allowed to precess at their characteristic frequencies. The second 90° is called a "mixing" pulse and essentially mixes the components of the transverse magnetization, i.e. it causes coherence transfer between coupled spins. After two times of Fourier transform, coupled spins are displayed as off-diagonal peaks.

DQCOSY (Double Quantum coherence COSY) has a pulse scheme similar to the COSY experiment, where the final 90° pulse has been replaced with a pair of 90° pulses to generate (the second 90° pulse) and select (the third 90° pulse) double quantum coherence. The pulse scheme is:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\frac{\pi}{2} - t_1 - \frac{\pi}{2} - \tau - \frac{\pi}{2} - t_2$$

The advantage of this experiment is that solvent peaks are eliminated and it simplifies peaks identification, but the tradeoff is that the signal-to-noise ratio is reduced by a factor of two.

TOCSY (TOtal Correlation SpectroscopY) is also based on J-transfer of magnetization between spins. The feature of the TOCSY is that magnetization transfer can occur through multiple bonds, not only through the three or four bond limit of the basic COSY experiment. The extent of magnetization transfer depends on the length of the mixing time, the magnitude of each of the coupling constants and the linewidth of the peaks. TOCSY also has the advantage that the crosspeaks are pure absorptive in-phase doublets in t_1 and t_2, making the experiment more sensitive and easier to interpret. The pulse scheme for the TOCSY experiment is:

$$\frac{\pi}{2} - t_1 - SL - t_2$$

where SL symbolizes a spin lock which holds the magnetization along one axis of the XY plane. The result of the spin lock is that the only effective remaining part of the Hamiltonian is the scalar coupling term.

Depending on the molecules and the desired information under interest, many different 2D experiments can be used. Many new experiments have become available and surely more sophisticated experiments will emerge as the demand for biomolecule structure study become more challenging.[6]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.1.5 Multi-dimensional NMR and 3D NOE-NOE

For 35mer oligonucleotides or 20kD proteins the 2D NMR experiment reaches its resolution limit.[7,8] However, higher resolution can be achieved by extending NMR experiments to a higher dimension. In principle, a 3D NMR experiment involves expanding a 2D NMR experiment into the third dimension to resolve the overlapped peaks. While many different types of 3D experiments have been developed and applied to protein,[7] 3D NOE-NOE can be used for DNA studies in a similar fashion. The pulse sequences for 3D NOE-NOE is shown in Figure 1.3. The pulse scheme of 3D NOE-NOE is basically the same as that of 2D NOESY, but two more pulses and an additional mixing time are added. The forth 90° pulse gives all the spins a second chance to transfer magnetization among their neighboring protons, and the magnetization transfer is optimized by selecting a suitable second mixing time (τ_2). The last 90° pulse bring all the spins back onto the XY plane for detection. Experimentally, the two mixing times are usually set equal. In this pulse sequence, there are three time variables, t_1, t_2 and t_3. After three times Fourier Transformation, the time domain signals are transferred into frequency domain signals. In this cubic 3D frequency domain, one frame (slide) of the cube is a single 2D NOSEY. Clearly 3D NOE-NOE is a set of stacking of many 2D's in the third dimension and it will provide higher experimental resolution than a single 2D NOESY, even though crosspeak intensities in 3D are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

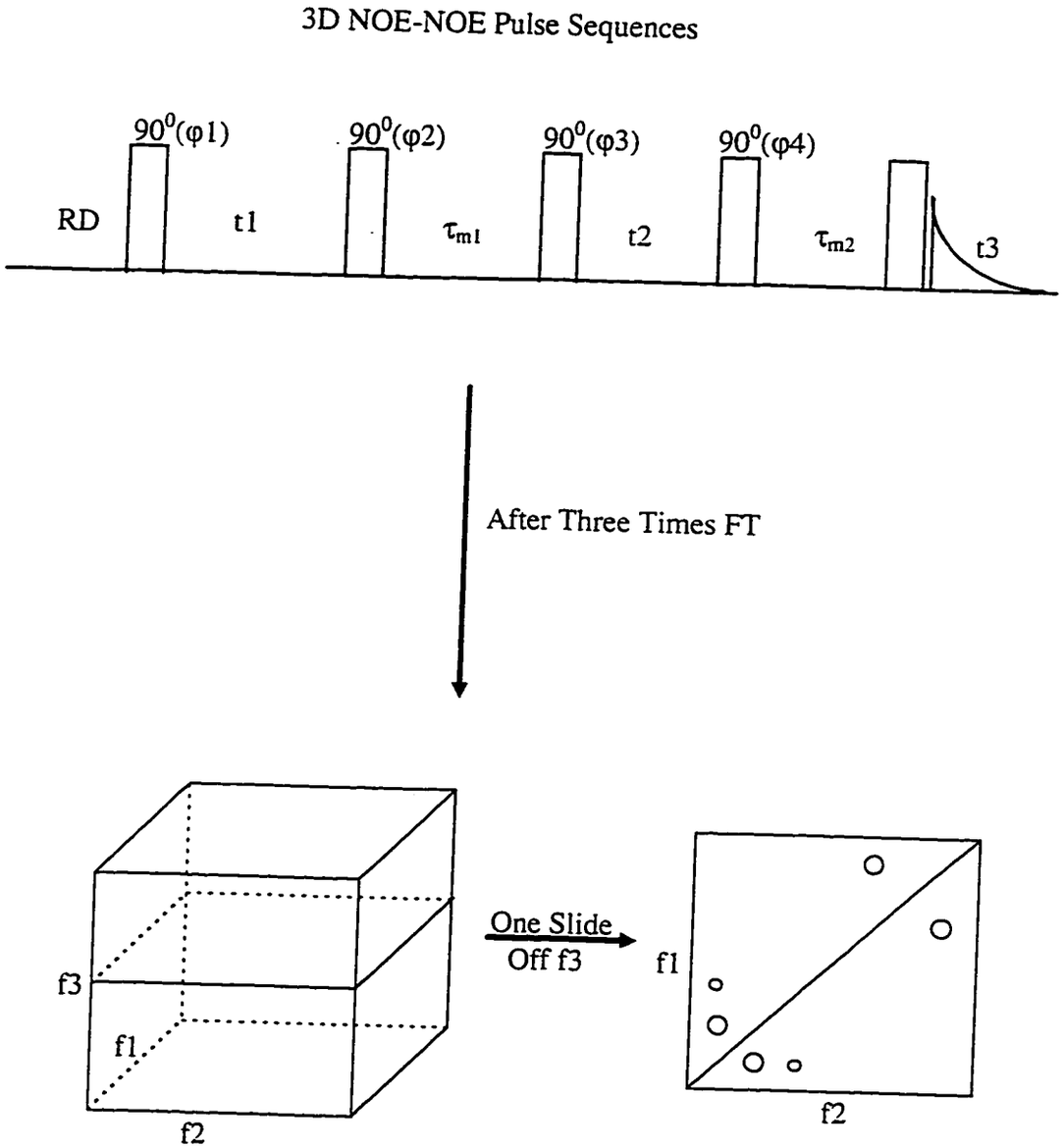


Figure 1.3 The Pulse sequences of 3D NOE-NOE. Its spectrum in frequency domain after three times Fourier Transformation, and one 2D plane cut off the 3D spectrum are shown.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

associated with higher experimental error. For a 3D NOE-NOE the intensities of cross-peaks, A_{ijk}, are proportional to the two magnetization transfer efficiencies, and it can be approximated as a product of the two individual intensities in 2D

$$A_{ijk} = A_{ij} A_{jk} \quad (1.3)$$

where A_{ij} and A_{jk} are the transfer efficiencies between protons i, j and j, k, respectively. Based on this, it is possible to back extract the 3D NOE-NOE intensities' information into 2D NOESY's (deconvolution). These extracted data can be used as 2D "experimental" data, and be combined with real 2D experiment data to refine biomolecule structures. This approach will be discussed in great detail in the next chapter. Several other methods for utilizing 3D NOE-NOE data in structure refinement have been suggested,[9,10] but the deconvolution method may be more promising due to its quality and efficiency for large DNA molecules.

1.2 Basics of nucleic acid structure

1.2.1 Basic structure of DNA

Single-stranded DNA is a polymer of mononucleotides unit composed by phosphodiester, β-D-2-deoxyribofuranose sugar and the four nucleus bases: adenine, guanine, thymine and cytosine (Figure 1.4). Double-stranded DNA consists of two single-stranded DNAs, which are anti-parallel each other and held together by inter-strand hydrogen bonding between the bases. Adenine pairs with thymine by two hydrogen bonds and guanine pairs with cytosine by three hydrogen bonds (Figure 1.5).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

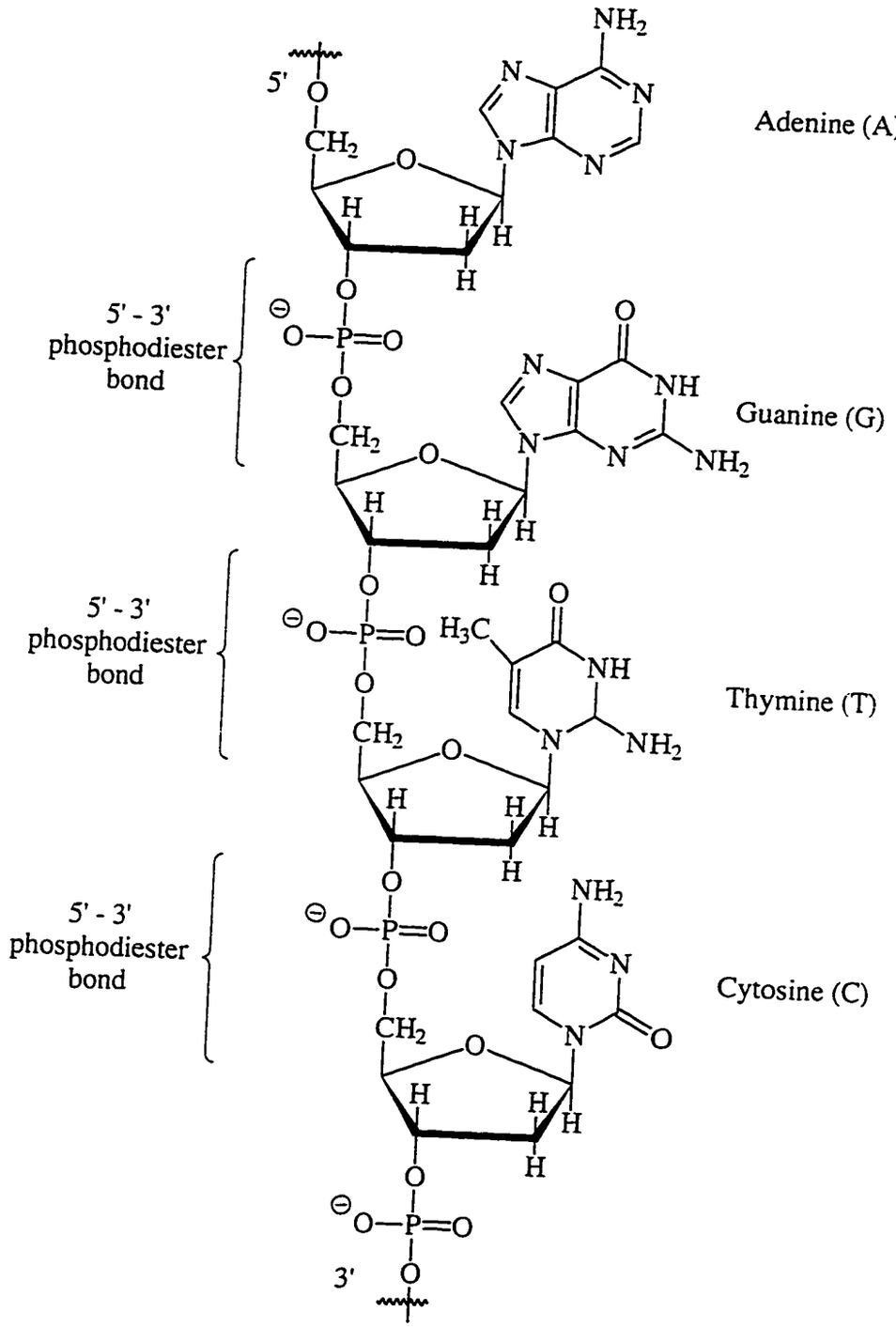


Figure 1.4 Structure of a polydeoxyribonucleotide[11]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

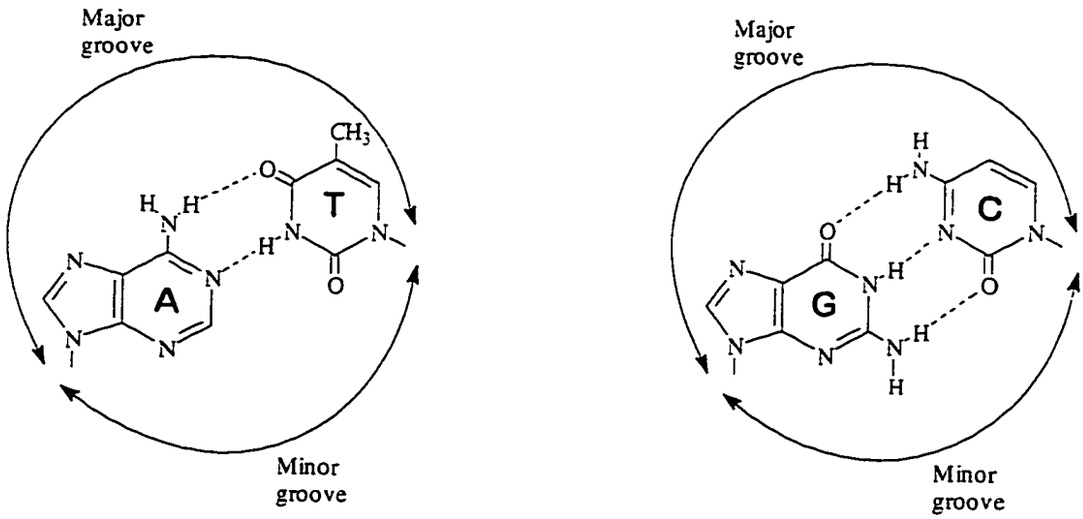


Figure 1.5 Watson-Crick purine-pyrimidine base pairing

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

This DNA duplex structure was first discovered by Watson and Crick in 1950's[12,13] by using X-ray fiber diffraction. This DNA duplex structure elucidation is one of greatest discoveries of the century. Under different conditions, double-stranded DNA is able to adopt at least three different conformations, A-form, B-form and Z-form (Figure 1.6). The structural parameters of each form are listed in table 1.1. A-form DNA exists in high salt concentration, Z-form DNA requires repetitive dinucleotide track, and by far the most common one is the B-form DNA.

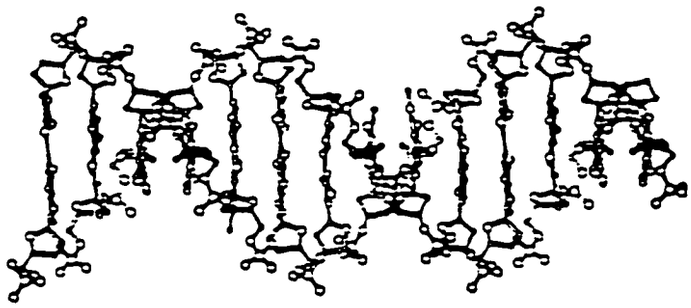
Also DNA has to be viewed as a dynamic molecule as all other molecules, because there are many dynamic motions associated with it, like propeller twist, helical twist, base roll and tilt, and slide[14] (Figure 1.7). These fast motions are averaged and may not show up when NMR technique is used to probe the DNA structure.

1.2.2 NMR parameters of DNA

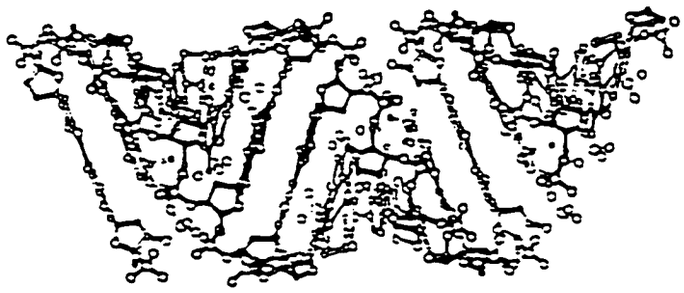
The determination of solution structure of DNA is based on the information of inter-proton distances, which are obtained from nuclei Overhauser enhancement spectroscopy, and bond angles, which are extracted from proton-proton coupling constants. These parameters are used as restraints in DNA molecular dynamics simulation to determine a "real" structure. The first step in extracting information from NMR spectroscopy is to make complete resonance assignments by sequential assignment methodology.[2] Before one starts the assignments, the table of Approximate ^{1}H Chemical Shift Ranges in Single Stranded and Duplex DNA, table 1.2,[2] should be consulted. Because the ^{1}H chemical shifts in DNAs are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A) B-form DNA



B) A-form DNA



C) Z-form DNA.

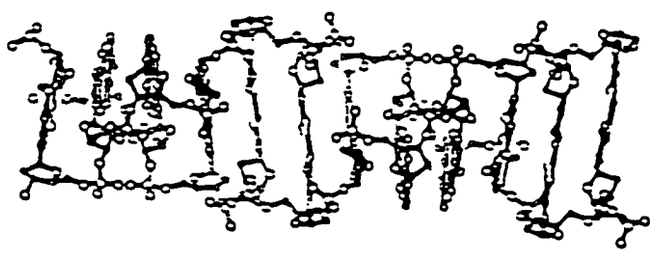


Figure 1.6 A-form, B-form and Z-form DNA[15]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1.1 Structural parameters for A-form, B-form and Z-form DNA

Helix Parameter	A-DNA	B-DNA	Z-DNA
Handedness	right	right	left
Bases/Turn	10.9	10.0	12.0
Height per Base	2.9 Å	3.4 Å	GC 5.3 Å CG 4.1 Å
Pitch	31.6 Å	34.0 Å	45.6 Å
Glycosyl Angle	anti	anti	C anti G syn
Sugar Pucker	3'C endo	2'C endo	C 2'C endo G 3'C endo
Base Pairing	W-C	W-C	W-C
Repeating Helix Unit	1 bp	1 bp	2 pb
Twist/base pair (bp)	33°	36°	GC -51° CG -9°
Axis Displacmnt	4 Å	0	-3 Å
Major Groove	very deep	deep, wide	shallow
Minor Groove	shallow	deep and narrow	very deep
Base Incline	13°	-2°	9°
Base Roll	6°	-1°	3°
Propeller Twist	15°	12°	4°

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

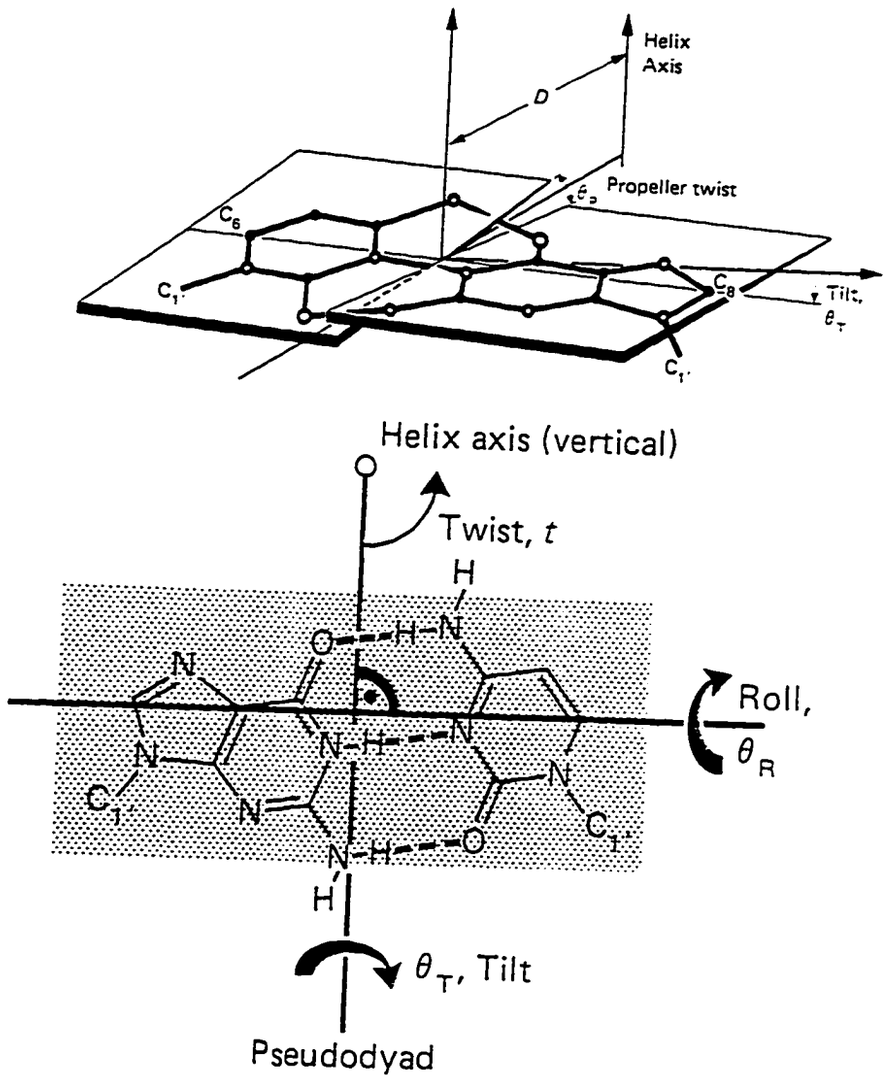
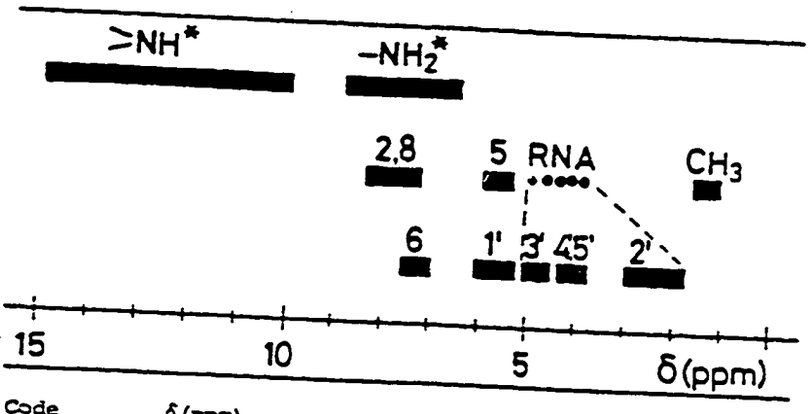


Figure 1.7 DNA local motions: twist, roll, tilt.[14]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1.2 Approximate ^1H chemical shift ranges in single stranded and duplex DNA and RNA fragments[2]



Code	δ (ppm)	Comments
2'	1.8–3.0	2'H, 2"H in DNA
4',5'	3.7–4.5	4'H, 5'H, 5"H in DNA
3'	4.4–5.2	3'H in DNA
•••••	3.7–5.2	2'H, 3'H, 4'H, 5'H, 5"H in RNA
1'	5.3–6.3	1'H
CH_3	1.2–1.6	CH_3 of T
5	5.3–6.0	5H of C and U
6	7.1–7.6	6H of C, T and U
2,8	7.3–8.4	8H of A and G, 2H of A
$-NH_2$*	6.6–9.0	NH_2 of A, C and G
≥NH*	10 – 15	Ring NH of G, T and U

* The rather wide ranges result from the sequence effects on the shifts and for the labile protons (identified by *, observable only in H_2O) also from hydrogen bonding.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

relatively constants. The assignment process begins with the cytosine's H5 and H6 scalar coupled crosspeak(s). These crosspeaks are quite intense since they are highly coupled. The NOE intensity of this H5/H6 proton pair is used as internal distance ruler because the cytosine ring is considered rigid in DNA duplex environment. Also usually evident in the correlated spectrum is the presence of coupling between the thymine methyl protons and its H6. Once a few of the cytosine and thymine base chemical shifts have been identified, one can continue with the sequential assignment method. Each base proton (A's H2, H8; G's H8; C's H6 and T's H6) is close enough falling into the NOE's 5Å cutoff to show NOESY crosspeaks with its neighboring protons on the same nucleoside residue: H1′, H2′, H2′′ and also will likely have crosspeaks with the H1′, H2′, H2′′ of the residue next to it in 5′ direction. The 5′ end residue can be easily identified since its base protons can only have crosspeaks with its own deoxyribose. By walking along the DNA from base proton to base proton, the assignments should become clear for all the H8, H6, H1′, H2′ and H2′′ protons[2] (Figure 1.8). Once these assignments are confirmed by checking crosspeaks for H1′ to H2′ and H1′ to H2′′, other regions of the spectrum can be checked for additional information on the other protons in the molecule such as crosspeaks between H3′ and H2′/H2′′, H6/H8 and H3′, H4′ and H5′/H5′′. To all the assignable protons, the entire data set must be self-consistent. Some other NMR experiments may be necessary to confirm the assignment analysis, like different mixing times NOESYs and TOCSYs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

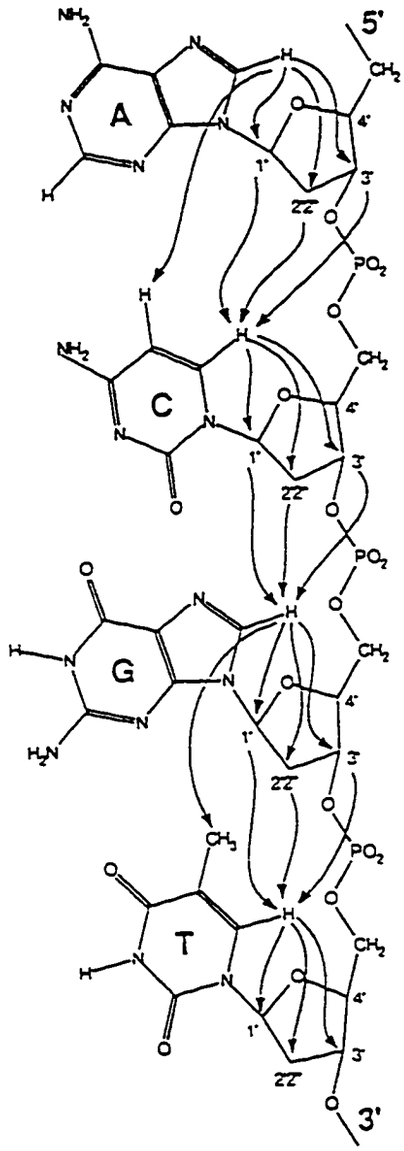


Figure 1.8 NOE connectivity along a DNA duplex strand. The arrow traces from base proton to sugar proton, and then to the next base proton in the 5′ to 3′ direction.[2]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.3 Iterative relaxation matrix/MD simulation methodology in structure refinement

Since the NOESY experiment is the most important experiment in terms of getting interproton distances of solution biomolecules,[16] several methods[17-20] have been proposed to use the NOESY crosspeak intensities as the interproton distance information to refine the corresponding biopolymer structure. Until a few years ago, the calculation of the interproton distances from the NOE data used the so-called two spin approximation. This assumption neglects the significant contribution of the nuclei surrounding the proton pairs under consideration (spin diffusion, magnetization leaking), even though this method has the advantage of lower computational demand. Recently, methods have been developed to use the experimental NOE intensities directly in structure refinement instead of interproton distances.[9,21] Efforts are still being made to improve the efficiency and quality of refinement, including saving computer time and improving the accuracy and precision of refinement.

Another alternative methodology in using NOE crosspeak intensities for structure refinement is the full relaxation rate matrix approach,[16] coupled with molecular dynamics simulation. MORASS was developed based on this approach, and it has been successfully used for oligonucleotide structure refinement with 2D NOESY data. Further expanding of this approach for utilizing 3D NOE-NOE is the topic of the next chapter.[22]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1.3.1 NOESY data and interproton distance calculation

The evolution of the NOE intensities as a function of mixing time is described by the Bloch equation[16,23]

$$\frac{d}{dt}V(t_m) = -\Gamma \bullet V(t_m) \qquad (1.4)$$

where Γ is a symmetrical relaxation rate matrix, and $V(t_m)$ is the NOE volume (intensity) matrix of the molecular system under consideration at mixing time t_m. The above equation can be solved either by Taylor expansion approximation,[16] or by treating it as an eigenvalue problem for precise solution. In the later approach, the formal solution to this equation is given by

$$V(t_m) = \exp[-\Gamma(t_m) \bullet V(0)] \quad (1.5)$$

which can be rearranged as follows

$$\Gamma = -(\frac{1}{t_m}) \ln \frac{V}{V_o} \qquad (1.6)$$

where V_o is a diagonal volume matrix at zero mixing time (no off diagonal peaks at zero mixing time as NOE has no time to build up yet). Finally, the Bloch equation can be solved by obtaining the following expression for the rate matrix

$$\Gamma = -(\frac{1}{t_m}) K[\ln(L)]K^{-1} \qquad (1.7)$$

where K is the eigenvector matrix and L contains eigenvalues of V/V_o, respectively. Each element of the rate matrix Γ is related to the corresponding spin spin distance by equations

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$\rho_{ii} = \sum_{j \neq i}^{n} \frac{\gamma^4 h^2}{10\left\langle r_{ij}^{\ 3} \right\rangle^2} [J(\omega) + 3J(0) + 6J(2\omega)] \quad (1.8)$$

$$\rho_{ij} = \frac{\gamma^4 h^2}{10\left\langle r_{ij}^{\ 3} \right\rangle^2} [6J(2\omega) - J(0)] \quad (1.9)$$

where γ is the gyromagnetic ratio for ^{1}H nuclei. J(ω) is the spectral density function describing the transition probabilities for a single rotational correlational time τ_c, and it is given below

$$J(\omega) = \left\{ \frac{\tau_c}{1 + \omega^2 \tau_c^{\ 2}} \right\} \quad (1.10)$$

This set of equations is the link bridging the NOE volumes with interproton distances. They are used repeatedly from structure (interproton distances) to NOE volume matrix and from NOE volume matrix to interproton distances (structure).

If a 2D NOESY experiment could provide a complete set of NOE data, then its back calculation to distances would be one step. But in reality, the experimental NOE data set is usually a sparse matrix, due to no NOE peaks for $d > 0.5$ *nm*. One solution to this approach is the hybrid relaxation matrix[24-27] (Figure 1.9). A reasonable initial model (usually A-DNA and/or B-DNA) is used to generate a complete simulated 2D NOE volume matrix $V_{simulated}$ with mixing time t_m from the relaxation rate matrix Γ. Experimental NOE data are merged with the simulated NOE volume gradually, creating a new hybrid volume matrix V_{hybrid}. When the $V_{simulated}$ is hybridized with the incomplete $V_{experiment}$, one of the two has to be scaled in order to make these two data sets comparable. This is usually done by using the cytidine H5/H6 crosspeak

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

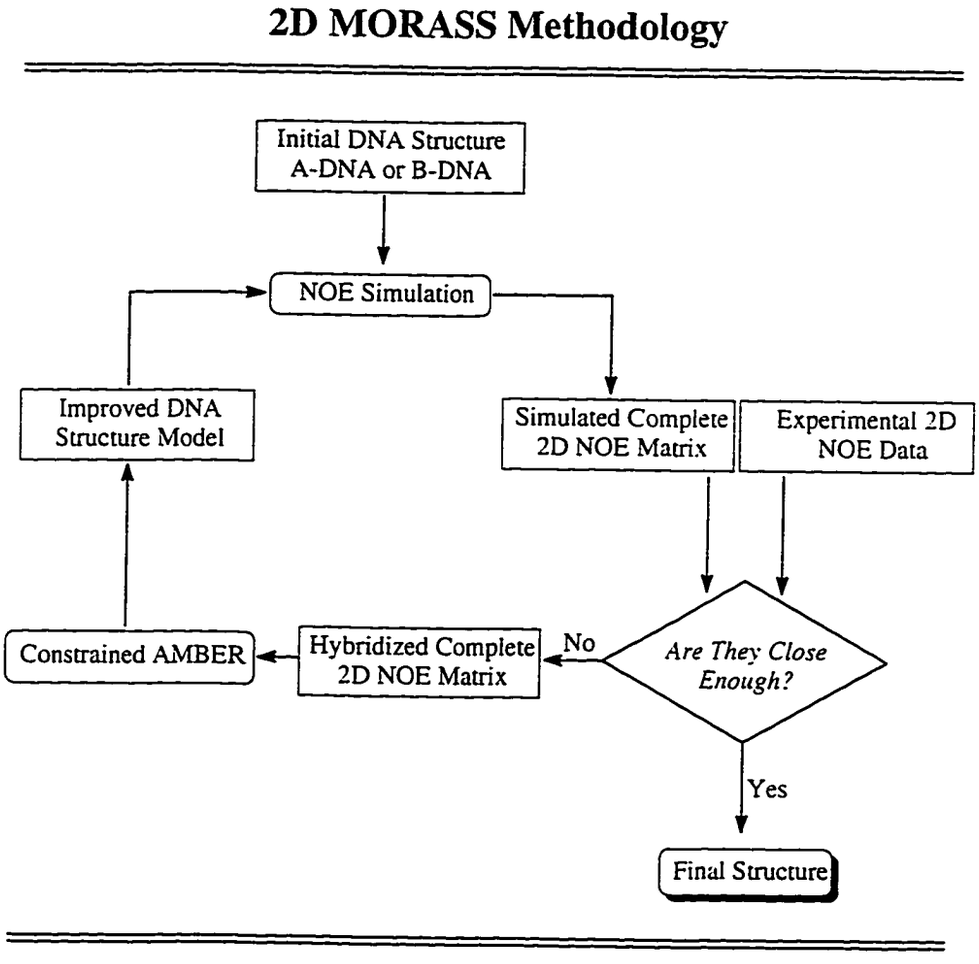


Figure 1.9 Flow chart of MORASS methodology[28]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

intensity as the internal ruler. This newer volume matrix V_{hybrid} can be used to back-calculate a newer relaxation rate matrix Γ, which in term can generate a newer structure after molecular dynamics energy relaxation. This second generation structure should be closer to the real solution DNA structure since it contains some experimental data. This newer structure is used for the second round simulation, and this loop is repeated until a set of criteria is met. During the simulated NOE and experimental NOE merging process, experimental data is only partially merged in, and the merging rate is based on the absolute difference between the experimental and simulated data pair. This is because the initial model structure needs to be gently re-shaped rather than one step dramatic forced change.

This refinement calculation algorithm was developed by Dr. Gorenstein's group with FORTRAN coding, and termed MORASS (Multi-spin Overhauser Relaxation AnalysiS and Simulation). Figure 1.9 shows the overview of 2D MORASS package.

1.3.2 Molecular dynamics simulation

If pure molecular mechanics and molecular dynamics (MM/MD) simulation can have some experimental data input, its results would make more sense and be more reliable. The process for extracting experimental data (interproton distances) from 2D NOESY for MM/MD simulation is discussed above. This section discusses the basic principles used in MM/MD simulation.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The theory used in MM/MD simulation assumes that all the motions associated with a biomolecule in solution is a classic motion, such that the classic Newton's equations can be applied to it[29]

$$\frac{dr_i(t)}{dt} = v_i(t) \qquad (1.11)$$

$$\frac{dv_i}{dt} = m_i^{-1} F_i(t) \qquad (1.12)$$

The Cartesian position vector r_i and the velocity v_i of atom i (with atomic mass m_i) are functions of time (t). The force F_i exerted on atom i by the other atoms in the molecular system is given by the negative gradient of the atomic interactions' function (V)

$$F_i = -\frac{\partial V[r_1(t), r_2(t), \ldots, r_N(t)]}{\partial r_i(t)} \qquad (1.13)$$

where the force field V for biomolecules is empirically assumed to have the following format

$$V = \sum_{bonds} \frac{1}{2} K_b [b - b_0]^2 + \sum_{bondangles} \frac{1}{2} K_\Theta [\theta - \theta_0]^2 + \sum_{improperdihedrals} \frac{1}{2} K_\varsigma [\varsigma - \varsigma_0]^2 +$$
$$\sum_{dihedrals} \frac{1}{2} K_\varphi [1 + \cos(n\varphi - \delta)] + \qquad (1.14)$$
$$\sum_{allpairs(i,j)} \left[\frac{C_{12}(ij)}{r_{ij}^{12}} - \frac{C_6(ij)}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}\right] + \sum_{H-bonds} \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r^{10}_{ij}}\right]$$

The first four terms represent contributions from bond, bond angles, improper dihedral angles and dihedral angles, respectively. The fifth term is a sum over all pairs of atoms and represents for the effective non-bonded interaction that including the van der Waals and the Coulomb interactions between atoms i and j with charges q_i and q_j at a distance

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

r_{ij}. The last term accounts for hydrogen bonding. In order to carry out the full calculations, all the force field parameters such as K_b, K_θ, etc, should be known or assumed. After equation 1.13 is integrated on the time scale the Cartesian positions of all atoms at a particular time point is known, and therefore the molecular motion trajectory (molecular configuration as a function of time) is known.

If molecular dynamics simulation is carried out with no restrains, it will simply generate a trajectory. But if experimental data, like NOE intensity and coupling constant, is used as restrains in the molecular motion calculation, this structure will be forced to move into a limited motional region. Structures in this motional region are termed "final structures" or "refined structures". Temperature also plays a role in structure refinement process, as sometimes temperature will be "elevated" to certain point and then "annealed" down for searching global energy minimum.

In MORASS package, the NOE intensities (distance restrains) are used in a pseudo-harmonic potential function as Figure 1.10. The energy term caused by the NOE restrains can be expressed as:

$$V_{NOESY} = \begin{cases} K_L[r-(r_0 - L_{err})]^2 & r < (r_0 - L_{err}) \\ 0 & \\ K_R[r-(r_0 + R_{err})]^2 & r > (r_0 + R_{err}) \end{cases} \qquad (1.15)$$

where K_L and K_R represent the lower and upper bond force constants, R_{err} and R_{err} are the upper and lower assumed error bars. r Is the actual interproton distance and r_0 the target distance restraint from the NOE.[30] This potential function shows the NOE

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

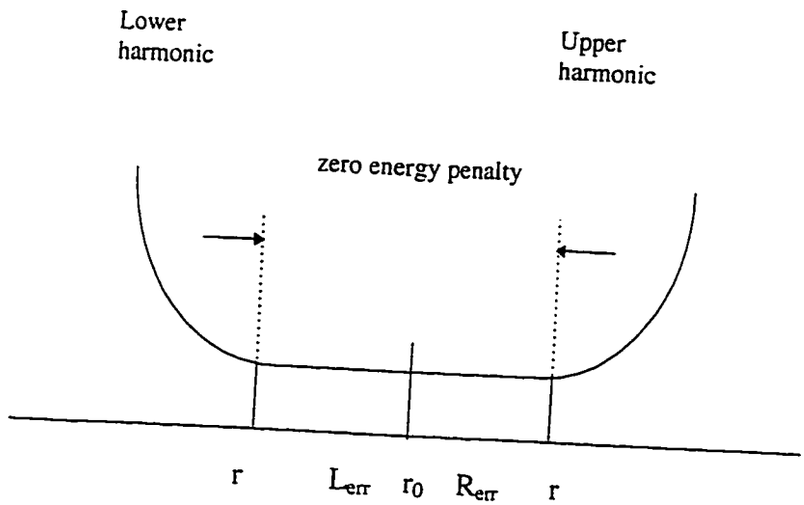


Figure 1.10 Pseudo-harmonic potential function used in distance restrained structure refinement

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

derived distance involvement in molecular mechanics and dynamics and as a driving force to push the initial model structure to the refined structure.

1.3.3 Iterative hybrid matrix/MD refinement

Schematically, the NOE data incorporation in molecular dynamics (MD) and the MORASS structure refinement process is shown in Figure 1.9. MORASS is used for 2D NOE matrix simulation from known structure and interproton distance calculation from a NOE matrix. AMBER[29] is used for MD simulation with flatwell modification. These two combined is known as the MORASS/AMBER refinement protocol. First, a reasonable initial model is built, and the 2D NOE volume matrix is simulated by MORASS. The simulated 2D NOESY is partially merged with experimentally available 2D NOE data according to preset criteria. Second, a set of interproton distances are extracted from the new NOE data matrix by MORASS, and this set of distances is used in AMBER MD run to generate an improved structure by using reasonable force constants and error bars. This newer structure is the starting point of the next MORASS/AMBER iteration. As this iteration refinement process goes on, finally all (or almost all) experimental NOE data are merged into the simulated 2D NOE volume matrix. The quality of the final structure is judged based on the agreement between simulated 2D NOE and experimental 2D NOE, the total energy, energy penalty and other related quantities.

Theoretically, if there are enough experimental 2D NOE data restrains, the final refined structure would be independent of the very initial model being selected.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Practically, refinement is usually started with two different models, like A-DNA and B-DNA. Once the two final "refined" structures are consistent each other in addition to the criteria mentioned above, then those structures are more likely the real structure which the biomolecule adopts in the solution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 2. STRUCTURE REFINEMENT VIA DECONVOLUTED 3D NOE-NOE DATA: THE HYBRID-HYBRID MATRIX METHOD

2.1 Introduction

When the size of a biomolecule becomes very large (larger than 15,000 Dalton), the 2D NMR spectrum becomes very crowded and eventually fewer and fewer significant assignments can be made. This problem can be greatly improved by extending two dimensional NMR to three dimensional NMR experiments. As in 2D NMR, the 3D NOE-NOE data contains interproton distance information that can be used as constrains for structure refinement.

The general background of 3D NMR development is given in chapter one. The focus in this chapter is to extend the 2D MORASS structure refinement methodology to 3D NOE-NOE, and the preliminary results of the verification of the 3D NOE-NOE hybrid-hybrid matrix method are presented here.

It has been shown that 3D NOE-NOE spectroscopy holds great promise for structure refinement of large biomolecules.[31] Several different approaches have been developed to refine structures from 3D NOE-NOE spectra.[17-20] Three-dimensional volumes (A_{ijk}) may be treated by the "two-spin approximation" method, $A_{ijk} \propto r_{ij}^{-6} r_{jk}^{-6}$. The volume depends upon the inverse sixth power of the distances between the two sets

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of spins i, j and j, k. Alternatively, a slightly more accurate method utilizes various terms of the Taylor series expansion of the exponential rate matrix.[17,32,33] However, it was demonstrated that at realistic mixing times, which allow reasonably accurate measurement of 3D NOE-NOE volumes, both the two-spin and the Taylor series expansion approximation methods can lead to considerable systematic errors.[31]

Distance constrains derived from either approximation method may be used in distance geometry[2,34,35] or restrained molecular-dynamics refinement.[30,36,37] Alternatively, a direct NOE gradient refinement method based upon the 3D NOE-NOE volumes has been used.[17] However, gradient methods are generally very computationally demanding. In the original Yip and Case paper,[9] the gradient method requires calculating the product of five matrices, which for large biomolecules is computationally prohibitive. More recent NOE gradient calculations based upon the first two terms of the expansion of the exponential rate matrix greatly reduce the time required.[17] Kaptein and co-workers[17] have shown that their method scales with n^3, where n is the number of spins. Therefore direct gradient refinement methods utilizing matrix calculations of the volumes hold great potential for refinement of large biomolecules from 3D NOE-NOE spectra.

The relaxation-matrix approach[16,23,28,38] avoids the two-spin approximation by employing a matrix eigenvalue/eigenvector solution to the Block equations. Cross-relaxation rates evaluated by a matrix method include effects from multiple relaxation paths, and therefore distances and structures derived from them are more precise and accurate.[16,28,39] However, while the 3D volumes may be more accurately calculated

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from the complete relaxation-matrix method,[31] accurate refinement methods from a 3D relaxation-matrix method have yet to be fully developed for large biomolecules.

For two-dimensional NMR, accurate distances can be directly obtained from the complete relaxation method by diagonalization of the 2D volume matrix which represents a 2D NOESY spectrum.[25,28] Unfortunately, the eigenvalue/eigenvector solutions are very sensitive to the accuracy and completeness of the NOESY volume matrix.[28] A hybrid matrix solution to this problem was originally proposed by Kaptein and co-workers[21,24] and implemented in several programs.[21,24,25,30] This hybrid matrix approach combines the information from the experimental NOESY volumes, A_{ijk}^{exp}, and calculated volumes, A_{ijk}^{calc}, derived from an initial structure. The hybrid volume matrix, A^{hyb}, contains the well-resolved and measurable cross-peaks from the experimental NOESY spectrum, while overlapping or weak cross-peaks and diagonals are from the calculated spectrum. This hybrid volume matrix is then used to calculate the rate matrix, and then the resulting distances taken from the cross-relaxation rates are utilized in a distance geometry or restrained molecular-dynamics refinement of the structure. This process is repeated until a satisfactory agreement between the calculated and observed cross-peak volumes is obtained. In various structure refinements, less than ten iterations are typically needed to achieve convergence with a family of structures consistent with the NOE data. A similar "hybrid-hybrid" relaxation-matrix method may also be extended to 3D NOE-NOE spectra. It represents a more computationally efficient method than the current gradient methods, avoids the assumptions of the various

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

approximation methods, and potentially provides a method for the refinement of larger biomolecules.

The study presented here is an extension of 2D MORASS structure refinement methodology (Figure 2.1). When the two mixing times are set equal in a 3D NOE-NOE spectrum, the 3D cross-peaks can be deconvoluted into 2D cross-peaks by simple division. The 2D data set thus obtained can be used in the standard 2D MORASS protocol. There are three advantages of using 3D NOE-NOE data in this way. 1) The full 3D matrix is not needed in the whole calculation procedure, and a linear table can serve the purpose well. This subsequently eases the computational requirement. 2) The well developed 2D MORASS refinement methodology can be utilized here to incorporate 3D NOE-NOE data into structure refinement. 3) It is already shown that the MORASS method is very sensitive to the size of the NMR experiment data set, so the very large size of 3D NOE-NOE data provides inherent advantage to the MORASS method, even though the accuracy of the data is reduced due to the 3D NOE-NOE volume integration error. Therefore the refinement convergence will be hopefully easier. In general, this method not only provides a simple and straightforward means to utilize 3D NOE-NOE data for structure refinement, but also avoids the two-spin approximation error and the prohibitively high computational demand by the gradient method.

So far, no other method has been tested for nucleic acid structure refinement by using 3D NOE-NOE data, therefore this method may provide a valuable tool for large

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

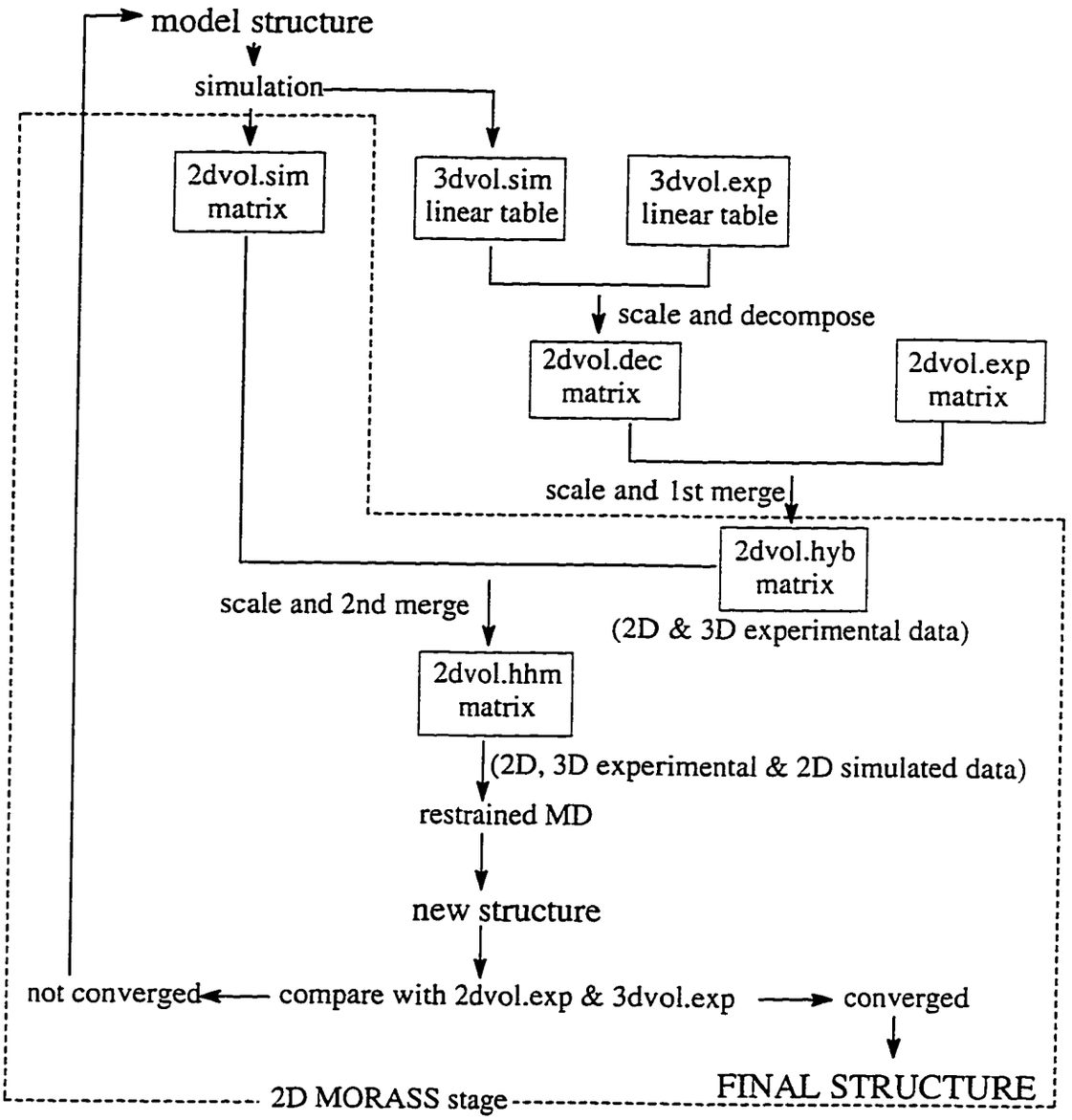


Figure 2.1 Flow chart of structure refinement by hybrid-hybrid matrix method of 3D NOE-NOE data analysis through the MORASS/MD protocol

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

nucleic acid structure refinement which structures can not be refined from 2D NOE data alone. Real structure refinement of a three-way-junction DNA by using 2D NOSEY and 3D NOE-NOE data sets is still under way. The following sections provide the theoretical discussion and computational simulation of the method.

2.2 Theory and method

The intensity of a 3D NOE-NOE cross-peak is proportional to the product of the magnetization transfer efficiencies of the corresponding two individual 2D cross-relaxations. The intensity of a 2D cross-peak between spins *i* and *j* in a multi-spin system can be described by the relaxation matrix *R*[3,16,38]

$$A_{ij}(\tau_m) = c\left[e^{-R\tau_m}\right]_{ij} A_{ii}(0) \qquad (2.1)$$

where τ_m is the mixing time, c is a proportional constant and $A_{ii}(0)$ is the equilibrium magnetization of spin i. A 3D NOE-NOE cross-peak intensity comes from two 2D NOE (two times magnetization transfer) and is given by

$$A_{ijk}(\tau_m^{(1)},\tau_m^{(2)}) = c\left[e^{-R\tau_m^{(2)}}\right]_{jk}\left[e^{-R\tau^{(1)}{}_m}\right]_{ij} A_{ii}(0) \qquad (2.2)$$

where $\tau_m^{(1)}$ and $\tau_m^{(2)}$ are the two mixing times. R is the relaxation matrix, and A_{ijk} represents a 2D plane cut through the 3D NOE-NOE spectrum, or:

$$A_{ijk}^{(3D)} = cA_{ij}^{(1)}A_{jk}^{(2)} \qquad (2.3)$$

where $A_{ijk}^{(3D)}$ is a single 3D peak intensity and A_{ij} and A_{jk} are the 2D peak intensities during the two mixing times. In order to simplify the calculation two mixing times can be set same: $\tau_m^{(1)} = \tau_m^{(2)}$. In this way, equation (2.3) can be simplified as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$A_{ijk}^{(3D)} = c\, A_{ij}\, A_{jk} \qquad (2.4)$$

After rearrangement, equation (2.4) can be rewritten as:

$$A_{ij} = A_{ijk}^{(3D)} / (A_{jk}\, S) \quad (2.5)$$

The implication of equation (2.5) is that a 2D peak intensity can be obtained from correlated one 3D peak and one 2D peak via a scaling factor *S*. We termed this process, in which 2D peak intensity information is extracted from 3D spectrum, deconvolution. Theoretically, 3D NOE-NOE spectra contain the same amount of information (peak intensity or proton distance) as 2D spectra do, such that we can gain no more intensity data. But practically, as biomolecules become larger and larger, the 2D NOESY spectra will not provide enough resolvable intensity information, which is necessary to refine the structure. In such circumstances, 3D NOE-NOE can offer much superior spectrum resolution and their peak intensity can be deconvoluted into 2D peak intensity, which could be used in the well developed 2D MORASS protocol for structure refinement.

In deconvoluting a 3D peak into a corresponding 2D intensity, the denominator (A_{jk}) in equation (2.5) can be derived from three sources. Firstly, if the 3D spectrum have enough resolution, the A_{jk} peak intensity can come from the cross-diagonal peaks by equation

$$A_{jk} = \sqrt{A_{kjk}} \qquad (2.6)$$

Secondly, it can come from 2D experimental data after proper scaling, if this A_{jk} is a resolvable in 2D NOESY. Thirdly, it can be a simulated 2D cross-peaks. This may seem not appropriate because the simulated data are based on the initial assumed model (A-DNA or B-DNA) and that simulated 2D peak intensities are not quite accurate. But

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

what must be emphasized is that structure refinement process is an iterative process. As the refined structure gets closer and closer to the final structure, the simulated 2D peaks, which already contain experimental 3D and 2D data information, will be more and more accurate and contribute greater and greater to the structure refinement process.

Because two independent 3D data sets are merged, the hybrid-hybrid matrix approach will rely heavily on careful scaling. One way to scale the experimental and theoretical volume matrices is to match the volumes of several "markers".[21,24,26] Some of the back-transfer volumes (A_{jkj}) in the 3D spectrum will hopefully be well resolved and correspond to spin pairs of fixed distance, much as proton pairs one would use as reference volumes in the 2D hybrid matrix or two-spin methods. Alternatively, all experimental volumes may be used for scaling. The ratio of the sum of these volumes gives the appropriate scale factor, S,

$$S = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{ijk}^{sim}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{ijk}^{exp}} \qquad (2.7)$$

where the summation is taken over all the experimentally integrated volumes. The scaling factor needs to be calculated in each iteration as the model structure changes in each iteration step. In this study, the linear regression statistics method was used for the scaling of both 2D and 3D data.

Theoretically, a 2D NOE matrix can be obtained from any single plane in the 3D spectrum. The i and j in equation (2.5) run through all spins in the system. However, a given individual plane in the 3D matrix is invariably incomplete due to numerical truncation, such that a 3D peak will not be observable experimentally if the distance

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

between either of spin pair i, j or j, k is greater than 5.0Å. However, since each plane will be incomplete in a region different from other planes, the full 2D data set will be recovered by deconvolution. This is done by deconvoluting each of the incomplete planes in the 3D matrix into separate incomplete 2D volume matrix elements using equation (2.5). Then the 2D planes are combined to form a more complete 2D NOE volume matrix. When more than one 2D plane contains an identical magnetization transfer between two spins, the peak volume is averaged to give out a single element $<A_{ij}>$. Errors are minimized in this treatment by averaging many computed A_{ij} values derived from many 3D volumes:

$$<A_{ij}>=\frac{1}{n}\sum_{k=1}^{n}\frac{A_{ijk}^{exp}}{A_{jk}} \quad (2.8)$$

When the spin number is more than 150, a three dimension array structure will be very large, so linear tables are used to arrange all the 3D experimental data: L1(20000) for 3D back-transfer peaks, L2(20000) for 3D cross peaks and Tb(20000) for 2D peaks (Figure 2.2). The linear tables add less computational requirements to the computation than the full matrix, because generally only a few thousands of additional 3D data are treated rather than many zero elements contained in the n^3-sized three dimensional array.

The schematic protocol of 3D MORASS is shown in Figure 2.1. The tabulated 3D NMR experimental data are decomposed into 2D NMR "experimental" data, and in this decomposing process some 3D simulated data may be needed and if so they are simulated from the initial model. After deconvolution the newly generated 2D

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

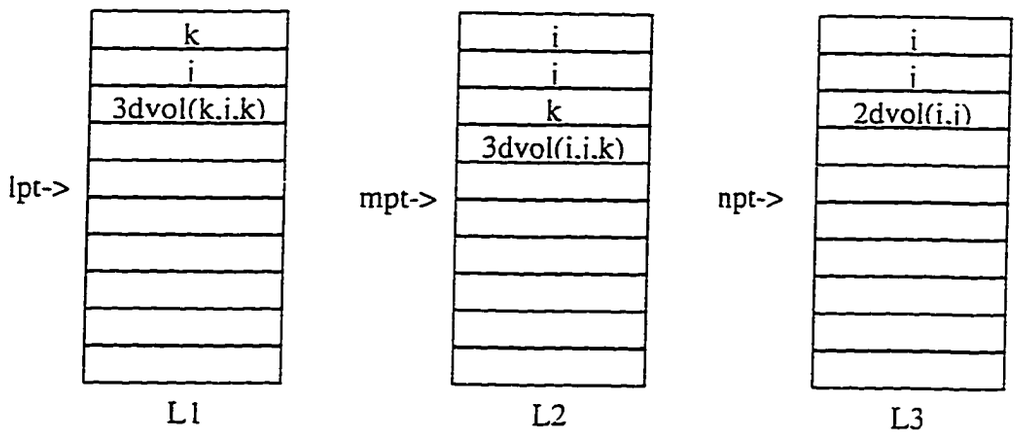


Figure 2.2 Scheme of linear table for 3D NOE-NOE volume data[40]

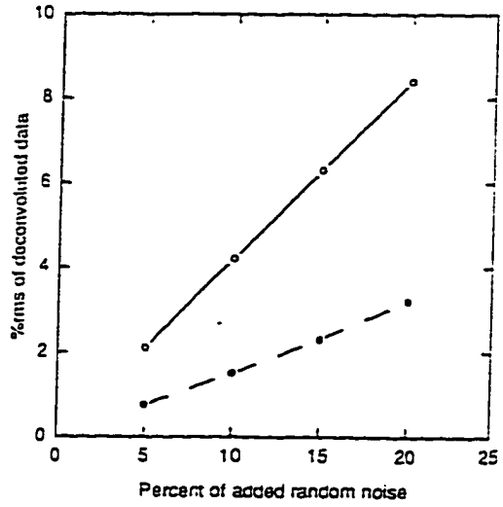


Figure 2.3 The %rms error in the volumes of the deconvoluted 2D volumes plotted against the amount of random noise added to the simulated 3D data set. The perturbed data were deconvoluted and the rms error was calculated for each plane (solid line) and the resulting rms after the planes were averaged together (dashed line). Only a single 2D volume is shown since the remainder of the data is virtually identical.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

"experimental" data are merged with real 2D experimental data set by proper scaling. Usually there are some common peaks which appear in deconvoluted 2D data sets and in real 2D experimental data sets, whose intensities are used as rulers for the first merge operation. The 2D hybridized experimental data matrix is merged with simulated 2D full matrix to generate a new 2D intensity full matrix which contains some experimental data. This merge is a process in which the experimental data is gradually combined with the simulated data, because one-step full merge (using all experimental data to replace the corresponding simulated data) will mathematically cause the matrix equation to become unresolvable and chemically make some parameters nonsense (e.g. bond lengths, bond angles unrealistic). The ruler used in the second merge is the H5 and H6 proton pair of cytidine because the distance between these two protons is "fixed" in a DNA structure. If there are more than one cytidine present in the structure of interest. the averaged distance could be used (end cytidines not included). The 2D intensity matrix can generate the relaxation rate matrix. The inter-proton distances, extracted from the rate matrix, are used in molecular dynamics calculation in order to release the structural strained energy. In this way, a newer structure is generated and if the quality of this structure meets some preset criteria, this structure is the final structure or one of the final structures, otherwise the newer structure is used as the newer model structure and next iteration step should be performed. Generally, less than ten iterative steps are needed to obtain convergence of the initial models.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2.3 Results and discussion

Before this hybrid-hybrid matrix method can be used as a developed method to refine biomolecules, its algorithm needs to be tested against a known structure (well refined structure) and its corresponding 2D/3D experimental data.

The first thing tested was the 3D NOE-NOE integration error minimization through averaging, equation 2.8. A 3D NOE-NOE spectrum of a 12-mer GG mismatch d(CGGGAATTCGCG)$_2$ (this structure had been well refined by 2D MORASS with 2D NOESY, final %RMS$_{vol}$=29) duplex was simulated, and 5-20% random integration noise proportional to the 3D volumes was added. The back-transfer volumes, A_{jkj}, were used to calculate an initial estimate of A_{jk} (=$[A_{jkj}]^{1/2}$). The 3D spectrum was deconvoluted with the method as described under Theory and Methods. The calculated averaged 2D volumes $<A_{ij}>$ were compared to the exact simulated values. Figure 2.3 shows a plot of the %RMS$_{vol}$ deviation of the A_{ij} data after deconvolution as a function of the added random noise. As expected, the resulting %RMS errors for a singled deconvoluted A_{ij} volume derived from the simulated (plus noise) 3D volumes are about one-half of the errors introduced into the 3D volumes. However, when the planes are linked together and the $<A_{ij}>$ values derived from the simulated (plus noise) 3D volumes are compared to the target A_{ij}, the error is dramatically reduced two to three fold, due to the effects of averaging. This result clearly tells us that even though the 3D NOE-NOE data are associated with large integration error, after averaging the data quality can be substantially improved, which in term improves the refined structure.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The second confirmation is the series of algorithm leading to the 3D volume simulation. Based on the algorithm discussed, forty-six representative true 3D experimental peaks ($f_1 \neq f_2 \neq f_3$) are compared against the data from the full matrix calculation and the NMR-derived-solution structure. These forty-six experimental 3D NOE-NOE spectral data (mixing time of 100 ms for both) of the 12-mer GG mismatch duplex were integrated using FELIX2.05 or 2.10 software (Biosym, Inc.). The structure of the 12-mer GG mismatch is the final refined structure. These data were chosen because both an A_{ijk} and a corresponding back-transfer volume A_{jkj} could be reliable integrated. As shown in Figure 2.4, a linear regression correlation coefficient of 0.91 was obtained. With regard to the 0.91 regression correlation coefficient, one need to bear in mind that 50% relative integration error in 3D NOE-NOE experimental data is common. Again, it is clear that the simulated data set is very well correlated to the experimental data set, and this indicates the validity of the 3D simulation algorithm.

Due to the incompleteness of the 3D data set and the requirement of the proposed the hybrid-hybrid algorithm (Figure 2.1), it is necessary to simulate some 3D volumes and deconvolute the 3D merged volumes into a 2D volume data set. It is necessary to test the calculation for the deconvolution process. The 12-mer GG mismatched duplex structure-derived 3D simulated data were deconvoluted into a 2D NOE volume matrix as shown in the flow chart (Figure 2.1), and the deconvoluted 2D NOE volumes are compared to the experimental 2D NOESY spectrum. Eighty-seven experimental 2D peak volumes are checked against the corresponding 2D peak volumes from deconvoluted simulated 3D volumes (Figure 2.5). The two data sets are consistent,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

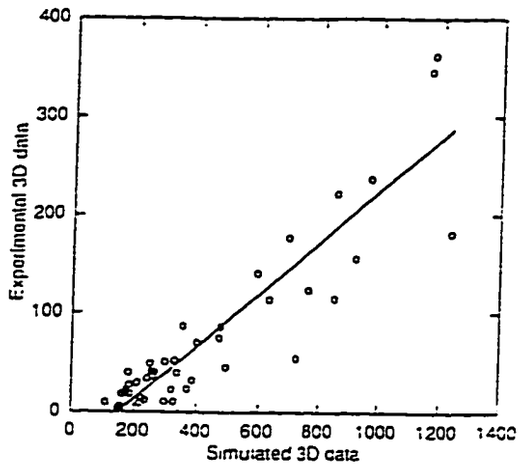


Figure 2.4 Plot of representative 3D experimental NOE-NOE volumes for the GG mismatch 12-mer d(CGGGAATTCGCG)$_2$ duplex versus the 3D simulated volumes. The data represent both 2D back transfer (V_{jkj}) and true 3D V_{ijk} volumes. The solid line is the best fit of the data from the linear regression with a correlation coefficient of 0.91.

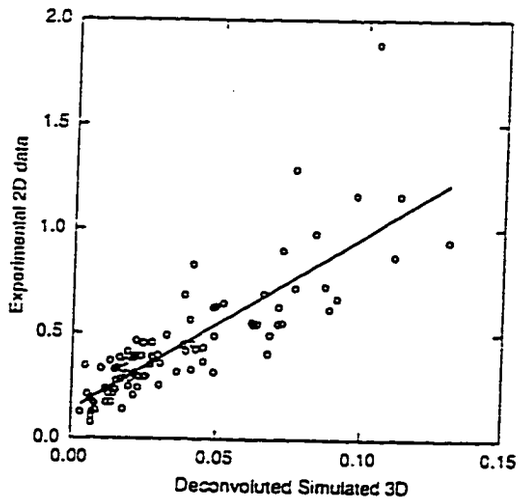


Figure 2.5 Representative experimental 2D NOE volumes for the 12-mer duplex (**PAT1**) plotted against the simulated and deconvoluted 3D volumes. The solid line is the best fit of the data from the linear regression with a correlation coefficient of 0.83.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with a correlation coefficient of 0.83. This good correlation coefficient suggests that the deconvolution algorithm is not only true in theory but also supported by the experiment result.

All these results suggest that the hybrid-hybrid matrix method of 3D NOE-NOE data analysis could provide a viable tool for refinement studies of large biomolecules. Other elements, for example the effect of thermal noise, are being tested. A major limitation in all 3D methods is the difficulty of accurately and efficiently integrating the many thousands of 3D volumes. Development of efficient software to automate the peak picking and integration of 3D data sets will be crucial.

Using this hybrid-hybrid matrix method of 3D NOE-NOE for a 35-mer three-way-junction DNA is under way, and results are not available yet due to the large number of peaks to assign and integrate.[41] But the pure simulation work done by Dr. Fan Zhu[42] did show that when compared with 2D MORASS (assume 2D has enough data to refine a structure) 3D MORASS method provides at least same or even better quality structure; the convergence is better for the refinement from different starting models. When a 2D NOESY can only provide scarce resolvable crosspeaks, the 3D MORASS becomes a better or even the only choice.

2.4 Conclusion

Simulations have shown that the highest accuracy and precision in structural refinement is achieved if both increased numbers and accuracy of the constrains are used.[39,43] The use of approximate methods has led investigators to apply liberal error

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

bars when using NMR-derived distances during constrained structural refinement. While it is true that many good-quality structures have been obtained from NMR,[44] especially with the development of 3D and 4D NMR.[45-48] It is obvious that using more and tighter constraints will achieve greater precision in the structures obtained by NMR.[39,49,50]

Thus, 3D NOE-NOE spectra hold the promise of giving more accurate structures given the vastly increased number of resolvable 3D NOE-NOE volumes.[17] As it has been shown[31] that approximation methods may not yield very accurate distances for the long mixing times which are required to achieve adequate magnetization transfer and signal-to-noise in these 3D spectra. Calculations based upon the complete relaxation rate matrix are well within current computational resources, even for larger macromolecules (<600 spins). Competing with the hybrid-hybrid matrix method described here are various direct gradient-NOE refinement methods.[17] Even with approximations to the gradient calculations, these scale as n^3, where the hybrid-hybrid matrix method is still basically an n^2 problem (diagonalization of the n×n volume matrix). It must be emphasized that NOE gradient refinements should utilize relaxation-matrix methods for calculating 3D volumes in order to achieve the highest accuracy for the refined structure.[31]

An additional advantage of this methodology is that software written for 2D NOESY data analysis does not need to be significantly modified since the 2D deconvoluted matrix replaces the experimental 2D NOESY volume matrix. The

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

analysis techniques described here are also general enough that they can be expanded to include even higher dimensional experiments, such as a 4D NOE-NOE-NOE spectrum.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

PART TWO: CHEMICAL NUCLEASE DESIGN, SYNTHESIS, ACTIVITY AND MECHANISTIC STUDY

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 3 INTRODUCTION TO ARTIFICIAL NUCLEASES

3.1 General Introduction

In recent years there has been a great deal of interest in designing artificial restriction nucleases,[51-85] which, like naturally occurring type II restriction enzymes, are able to cut DNA efficiently at unique sites. Nucleases, particularly the sequence-specific restriction endonucleases, have become indispensable tools in modern biochemistry and molecular biology. DNA recombination technology, which has revolutionized molecular biology and molecular genetics, makes significant use of enzyme-mediated sequence-specific cleavage. However, the usefulness of restriction enzymes is limited by their small recognition site sizes (4-8 base pairs).[86] Because of the small recognition site most restriction enzymes cleave with too high a frequency to allow analysis of large chromosomal DNAs, where separation into a manageable number of fragments is essential. Most restriction enzymes have recognition sites of 4-6 base pair (bp), corresponding to cleavage frequencies of once in every 256 bp for a 4 bp recognition site (4^4=256) to once in every 4096 bp for a 6 bp recognition site (4^6=4096). Because a single mammalian chromosome can contain more than 100 million base pairs, these natural restriction enzymes will generate far too many fragments to be analyzed by current available methods. A second problem is that although close to 3000

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

restriction enzymes have been discovered, the number of different sequences they recognize is only about 200 due to the occurrence of multiple enzymes (isoschizomers) that recognize the same sequences. Additional recognition sequences are likely to be added to the existing pool in the future, but there will remain many sequences for which no natural restriction endonuclease is available. Another limitation is that restriction enzymes do not in general cleave single-stranded DNA or RNA.

In addition to their potential as cloning tools, artificial nucleases may find applications as a) tools for cutting large DNA at unique selected sites for genomic mapping; b) chemotherapeutic reagents to treat certain genetic diseases; c) models for hydrolytic based DNA cleavage studies to provide insight into the mechanism(s) by which naturally occurring restriction enzymes function. In addition, redox based DNA cleavage study may provide insight into DNA damage processes caused by certain oxidants present in living cells, as well as provide additional insight into the interaction between small organic molecules and DNA molecules, necessary to design efficient DNA targeted drugs.

The current generation of artificial nucleases under development in our laboratory is composed of two parts: a sequence recognition part and a nucleotide cleaving part. The sequence recognition part provides the means to target unique sites on the substrate, while the cleaving agents carry out the substrate cleaving function. So far there have been many molecules designed and/or discovered which have DNA/RNA cleaving capability. These are categorized as shown in Figure 3.1. The sequence recognition part can be small molecules, oligonucleotides and proteins/peptides, while

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

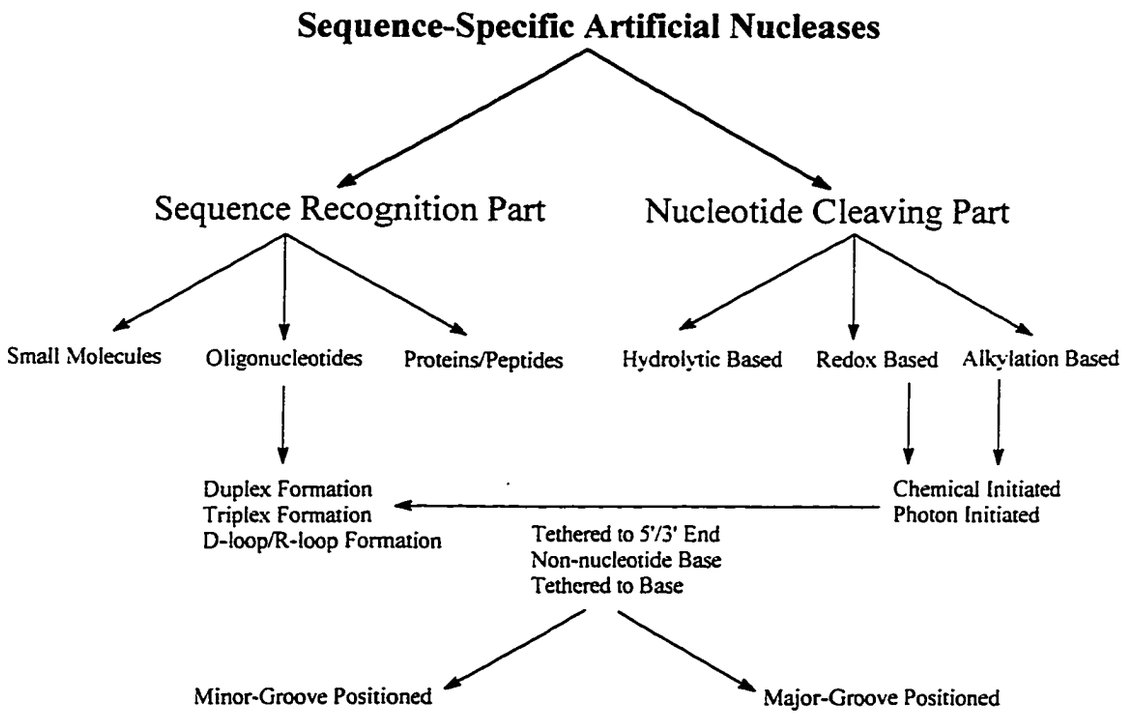


Figure 3.1 General scheme for artificial nuclease design.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the nucleotide cleaving part can be hydrolytic-based, redox-based or alkylation-based cleaving agents. Since our work has focused on oligonucleotides and transition-metal-complex conjugates, this type of artificial nucleases will be discussed in detail in the next section.

3.2 Artificial nucleases composed of oligonucleotide and transition-metal-complex conjugates

Oligonucleotides are particularly useful for recognizing defined sequences of single-stranded DNA, RNA and duplex DNA. The sugar-phosphate backbone provides a natural framework for building specific hydrogen-bonding interactions using the functional groups of the bases via Watson-Crick base pairing.[12,13] In addition, automated solid phase oligonucleotide synthesis has become sufficiently economical and simple,[87-90] that incorporating a modified phosphoramidite into an oligonucleotide is routine. Finally, many methods have been developed for nucleoside/nucleotide modification and derivatization, [84,91-101] so these methods can be utilized for covalently tethering nucleolytic agents to targeting oligonucleotides. According to the complementarity of carrying oligonucleotide of the cleaving agent, single-stranded DNA can be recognized by duplex formation[52,78,82,102,103] and double-stranded DNA can be recognized by triplex formation.[102,104-106]

Reports on the use of redox active metal complexes as chemical nucleases have been numerous over the past two decades.[52,53,55,57,58,66,68-70,72,74-76,78,107-116] Typical examples of transition metal mediated DNA redox cleaving agents are EDTA-

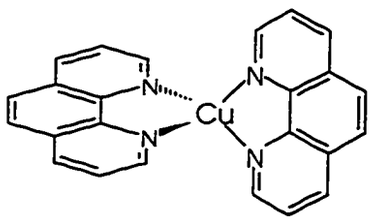
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Fe,[52,53,57,117,118] 1,10-phenanthrolene-Cu[55,58,72,78,107-109] and metalloporphyrin[66,104,119-122] (Figure 3.2). However, targeted scission using oligonucleotide and transition-metal-complex conjugates is relatively limited,[53,54,56-58,108,123-125] particularly for those conjugates in which transition-metal complexes are positioned *within* the grooves of DNA duplex.[51,57,84] Three representative examples of oligonucleotide transitional-metal-complex conjugates used as chemical nucleases are discussed below.

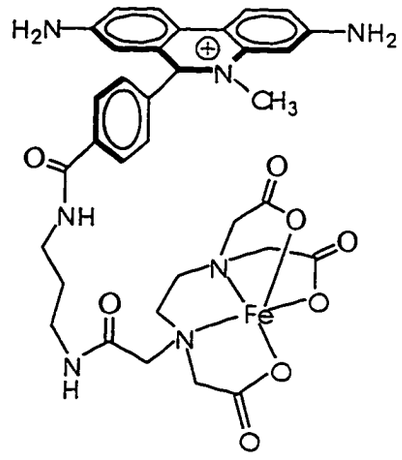
Oligonucleotide-EDTA conjugates. One of the most widely studied and used DNA cleavage reagents has been EDTA-Fe(II).[52,53,57,105,106,117,126-132] After a series of studies led by Dervan's group on tethering EDTA onto DNA minor-groove-binding molecules, like methidium[117,127] and penta-N-methylpyrrolecarboxamide,[129-132] several groups explored covalently linking EDTA onto a targeting oligonucleotide as potential sequence-specific artificial nuclease.

In 1984, Boutorin and colleagues[53] were first to successfully incorporate an EDTA ligand onto an oligonucleotide (Figure 3.3). This EDTA ligand was attached to the 5′-phosphate of an oligothymidylate octamer through a propylamine linker. The cleavage experiments using equimolar quantities of poly(dA) (200 nucleotide residues) and EDTA-conjugated oligonucleotide resulted in 50% cleavage of the substrate and 14% self-cleavage of the targeting EDTA-Fe(II)-poly(dT) following 5 hours incubation at 20°C. The self-cleavage is presumably due to the diffusible nature of the oxidative species produced during the reaction. Under the same reaction conditions the EDTA-Fe(II) complex cleaved poly(dA) considerably slower than the EDTA-Fe(II)-poly(dT). In a competition experiment, addition of poly(dT) to a reaction mixture of poly(dA) and

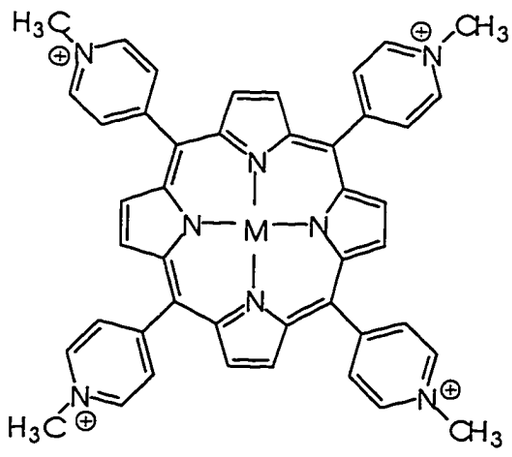
Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



Phenanthroline Copper(I)



Methidiumpropyl-EDTA-Iron(II)



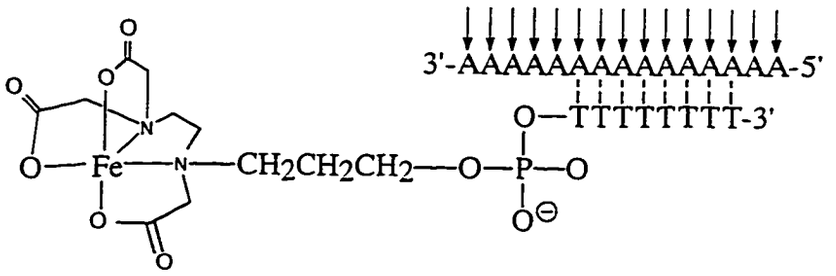
M = Fe 3+ , Mn3+ , Co3+ , Zn 2+ , Cu2+ , Ni 2+

meso-Tetrakis(4-N-methylpyridiniumyl)porphyrin

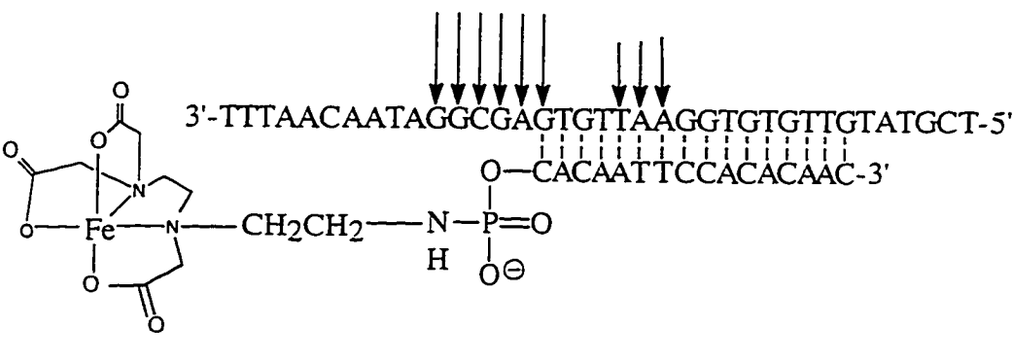
Figure 3.2 Typical transition metal ion complexes which function as redox-active DNA cleaving agents

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Boutorin et al:[53]



Chu et al:[56]



Dreyer et al:[57]

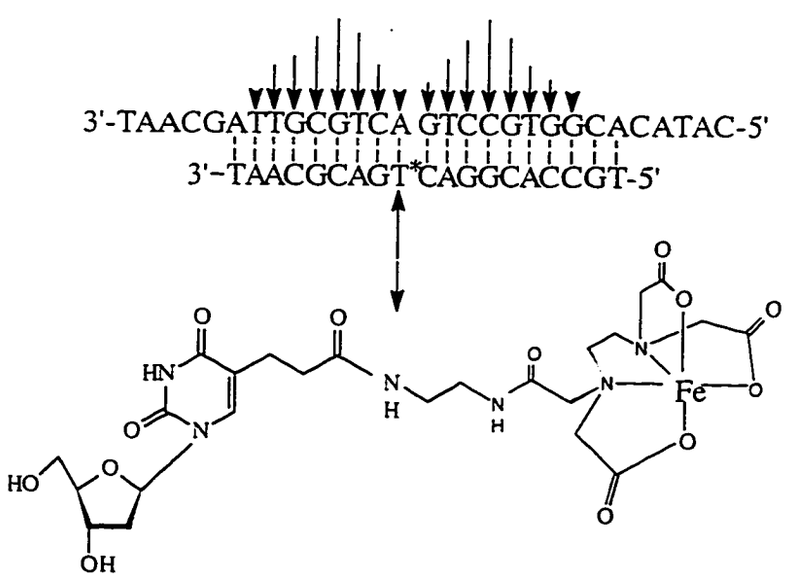


Figure 3.3 Oligonucleotide-EDTA conjugate as an artificial nuclease

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

EDTA-Fe(II)-poly(dT) led to decreased cleavage of the poly(dA). These results demonstrate that the cleavage of the poly(dA) substrate resulted from the binding of EDTA-Fe(II)-poly(dT) and not simply non-specific cleavage as in the case of poly(dT). In addition, incubation of the EDTA-Fe(II)-poly(dT) in the absence of target under the same reaction conditions resulted in 50-60% degradation of the self-conjugate according to the analysis by ion-exchange chromatography. Finally, the cleavage reaction proceeded only in the presence of the reducing agent dithiothreitol and dissolved oxygen.

In 1985, Orgel and colleagues[56] prepared a similar conjugate, attaching the EDTA ligand to the 5′-phosphorylated end of a 16-mer oligonucleotide via an ethylenediamine linker (Figure 3.3). The targeting 16-mer oligonucleotide was complementary to a 37-mer target at bases 7-22. Using a 7-70 fold excess of the oligonucleotide-EDTA conjugate the target was cleaved in 15% yield over a 9 nucleotide span. The substrate cleavage pattern showed that there were two loci flanking the linker position. In experiments in which the labeled 37-mer substrate reacted with Fe^{2+} alone or Fe^{2+} together with the 5′-P-16mer and free EDTA no cleavage was seen. These results suggest that the 15% cleavage of the 37-mer resulted from the binding of the 5′-EDTA-P-16mer. Although no quantitation of the self-damage was given, a second aliquot of oligonucleotide conjugate increased the cleavage yield by 20%, which implies that the targeting strand self-damage did occur during the cleavage reaction. The two loci cleavage pattern was explained by the authors to be a result of the ethylenediamine linker extending fully in either directions from the linker position.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Again, the nuclease activity of 5′-EDTA-Fe(II)-P-16mer required both reducing agent, dithiothreitol, and molecular oxygen.

In 1985, Dervan and Colleagues[57] positioned EDTA-Fe in the major groove of a DNA duplex by covalently linking it to C5 of a deoxyuridine in a 19-mer targeting oligonucleotide (Figure 3.3). This was the first attempt to attach a ligand with nuclease-like activity at a position other than the 3′- or 5′-end. Their target substrate was a 167-bp EcoRI/RsaI restriction fragment of plasmid pBR322. With a 500 fold excess of the EDTA-19mer conjugate to the target duplex they were able to cleave the target strand over a 16 nucleotide span around the site of modification. The cleavage pattern consisted of two loci of equal intensity, with maxima occurring four nucleotides to the 5′ side and three nucleotides to the 3′ side of the modification site. Their explanation was that the two cleavage loci arose from two conformations of the tethered EDTA-Fe(II) in the major groove of DNA duplex. One extended upstream and the other downstream, and the EDTA-Fe(II) in these two conformations was not quite equidistant from the modified base. Each cleavage locus extended over seven or eight nucleotides, suggesting that a diffusible species was responsible for cleavage. In a previous study with penta-N-methylpyrrolecarboxamide-EDTA•Fe(II), which binds in the *minor* groove of DNA duplex, the cleavage locus extended over four or five base pairs.[129-132] The greater number of cleavage sites produced by the major-groove-positioned EDTA-Fe(II) reflects the greater range available to a diffusible reactive species generated in the *major* groove of DNA duplex. The cleavage yield was not given, but the authors did comment on the inherent problem of self-cleavage. Experiments of the modified

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

oligonucleotide in the presence of dithiothreitol and absence of target strand showed that the conjugate was totally self-cleaved in less than one hour.

Using EDTA-Fe(II) equipped oligonucleotide Dervan's group was also able to cleave large double-stranded DNA by triplex formation.[105,106,133] In general, oligonucleotides 15-20 bases in length equipped with an EDTA-Fe moiety produced double-strand breaks with efficiencies ranging from 5 to 25%, and cleaving sites that spanned at least six nucleotides.

EDTA is a universally accepted ligand for iron-mediated redox reaction, but from the above results it is clear that EDTA may not be a good choice if one wants to position an iron atom tightly against a DNA molecule and obtain high cleavage specificity. We have chosen bidentate ligands bipyridine and phenanthroline instead of hexadentate EDTA for positioning an iron atom in a DNA duplex. In order to shift the mechanism from hydroxyl radical to iron-oxo it is necessary to associate the iron complex with DNA in such a way that the iron has direct access to a deoxyribosyl hydrogen atom. The nuclease activity study of iron mediated oligonucleotide-bipyridine conjugate is the subject of the next chapter.

Oligonucleotide-phenanthroline conjugates. Since the late 1970s when Sigman and colleagues discovered that phenanthroline-copper has nuclease activity,[134,135] several groups have tried to make sequence-specific nucleolytic oligonucleotide-phenanthroline conjugates.[51,55,58,72,78,84,107-109] The first example of a targeted scission reagent using phenanthroline-copper was synthesized by linking a 21-mer at its 5′-phosphorylated end to 5-glycylamido-1,10-phenanthroline through a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

carbodiimide condensation[54] (Figure 3.4). Two methods of assaying the cleavage reaction have been employed. In the first, the single-stranded DNA of the phage M13mp8 was the target, and cleaved sites were detected using a primer-extension assay. In the second, the substrate was an EcoRI fragment 3′ labeled in the non-template strand. After denaturing and re-annealing to a 53 fold concentration of the oligonucleotide-1,10-phenanthroline adduct, the substrates were cleaved over a continuous six nucleotide span at positions 20-25 in the presence of cupric ion and 3-mercaptopropionic acid. Hydroxyl radical-like species was proposed to be responsible for the cleavage reaction, but the centered 6 nucleotide cutting pattern was different than that caused by an oligonucleotide-EDTA adduct which generated two cleavage loci. The authors suggested that the hydroxyl radical-like species generated by phenanthroline-copper was less diffusible than that generated by EDTA-iron.

Similar work on oligonucleotide phenanthroline conjugates has been reported by others. Francois et al.[58] prepared oligonucleotides with phenanthroline on one end and acridine on the other, and showed that this oligonucleotide conjugate could form a triplex and cleave both strands of a duplex. Hélène et al.[136] attached phenanthroline to the 5′-end of an α-octathymidylate, and used this α-oligonucleotide conjugate to target a 27-mer DNA. The 27-mer DNA cleavage pattern showed a parallel orientation of the strands. In contrast, cleavage of $(rA)_{10}$ by the α-octathymidylate conjugate indicated an anti-parallel orientation of the strands. Chen et al.[55] have created an oligoribonucleotide with modified uridines on which phenanthroline was attached through a tether to C5. This oligonucleotide can cleave single-stranded DNA and nick and cleave double-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

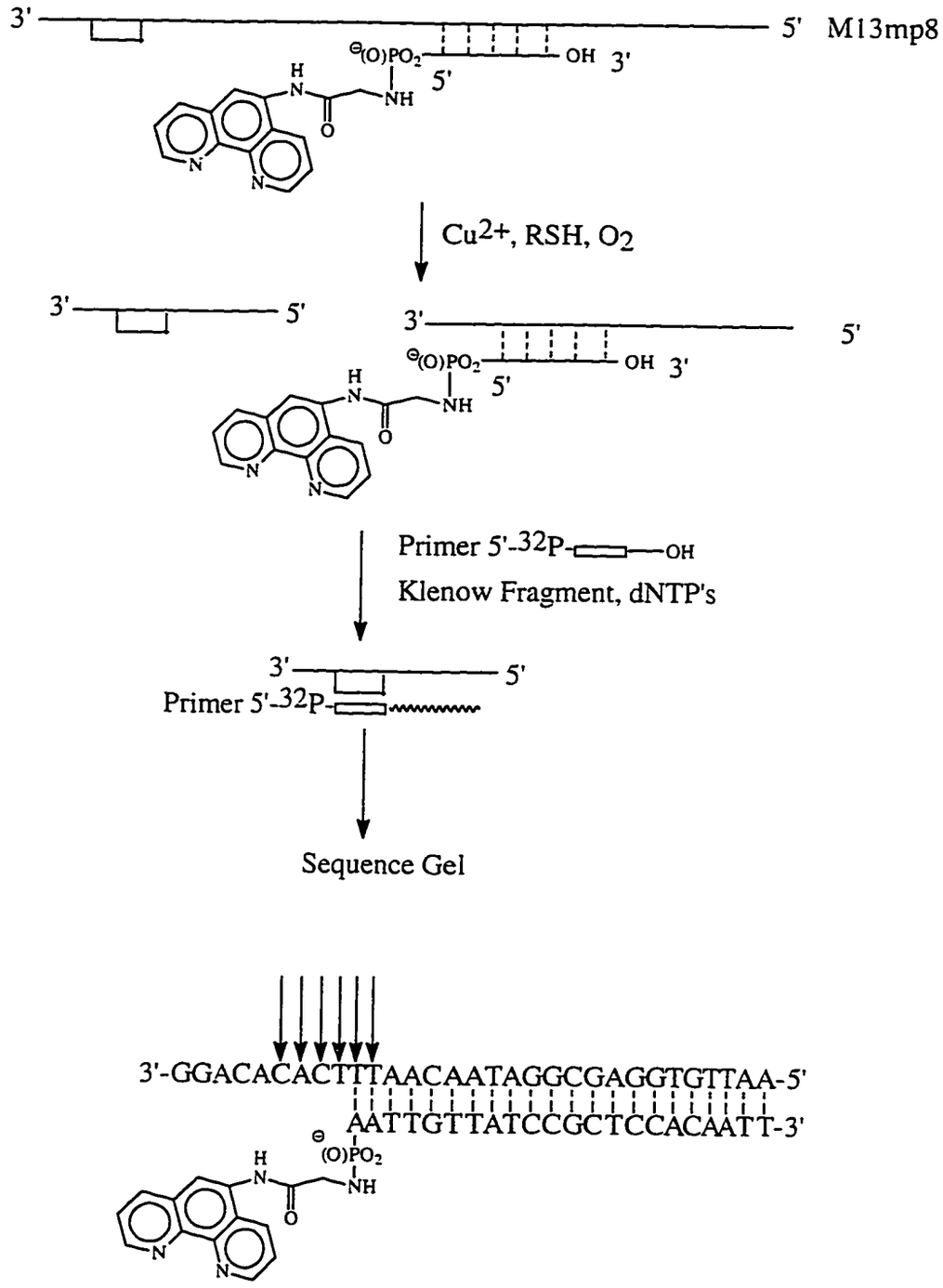


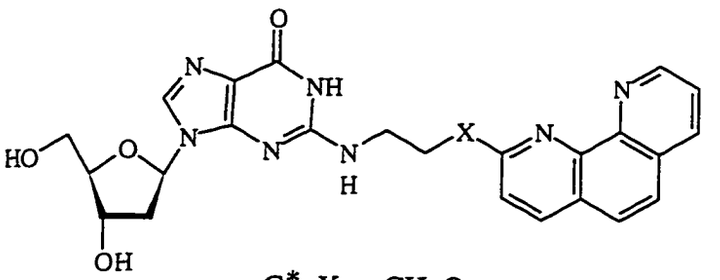
Figure 3.4 Oligonucleotide-phenanthroline conjugate as an artificial nuclease[54]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

stranded DNA through R-loop formation. In general, all these oligonucleotide-phenanthroline conjugates showed less than 20% cleavage efficiency and a six nucleotide cleavage span.

In contrast to the phenanthroline studies cited above, in which typically the ligand was appended at the 5′-end of an oligonucleotide, in our laboratory N. Gerry[51,84] constructed a modified oligonucleotide in which the metal-binding ligand was designed to specifically localize *within* the minor groove of a duplex nucleic acid. By linking phenanthroline to N-2 of deoxyguanosine and C2 of phenanthroline via -$CH_2CH_2CH_2O$- and -$CH_2CH_2NH(C)O$- tethers, the ligand is forced to project along the minor groove passing close to abstractable hydrogen atoms (H1′, H4′ and H5′/H5′′) of the target strand deoxyribosyls (Figure 3.5). When the phenanthroline ligand is thus covalently positioned in the minor groove, the oligonucleotide-phenanthroline conjugate does not show nuclease activity in the presence of copper ion and a reductant (3-mercaptopropionic acid or sodium ascorbate). The most probable cause of the inactivity of phenanthroline-copper in this system was the substitution of the phenanthroline at the 2-position. It has been demonstrated that substitution of phenanthroline at this position severely limits its redox activity in 2:1 phenanthroline copper complexes due to steric interaction that prevent formation of the square planar Cu^{2+} complex.[137] Although the modified oligonucleotides only employ one phenanthroline for chelation of the metal. the steric bulk of the ligands occupying the other two coordination sites must be great enough to prevent redox cycling. On the other hand, this oligonucleotide-phenanthroline conjugate showed relatively high nuclease activity in the presence of Fe(II) and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



G*: X = -CH_2O-

G^: X = -NHC(O)-

Oligonucleotide Substrate

3'-CTTTTTCTAGAGTCACCATAAACACTCGGTC-5'

Oligonucleotide-Phenanthroline Conjugates

O1 5'-TCTCAGTGGTATTTG-3'

O2 5'-TCTCAG*TGGTATTTG-3'

O3 5'-TCTCAGTGGTATTTG*T-3'

O4 5'-TCTCAGTG*GTATTTG-3'

O5 5'-TCTCAGTGG*TATTTG-3'

O6 5'-TCTCAG*TGG*TATTTG-3'

O7 5'-TCTCAG^TGGTATTTG-3'

O8 5'-TCTCAGTGGTATTTG^T-3'

O9 5'-TCTCAGTG^GTATTTG-3'

O10 5'-TCTCAGTGG^TATTTG-3'

O11 5'-TCTCAG^TGG^TATTTG-3'

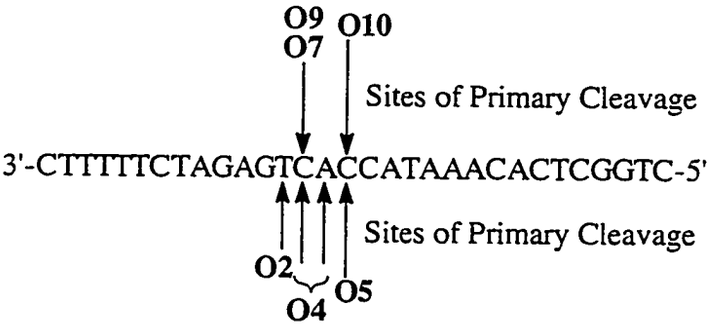


Figure 3.5 Minor-groove-positioned oligonucleotide-phenanthroline conjugate as an artificial nuclease[51,84]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dithiothreitol. When the phenanthroline was located near the 3′-end of the targeting oligonucleotide, a 31-mer substrate was cleaved over a 10 nucleotide span. However, sequences in which the phenanthroline was tethered near the middle of the targeting oligonucleotide led to highly specific cleavage concentrated at one or two sites. Cleavage efficiency up to 69% was achieved at a five fold excess of targeting strand. The self-cleavage of this conjugate was less than 10%, but other oxidative damage to the phenanthroline could not be excluded. This work illustrates the potential for further development of redox-based artificial nucleases. Particularly, it showed the possibility to achieve high DNA cleavage specificity by using an α,α′-diimine bidentate ligand to position an iron tightly against the opposing sequence.

Oligonucleotide-porphyrin conjugates. Cationic porphyrin transition-metal complexes are able to bind to DNA and induce cleavage either chemically or photochemically. Their conjugation with sequence-specific oligonucleotides has been exploited by a number of research groups.[64,66,102,119,138-140] Le Doan and colleagues[63,138,139] attached porphyrin to the 3′-end of oligothymidylates. After complexing with Fe(III), this oligonucleotide conjugate can cleave poly(dA) target in yields up to 37%. Addition of a second fresh aliquot of the conjugate after one hour reaction time resulted in total cleavage close to 60%. This conjugate nuclease activity required the presence of the reducing agent dithiothreitol. It was found that during the cleavage reaction this oligonucleotide conjugate was not self-cleaved. However, comparison of the conjugate UV spectra between reactions in the presence of substrate and in the absence of the substrate indicated conversion of the porphyrin ligand to an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

unidentified compound. In the presence of ascorbate or dithiothreitol as reducing agent, the conjugated porphyrin-Fe(III) complex not only resulted in substrate cleavage, but also cross-linking. Using 3-mercaptopropionic acid as reducing agent, they were able to achieve a cleavage specificity of 80% at one base in the seven nucleotide cleavage span. But the total cleavage yield plus cross-linking was only 1-2%.

In other similar reports, Frolova et al.[140] linked a porphyrin to the 3′- and 5′-end of a 16-mer oligonucleotide. With 100 fold excess of the conjugate they were able to obtain 20-60% cleavage of a complementary oligonucleotide in the presence of hydrogen peroxide. The cleavage efficiency was relatively high, but the cleavage covered several regions on the substrate. Recently Meunier and co-worker[104] used oligonucleotide-porphyrin•Mn(III) conjugate to cleave double-stranded DNA target present in the HIV-1 genome (Figure 3.6). They found the nature of the linker between the 5′-end of the targeting oligonucleotide and the tethered porphyrin to be a determining factor of the cleavage efficiency. If the linker is an aliphatic chain with different length, Figure 3.6, no efficient cleavage was observed. However, when the linker is a spermine molecule, cleavage yields as high as 80% were observed, in the presence of potassium monopersulfate, ($KHSO_5$). The high cleavage efficiency of spermine linked modified oligonucleotides was considered to be due to the enhanced stabilization of the triplex by the protonable spermine linker. But the cleavage still occurred over a seven nucleotide span on one strand and two cleavage loci on the other.

Overall, among the artificial nucleases reported so far only a few give cleavage yields higher than 50%, and they usually show a cutting pattern spanning several

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

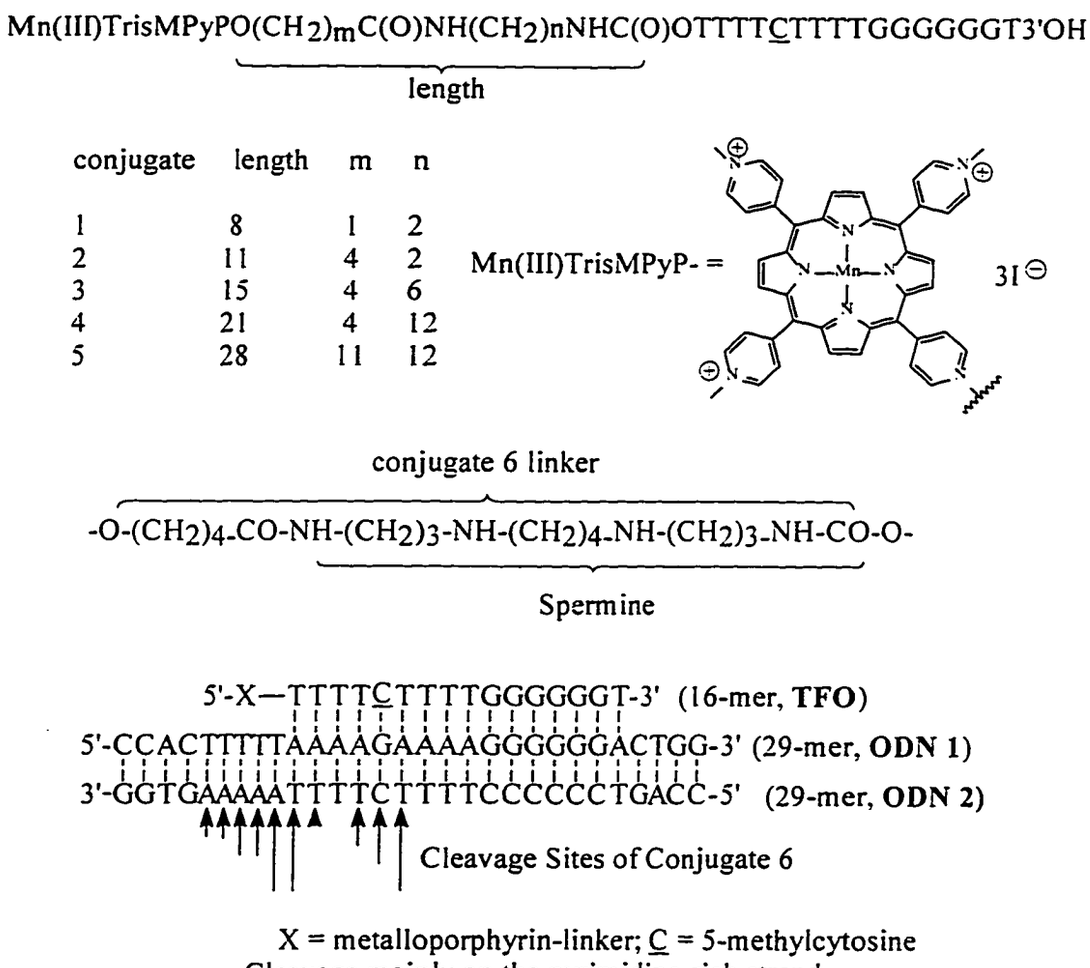


Figure 3.6 Oligonucleotide-porphyrin conjugate as an artificial nuclease[104]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

nucleotides and sometime two cleavage loci. Our laboratory have been working on improving artificial nuclease sequence-specificity and cleavage efficiency by covalently positioning an α,α′-diimine bidentate transition-metal ligand *within* the grooves of DNA duplex. The schemes of DNA fragmentation resulted from groove-positioned deoxyribosyl proton abstractions are shown in Figure 3.7, Figure 3.8 and Figure 3.9. The minor-groove-positioned phenanthroline-Fe(II) nuclease study has been reported[51,84] from this laboratory, and the major-groove-positioned bipyridine-Fe(II) nuclease study[141] will be fully discussed in the following chapters.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

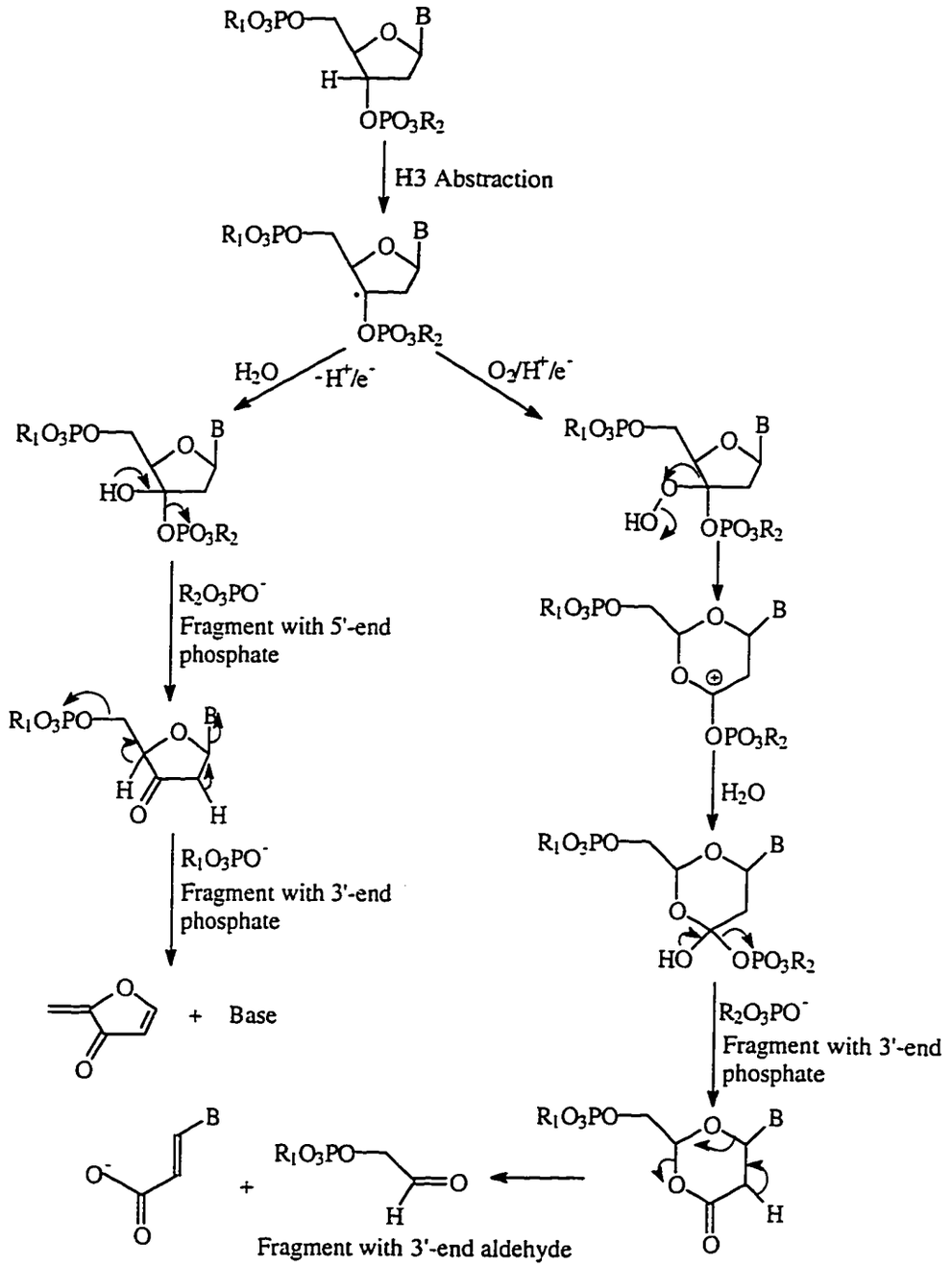


Figure 3.7 DNA fragmentation patterns via H3´ hydrogen abstraction in the major groove[142]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

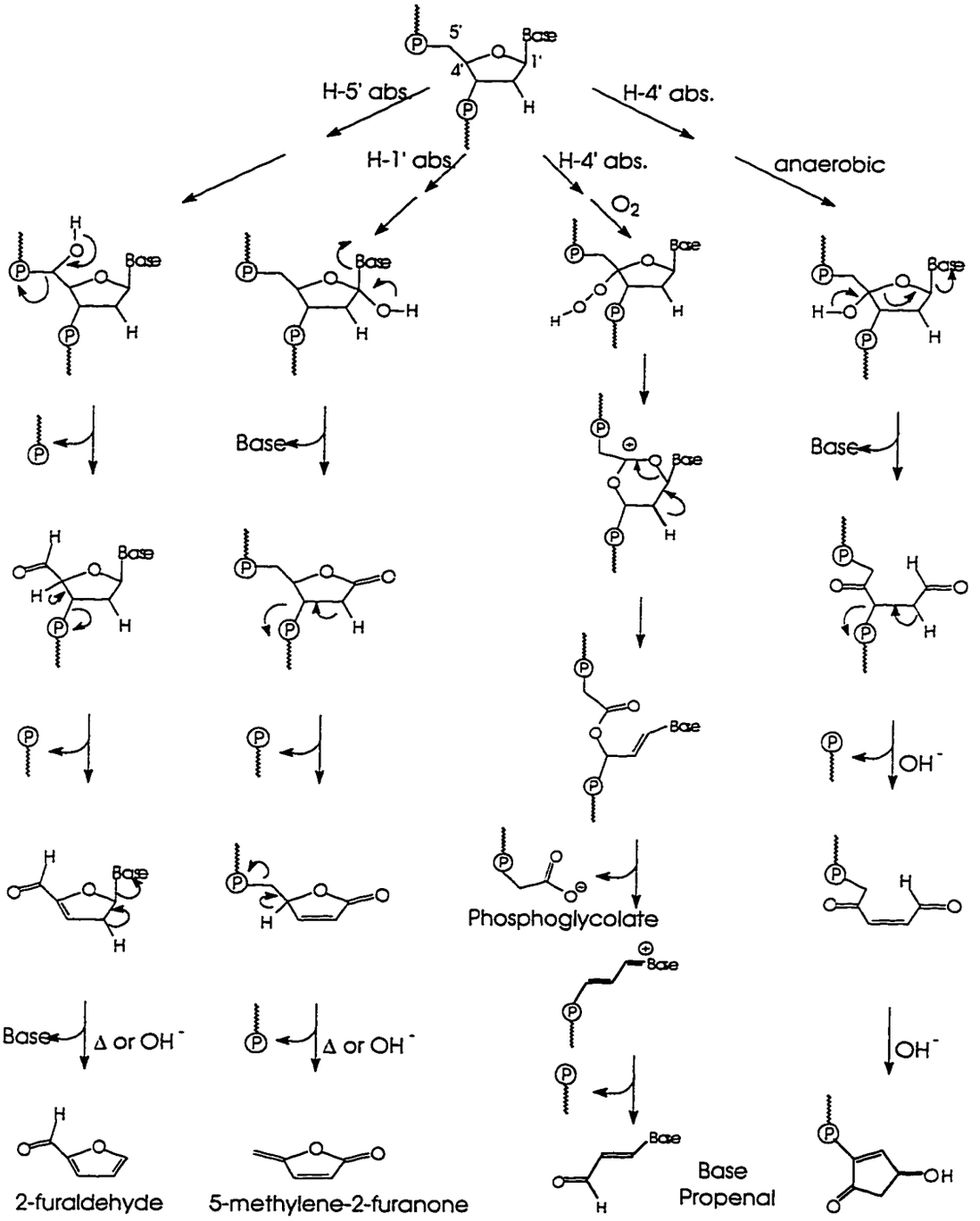


Figure 3.8 DNA fragmentation patterns via minor groove hydrogen abstraction in a DNA duplex[143-148]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

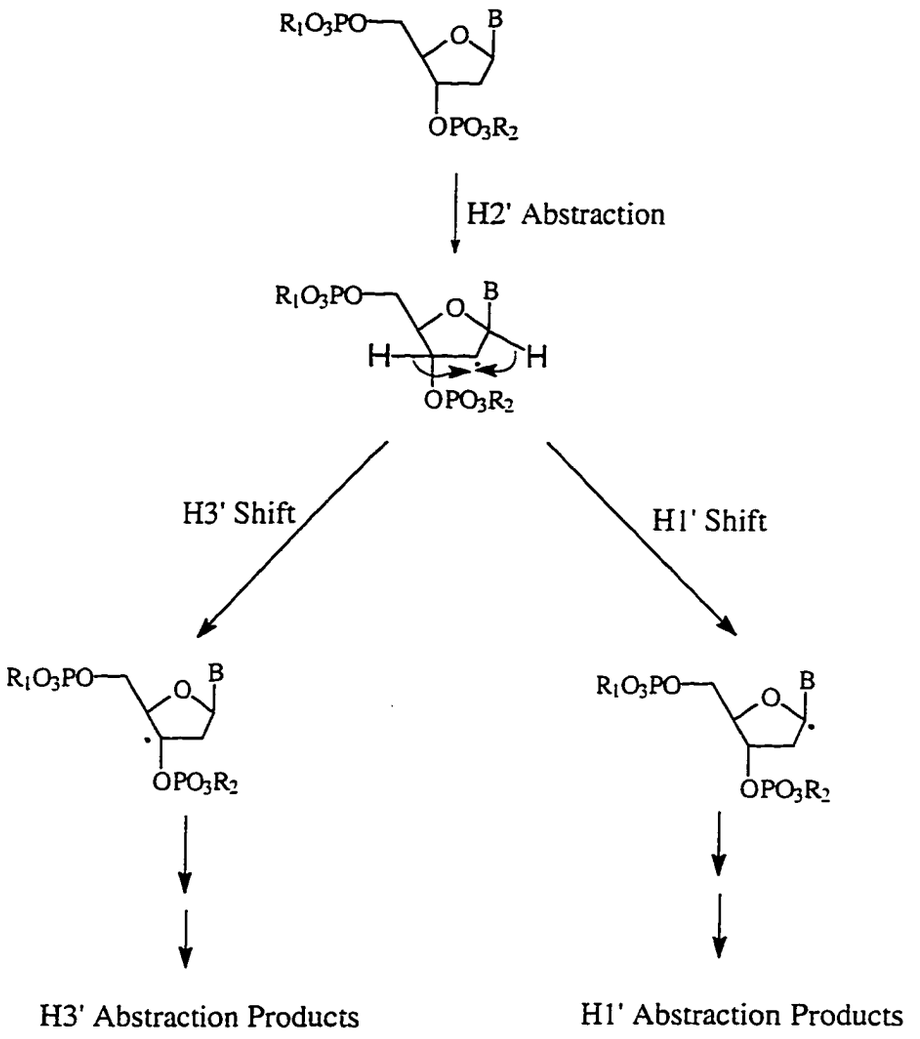


Figure 3.9 DNA fragmentation via H2´ hydrogen abstraction in a DNA duplex[149]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 4. CHEMICAL NUCLEASE DESIGN, SYNTHESIS, ACTIVITY AND MECHANISTIC STUDY

The ultimate goal of the artificial nuclease project is to develop a class of highly sequence-specific and efficient synthetic nucleases, which would have broad application in both chemotherapeutics and molecular biology. Many reports in literature have shown that redox active metal-mediated DNA cleavage by dioxygen has the potential to be used as artificial nucleases, but there are only a few reports[57,150] about exploiting the DNA cleavage with a covalently linked and major groove positioned metal complex.

There are four potential advantages in positioning a nuclease active linker in the major groove. 1) Pyrimidine nucleoside C5 modification chemistry is well developed.[93,94,96-101] 2) Major groove located linkers do not significantly decrease DNA duplex stability.[151] 3) There may be less limitation on ligand size and geometry in the wider major groove. For example, the phenanthrolene-copper does not show nuclease activity when it is covalently positioned in the minor groove.[84] This may be because the geometry interchange between Cu(II) (square planar) and Cu(I) (tetrahedron) is sterically locked in the minor groove during the copper complex mediating electron transfer.[84] When positioned in the major groove, the metal complex may not experience this geometry limitation. 4) The wider major groove provides better access for reagents

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in bulk solution to the metal-complex reaction center in the groove. With this hypothesis in mind, we set out to find an appropriate linker for tethering a transition metal complex to the C5 position of a deoxyuridine. Moieties appended on C5 will be positioned in the major groove when a DNA duplex is formed. Our initial specific objectives are: 1) find the appropriate linker through rational design and molecular modeling; 2) synthesize the C5 modified deoxyuridine, and incorporate it into an oligonucleotide; 3) assess the nuclease activity of the newly constructed chemical nuclease; 4) investigate the DNA cleavage mechanism if the new chemical nuclease does have nuclease activity; 5) explore applications for chemical nuclease in molecular biology.

4.1 Design of major-groove-positioned and redox-based chemical nucleases

Due to both the extensive pioneering work of the C5 derivatization of pyrimidine in this research group[93,94,96-101] and the geometrical location of the C5 in a DNA duplex, we have selected this site for tethering a transition metal ligand. 2'-Deoxyuridine was specifically chosen to carry a redox active complex. There are many known transition metal ligands: EDTA, phenanthrolene, porphyrin and bipyridine. EDTA[53,57,117,127] is the generally accepted ligand for iron mediated redox reactions. But it is not a good choice if one wants to position an iron atom tightly against a DNA substrate. Metalloporphyrins[64,66,104,115,119-121] have been quite extensively studied by other groups, but in general they do not have high DNA cleavage efficiency and specificity. Covalently linked phenanthrolene complexes have been investigated in this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

laboratory as part of another chemical nuclease project.[84] We have selected bipyridine as the metal ligand, because bipyridine is a good ligand for metals that can be redox mediated cleavers, and the synthesis for tethering bipyridine to a nucleoside is relatively reasonable. The bipyridine-metal complex linked to deoxyuridine C5 may be able to reach the phosphate oxygen(s) on the substrate DNA with an appropriate tether. If so, it may enhance this artificial nuclease's specificity because the active oxidative species generated *in situ* would be localized. The tether between the C5 of a 2′-deoxyuridine and the bipyridine was determined by modeling (Figure 4.1).

From a synthetic point of view, an amide linker is convenient for linking a bipyridine to the C5 side chain. In addition, a terminal amino group on the side chain may be useful for linking other ligands. To determine optimum tether length, possible tethers were modeled using Quanta/CHARmm (Molecular Simulations. Inc.) on a Silicon Graphics IRIS 4D/120 GTX computer workstation. The initial sequence chosen for modeling was d(AAAAAGATCT*CA), where the T* is the C5 derivatized 2′-deoxyuridine. The sequence chosen here is based on one of the long range goals of this work: cleaving RNA. This 12-mer is complementary to a sequence in rabbit β-globin mRNA that has been shown to be in a region relatively free of secondary structure. The eventual use of rabbit β-globin mRNA as a target will permit us to assay the specificity and efficiency of the bipyridine tethered chemical nuclease in cleaving mRNA under native conditions. However, DNA was initially used as target for the chemical nuclease study of bipyridine transition-metal complex. Consequently, the B-form DNA duplex

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

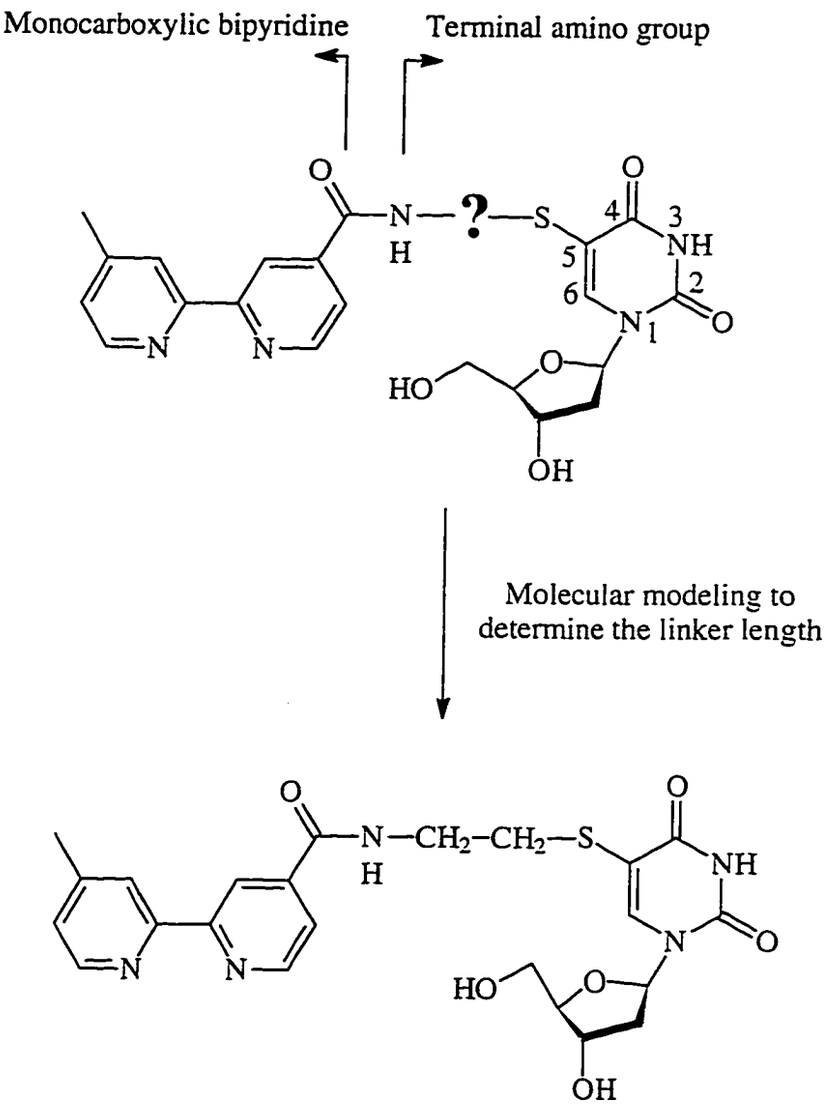


Figure 4.1 Design of a major groove positioned bipyridine tethered oligonucleotide

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

model was used for modeling because under native conditions double-stranded DNA forms the B-form duplex. The molecular modeling results provide only a rough guide, because the modeling was done in the absence of bound metal. Since bipyridine is a bidentate ligand, a metal ion has additional coordination sites (depending on the nature of the metal ion) open. Under the DNA cleavage conditions, the potential ligands which could occupy these sites are: water; anions present in solution; reductant(s), if required by the cleavage reaction; phosphate(s) on the substrate DNA and the heteroatoms on the nucleic bases. Tethers with different lengths were incorporated into the C5 side chain of a 2′-deoxyuridine for modeling. Energy minimization was accomplished by using the Adopted-Basis Newton-Raphson routine. The energy minimized structures were generated, and molecular dynamics simulations were run to generate the molecular motional trajectories (collective view of molecular structures on time scale). Based on the structural trajectories, the -$SCH_2CH_2NHC(O)$- tether was chosen for linking the nucleic acid base and the bipyridine ligand. Two views of a DNA duplex with the modified 2′-deoxyuridine are shown in Figures 4.2 and 4.3.

The modeling results indicated that a 2-carboxamidoethylthio linker could position the bipyridine ligand in close proximity to a phosphate and deoxyribosyl H3′ proton on the complementary strand. The structural dynamic simulation shows that the tethered bipyridine moves around a three nucleotide span of the substrate. This major groove positioned bipyridine has a certain degree of rotational freedom in a DNA duplex. If the bipyridine-metal complex does bind to a phosphate on the complementary

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

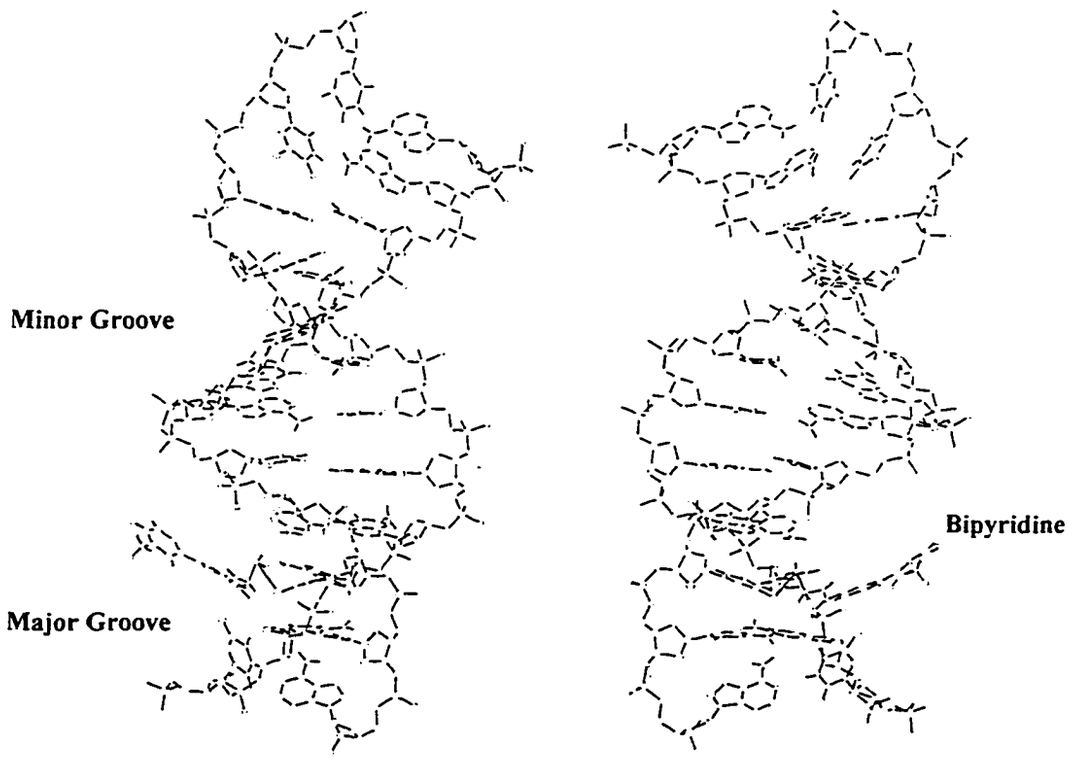


Figure 4.2 Bipyridine covalently positioned in the major groove of a DNA duplex.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

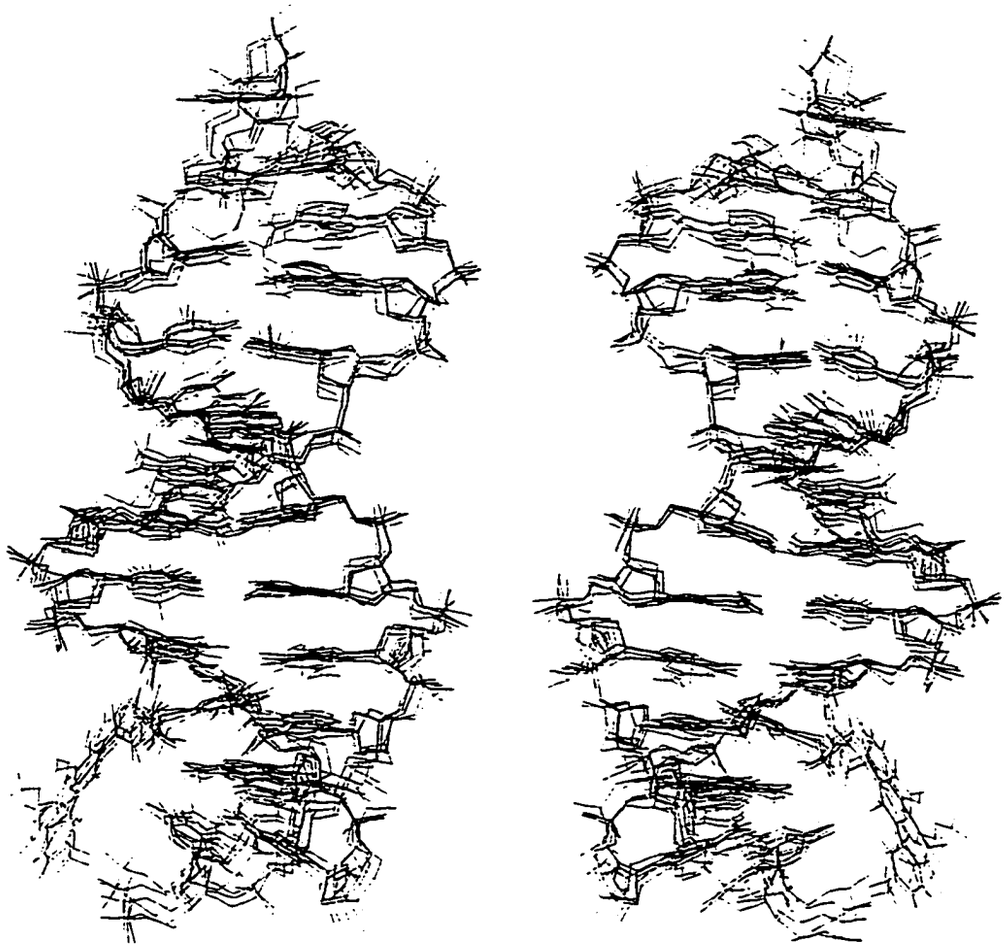


Figure 4.3 Stereo views of the overlay of a few dynamic motion frames of the bipyridine tethered oligonucleotide duplex. (left: major-groove view; right: minor-groove view)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

strand, it would be tightly positioned against the substrate and would possibly yield better specificity.

4.2 Synthesis

Solid-phase oligonucleotide synthesis is used to assemble target and targeting oligonucleotides.[87] Consequently, the bipyridine modified nucleoside needed to be transformed to a phosphoramidite for incorporation into an oligonucleotide.

4.2.1 5-[2-(4′-methyl-2,2′-dipyrid-4-yl-carboxamido)ethylthio]-2′-deoxyuridine phosphoramidite

The scheme for synthesis of **7** is outlined in Figure 4.4. 2′-Deoxyuridine was first functionalized at C5 by reaction with mercury acetate to form 5-(chloromercuri)-2′-deoxyuridine.[91,92] 5-(Chloromercuri)-2′-deoxyuridine and 5-iodo-2′-deoxyuridine are two of the most versatile intermediates for deoxyuridine modification at C5.[93,94,96-101] Compound **2** was coupled with N,N′-bis(trifluoroacetyl)cystamine (**10**) with one equivalent of lithium palladium chloride to yield nucleoside **3**. The trifluoroacetyl group was removed by ammonium treatment after the coupling reaction to yield nucleoside **4** which contains a terminal amino group which can be linked to variety of moieties. Bipyridine was attached to the free amino via an amide bond. 4,4′-Dimethyl-2,2′-dipyridine was subjected to controlled oxidation[152] to oxidize one of the two methyl groups to the monocarboxylic acid derivative, **8**, in 50% yield. The monocarboxylic acid derivative was activated by N-hydroxylsuccinimide before it was coupled

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

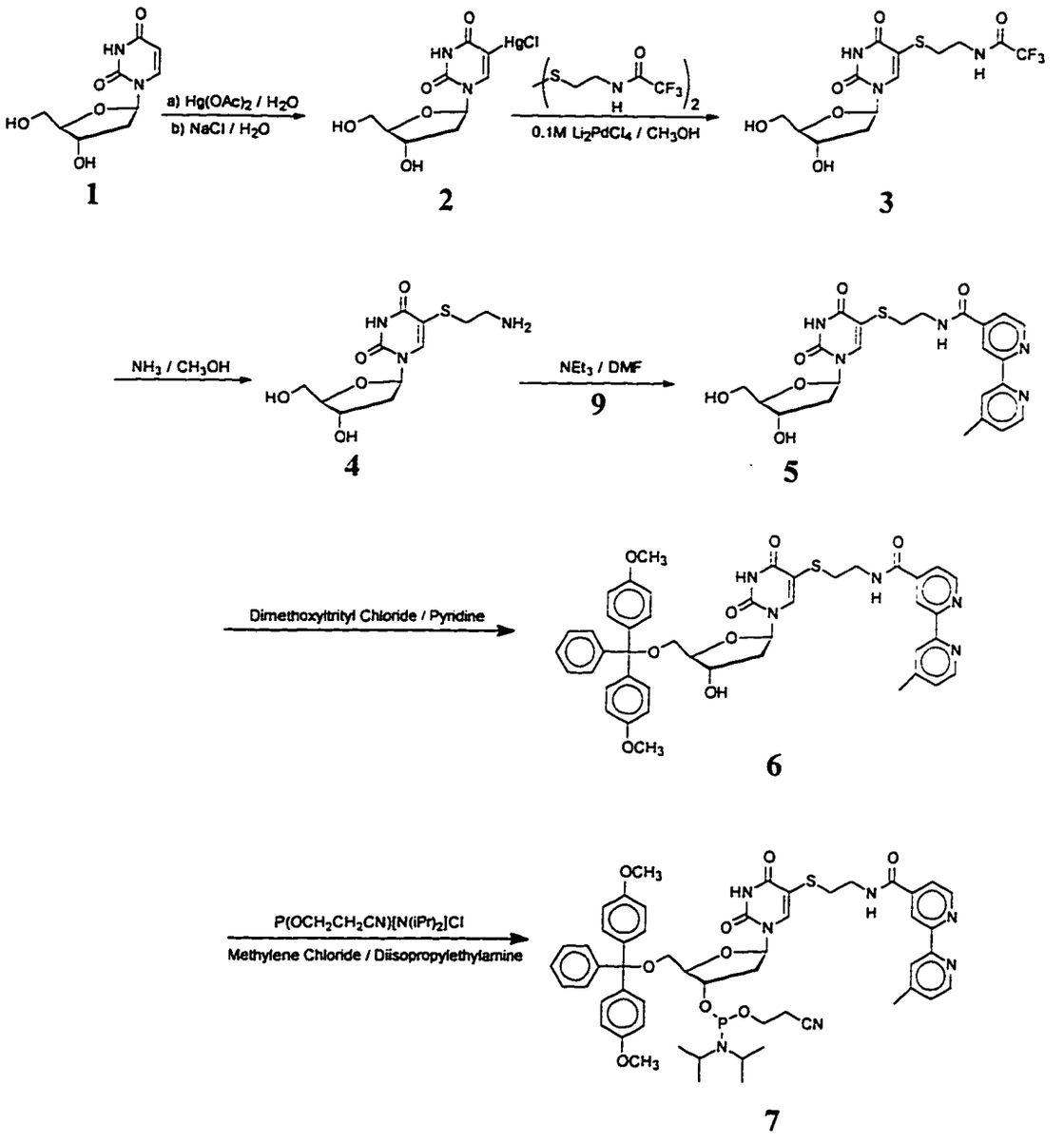


Figure 4.4 Synthesis of C5 bipyridine tethered 2′-deoxyuridine phosphoramidite (7)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

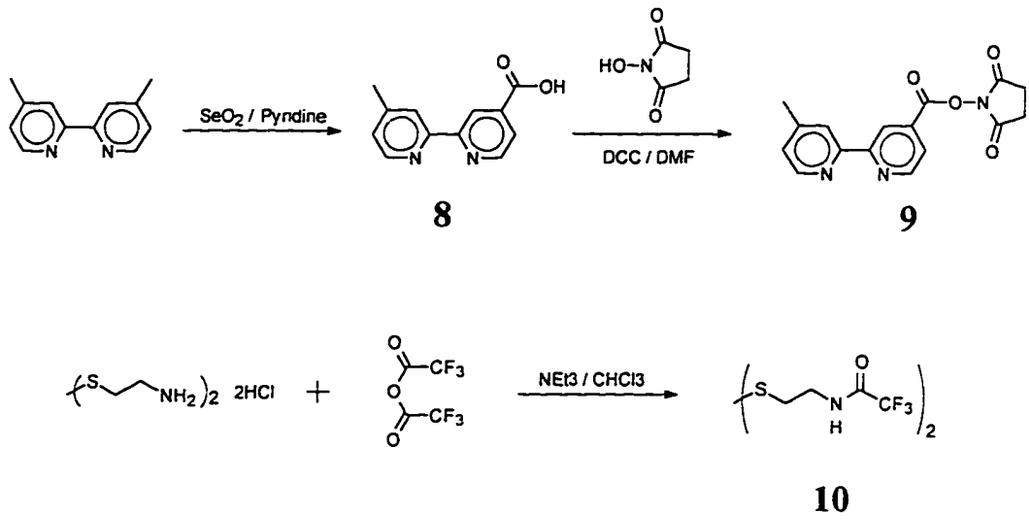


Figure 4.4 (continued)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to the terminal amino of the deoxyuridine analogue, **4**, to yield compound **5**. Compound **5** was converted into a protected phosphoramidite by sequential reaction with dimethoxytrityl chloride and chloro-β-cyanoethoxyl-N,N-diisopropylaminophosphine. The bipyridine tether did not interfere with the coupling and deprotection processes during the solid phase oligonucleotide synthesis. The incorporation of modified 2′-deoxyuridine was confirmed by oligonucleotide enzymatic digestion and HPLC analysis.

4.2.2 3′-Deuterio-2′-deoxyadenosine phosphoramidite

A ligand linked by a short tether to C5 occupies the major groove of a DNA duplex and appears unlikely to be capable of reaching the minor groove protons on the deoxyribose ring: H1′, H4′, and H5′. On the other hand, deoxyribosyl protons that are located in the major groove, H3′ and β-H2′, are easily accessible. Theoretically, H3′ abstraction is energetically more favorable than H2′ abstraction.[153] Experimentally, Barton's group has observed products which presumably come from H3′ abstraction.[150] They have described a chiral rhodium(III) complex which binds in the major groove, and upon photochemical initiation cleaves the bound DNA. We have positioned the bipyridine-Fe(II) in the major groove and observed an 80% substrate DNA cleavage yield. It is very likely that this substrate scission is also initiated via the H3′ abstraction in the major groove. Direct evidence for H3′ abstraction would possibly be obtained by measuring an isotope effect at this site. Consequently we labeled C3′ of a 2′-deoxyadenosine at the target site with ^{2}H.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The synthesis of 3′-deuterio-2′-deoxyadenosine has not been reported. However, 3′-deuterioadenosine has been synthesized[154] and a short route for converting adenosine into 2′-deoxyadenosine is known.[155] Our ^{2}H3′ labeling scheme is shown in Figure 4.5.

The synthesis of 3′-deuterioadenosine (**15**) proceeds by way of 3′-ketoadenosine (**12**) as an intermediate, consequently one might presume that 2′-deoxy-3′-deuterioadenosine may be prepared by way of 2′-deoxy-3′-ketoadenosine. In fact, this is not possible because of the instability of the 2′-deoxy-3′-ketoadenosine. Adenosine may be selectively protected by dimethylisobutylsilyl chloride at the 2′ and 5′ positions in pyridine solution under N_2 atmosphere. With a 48 hours reaction time, the desired compound **11** was obtained in 50% yield. The other three byproducts (all three hydroxyls protected, 3′ and 5′ hydroxyls protected and only 5′ hydroxyl protected) were obtained in much lower yield. No base amino silylation was observed in this reaction. The 3′ hydroxyl in the protected adenosine (**11**) was oxidized to a keto group by chromium trichloride/ethyl acetate/pyridine (1:3:6). The oxidized product (**12**) was extracted into diethyl ether. The isolated yield was 65% for this oxidation step. Other oxidizing agents have also been reported[154] for this particular oxidation step, but we found that chromium trichloride procedure was relatively convenient. The ^{2}H3′ atom was introduced by reducing the C3′ keto to a hydroxyl with a deuteride reducing agent. The goal in this step was to control the stereochemistry at C3′ in order to obtain ribose rather than xylose. Robins' group found[154] that this could be achieved quantitatively if the C5′ hydroxyl is freed before reducing the C3′ keto with sodium borodeuteride,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

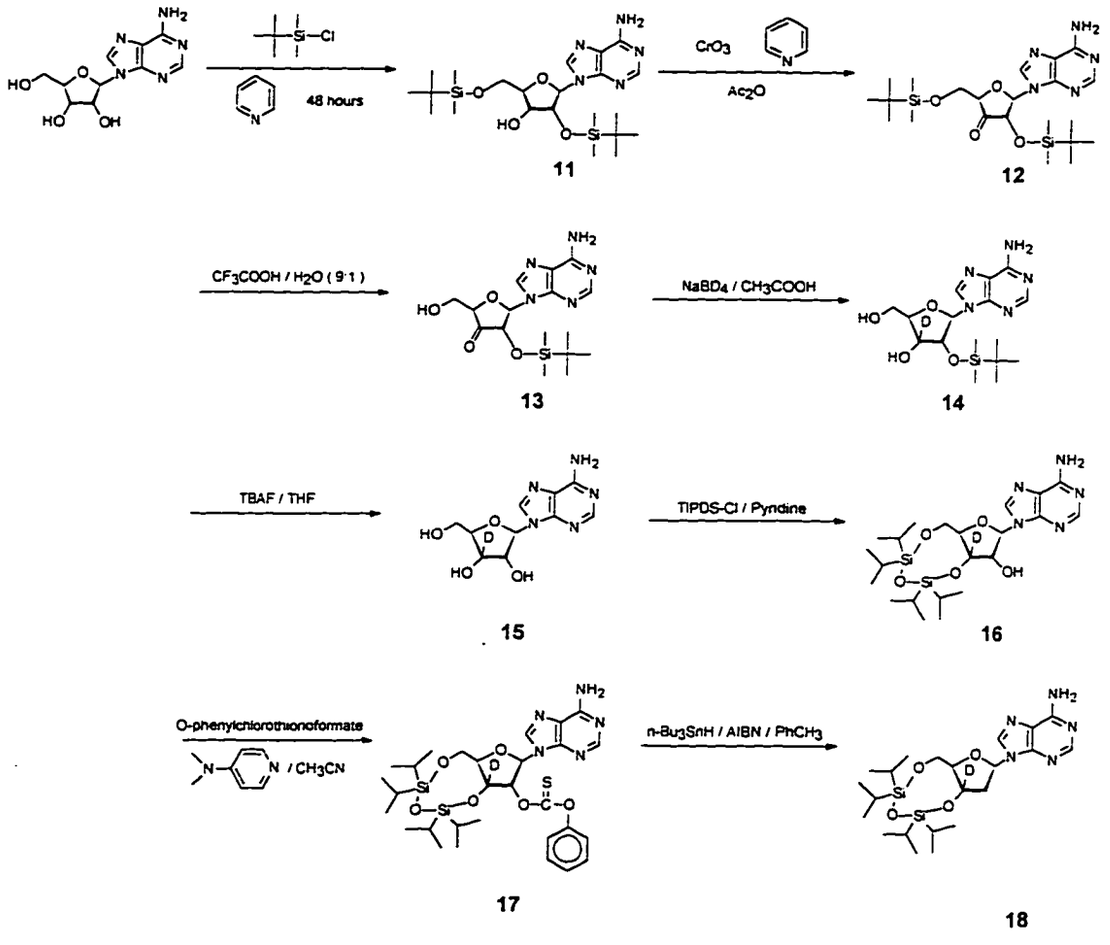


Figure 4.5 Synthetic scheme for D3′ labeling of 2′-deoxyadenosine phosphoramidite (**23**)[154-159]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

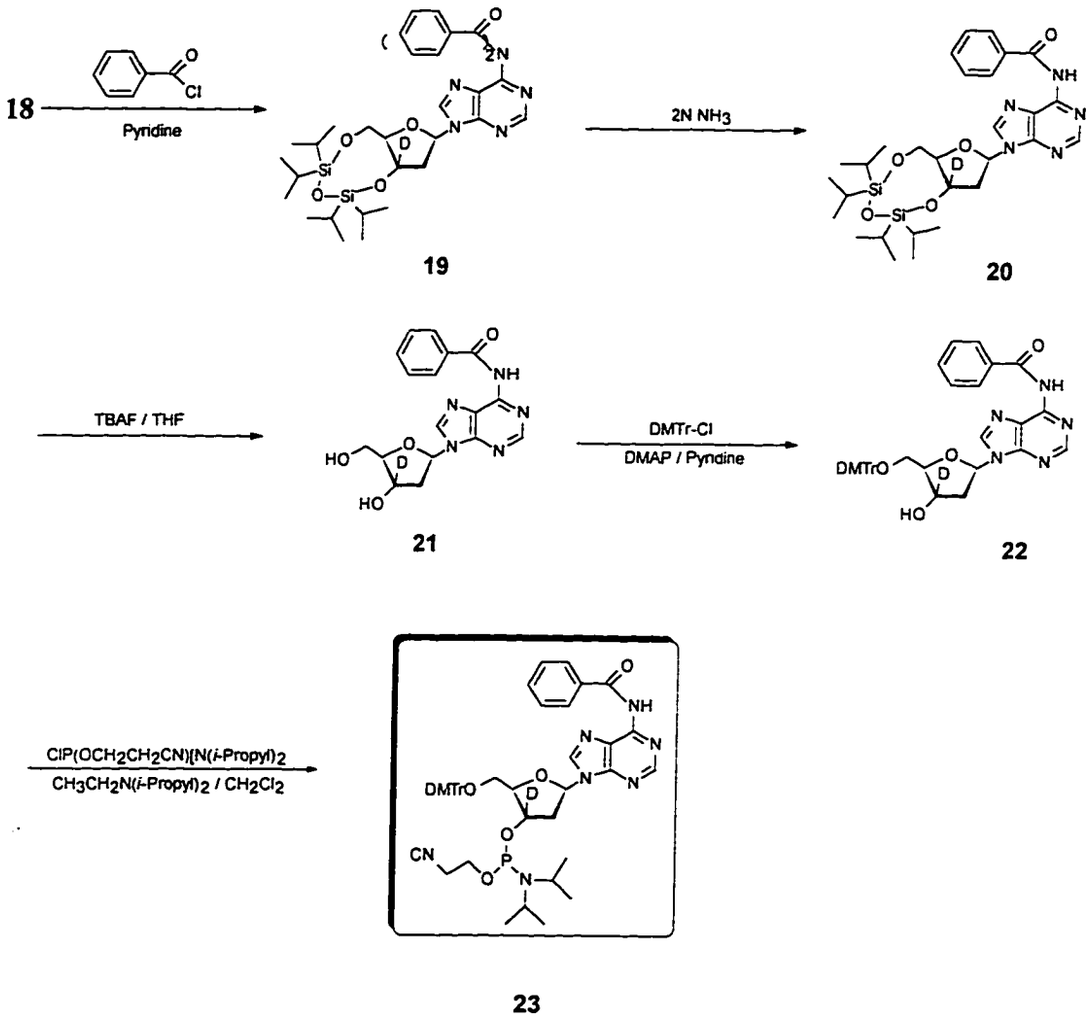


Figure 4.5 (continued)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

otherwise the xylose sugar is the major product. This stereochemistry control likely resulted from the coordination between the free C5′ hydroxyl and boron during the reduction, in which the incoming deuteride attacks the C3′ keto from the C5′ side of the sugar ring. The C2′ silyl protecting group can be removed by treatment with tetrabutylammonium fluoride in tetrahydrofuran, and the resulting product, 3′-deuterioadenosine (**15**), was checked against adenosine by NMR. The ^{1}H NMR spectrum of 3′-deuterioadenosine showed that in DMSO, H2′ gave a triplet with coupling constant 6Hz, and H4′ a triplet with coupling constant 3Hz. The OH3′ proton appeared as a singlet while the H3′ proton was absent. In contrast, the ^{1}H NMR spectrum of adenosine in DMSO showed that H2′ was a quartet with coupling constant 6Hz, H4′ a quartet with coupling constant 3Hz, OH3′ a doublet with coupling constant 4.5Hz, and H3′ a multiplet. This is consistent with $^{2}H3'$ incorporation. The next step was to remove the 2′-hydroxyl group from the deuterated adenosine.

In order to selectively remove the 2′-hydroxyl group, the 3′ and 5′-hydroxyl groups have to be selectively protected. This selective protection was carried out by using tetraisopropyldisiloxyl-1,3-dichloride, which was developed by Markiewicz.[159] The 2′-hydroxyl was transformed to a O-phenylthionoformate ester with O-phenylchlorothionoformate to yield purified compound **17** in 85% yield. Reduction with tributyltin hydride/AIBN/toluene yielded nucleoside **18** in quantitative yield.

The 6-amino of adenine was protected by benzoylation and the silyl protecting group removed by treatment with TBAF in THF. The product **21** was converted into phosphoramidite **23** by the reactions: dimethoxytritylation on the 5′-hydroxyl and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

phosphitylation on the 3′-hydroxyl. This final product was checked by high resolution FAB-MS and ^{31}P NMR, and found to sufficiently pure for incorporation into an oligonucleotide.

4.3 Synthesis of bipyridine-tethered 2′-deoxyuridine and 3′-deuterio-2′deoxyadenosine modified oligonucleotides

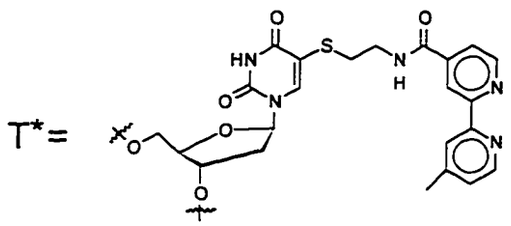
After the syntheses of both 5-[2-(4′-methyl-2,2′-dipyrid-4-yl-carboxamido)ethylthio]-5′-*O*-dimethoxytrityl-3′-*O*-β-cyanoethyl-N,N-diisopropylphosphoramidyl-2′-deoxyuridine (**7**) and 6-benzoylamino-9-(3-deuterio-3-β-cyanoethoxyl-N,N-diisopropylaminophosphoramidyl-5′-dimethoxyltrityl-β-D-ribofuranosyl)-purine (**23**), the phosphoramidites were successfully incorporated into oligonucleotides of different sequences by using the solid-phase phosphite-triester oligonucleotide synthesis protocol.[87] The stepwise coupling efficiency of the modified oligonucleotides averaged at 97%. These modified oligonucleotides were checked on a MALDI-TOF mass spectrometer, and their masses were consistent with theoretical values (Table 4.1). Further confirmation of the incorporation of T^* into the designed sequences was performed by G. Wang. The T^* containing oligonucleotides were digested by snake venom phosphodiesterase and bacterial alkaline phosphatase, and then the digest analyzed on reverse phase HPLC. The results of the analysis showed that the five HPLC peaks were consistent with the external T^* standard and the other four standard nucleosides. In addition, the peak integration was also consistent with the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

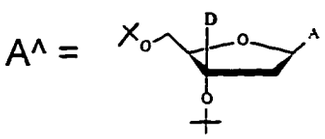
Table 4.1 MALDI-TOF results of bipyridine and 2′-deoxy-3′-deuterioadenosine modified oligonucleotides[a, b, c]

DNA sequences	Mass (found)	Mass (calc.)	Mass (diff.)	% error
5′-CGATAAGCTTGATAT*CG-3′ (**JC33**)	5465.08	5466.51	-1.43	-0.03
5′-CCGCGGTGGAGCT*CC-3′ (**JC31**)	4829.99	4826.92	+3.07	+0.06
5′-GATCT*CAGT-3′ (**A9**)	2977.43	2970.81	+6.62	+0.22
5′-AAGATCT*CAGTGG-3′ (**A13**)	4252.77	4255.69	-2.92	-0.07
5′-TCAGT*GGTA-3′ (**B9**)	3010.43	3010.85	-0.42	-0.01
5′-TCTCAGT*GGTATT-3′ (**B13**)	4208.11	4211.92	-3.81	-0.09
5′-CTGGCTCACAAATACCAC TGA^GATCTTTTTC-3′ (D3′-A21 **O1**)	9409.03	9405.30	3.73	+0.04

a:



b:



c: The error in mass measure is typically around 0.2%.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

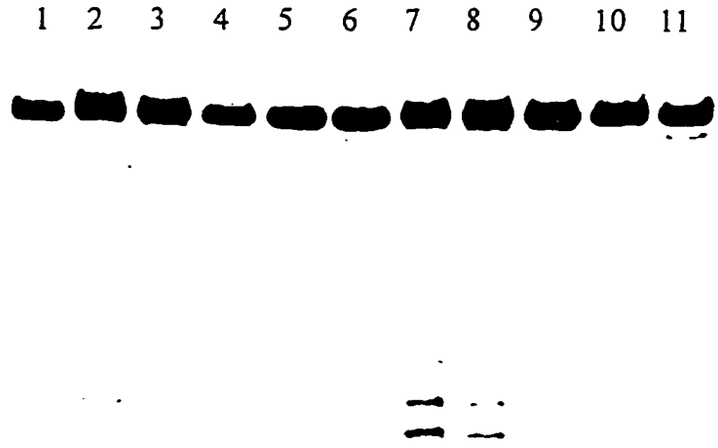
designed oligonucleotide sequences. These results indicated that the bipyridine tethered phosphoramidite can be successfully incorporated into an oligonucleotide and can survive during the deprotection post-synthesis workup. During all oligonucleotide manipulations, care was taken to avoid nuclease contamination and introduction of metal ions. Figure 4.11 shows the $T^{\bullet}$ oligonucleotides and their corresponding substrates. Oligonucleotide **O2** is complementary to a sequence in rabbit β-globin mRNA, which was chosen because it occurs in a region predicted to be relatively free of secondary structure. The oligonucleotide **JC31** is complementary to a sequence in the gene marker of pBluescript™ plasmid. This plasmid was to be used in exploring molecular biology applications for our chemical nucleases.

4.4 Nuclease activity assay and mechanistic study

4.4.1 Iron mediated single-stranded DNA cleavage

It has been shown that EDTA-Fe(II) nuclease activity requires the presence of a reductant as a co-reactant.[84,127] We initiated the nuclease activity assay of **O2** with Fe(II)/ascorbate and Fe(II)/dithiothreitol (DTT) at various reductant concentrations. However, we found that **O2** had virtually no nuclease activity in these Fe(II)/reductant systems (Figure 4.6). This result suggested that the bipyridine modified oligonucleotide could not cleave DNA by the mechanism proposed for EDTA-Fe. We then turned to the classic Fenton reaction, in which a mixture of ferrous ion (or ferrous complex) and hydrogen peroxide can produce a strong oxidative species. We found that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



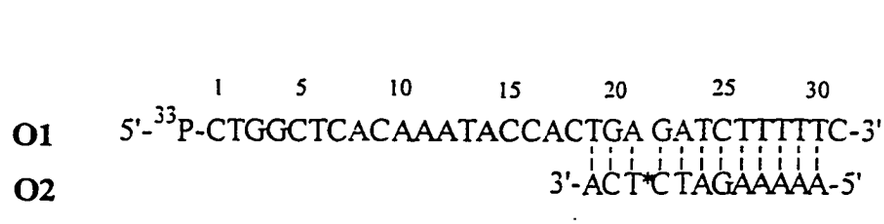


Figure 4.6 Oligonucleotide **O2** mediated cleavage of substrate **O1** in the presence of reductants. Lane 1: control, labeled **O1** only; lane 2-6: in the presence of DTT at concentrations of 20mM, 4mM, 1mM, 0.1mM and 0.01mM, respectively; lane 7-11: in the presence of ascorbate at concentrations of 20mM, 4mM, 1mM, 0.1mM and 0.01mM, respectively. Other conditions are: 0.9 pmol **O1**, 4.5 pmol **O2**, 50mM Tris/50mM NaCl buffer with pH 7.4, [Fe(II)] = 0.01mM, and 3 hours reaction time at room temperature.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in the presence of ferrous ion and hydrogen peroxide the bipyridine modified oligonucleotide **O2** had unique nuclease activity. With a five fold excess of the chemical nuclease, **O2** cleaved 80% of the substrate DNA localized over a three nucleotide span (Figure 4.7, Figure 4.8).

Chemical nuclease cutting efficiencies reported in the literature are often below 50%, and in many cases appear to require large excess of the chemical nuclease. The low efficiency of redox based chemical nucleases[57,84] was attributed largely to self-cleavage of the targeting strand during cleavage reaction. The high cleavage efficiency of **O2** suggests that this bipyridine modified oligonucleotide may not have a self-fragmentation problem. The self-damage of **O2** was assessed by labeling **O2** at the 5′-end and using unlabeled **O1** as substrate. The fragmentation of **O2** can be visualized when this oligonucleotide is labeled. The remarkable finding of this experiment was that very little self-cleavage of **O2** was observed (Figure 4.9). This result is consistent with the high efficacy observed for the cleavage of **O1** by **O2**.

One of the goals in designing chemical nucleases was to achieve catalytic activity. If a chemical nuclease molecule survives during a redox process, it may be possible to re-use the molecule in subsequent rounds of the cleavage reaction. Because the bipyridine modified oligonucleotide retains its molecular backbone integrity during the cleavage reaction, it was of interest to determine why **O2** was unable to cleave the remaining 20% of substrate, and why it did not function catalytically. The experiment designed to probe the nuclease activity of **O2** following the cleavage reaction was carried out as follows. A cleavage reaction was carried out as stated in the procedure

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

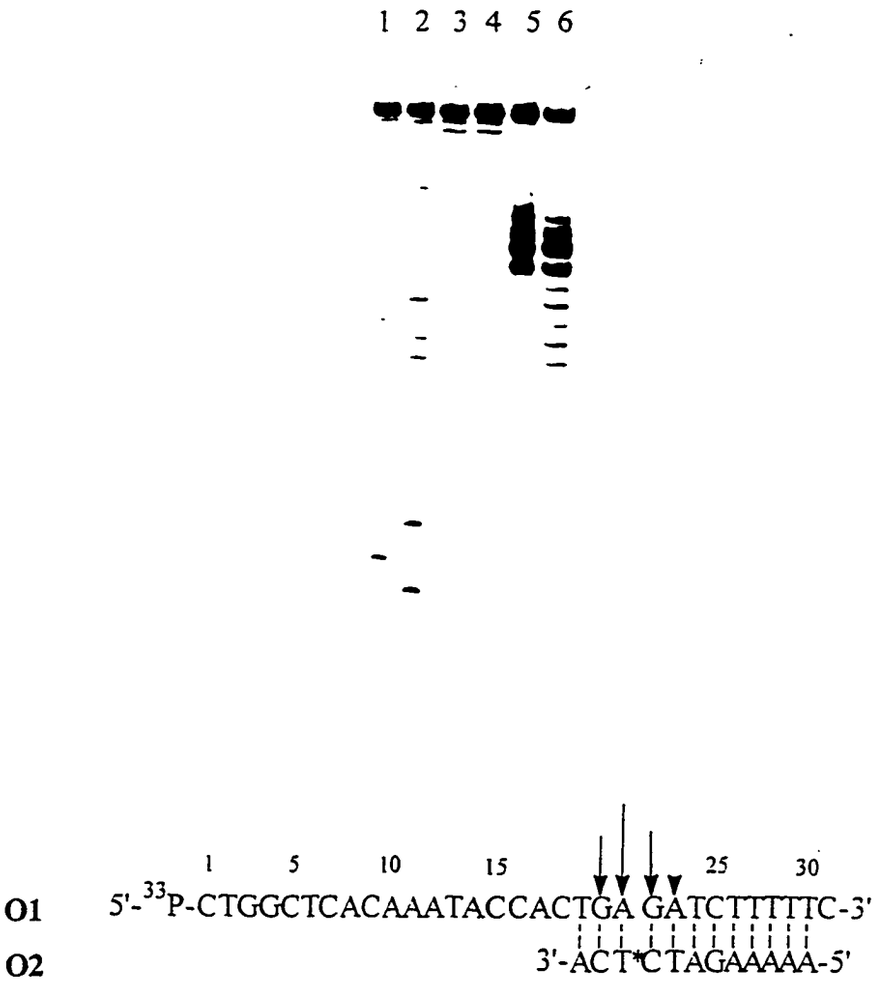


Figure 4.7 Oligonucleotide **O2** mediated cleavage of substrate **O1** in the presence of hydrogen peroxide. Lane 1: Maxam-Gilbert sequencing (A + G); lane 2: Maxam-Gilbert sequencing (T + C); lane 3: **O1**; lane 4: **O1** subjected to Fe^{2+} and H_2O_2 in the absence of **O2**; lane 5: reaction of **O1** with Fe^{2+} and H_2O_2 in the presence of **O2**; lane 6: same as lane 5, but the reaction products were treated with 1 M piperidine at 90°C for 30 minutes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

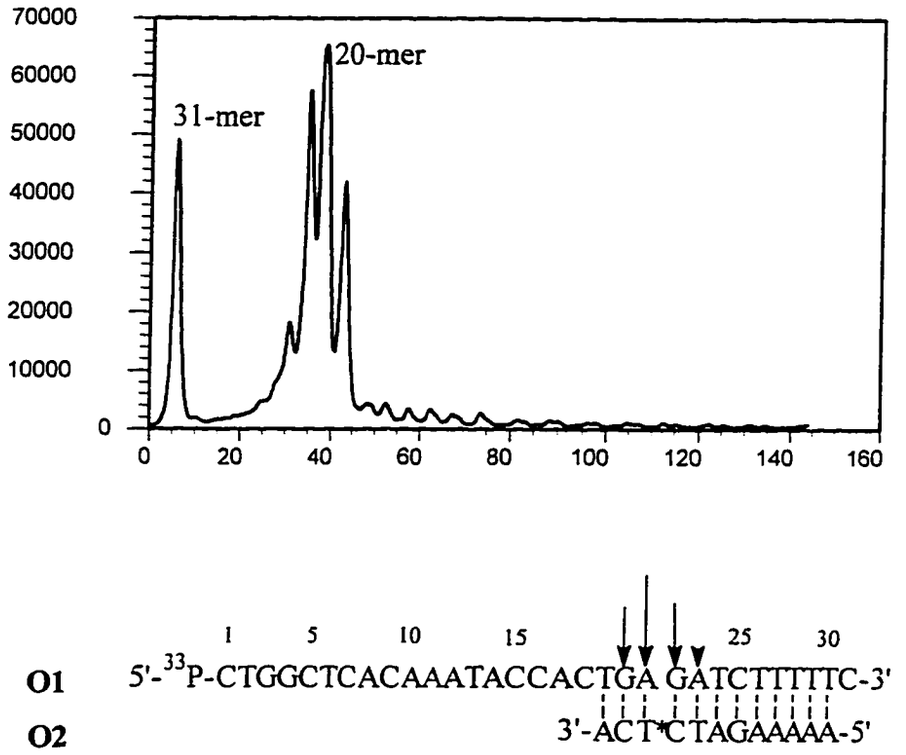
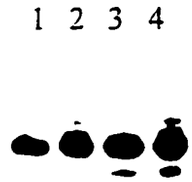


Figure 4.8 Phosphor Imaging™ quantitation of substrate oligonucleotide **O1** cleavage distribution. Profile of lane 5 in Figure 4.7 on previous page.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



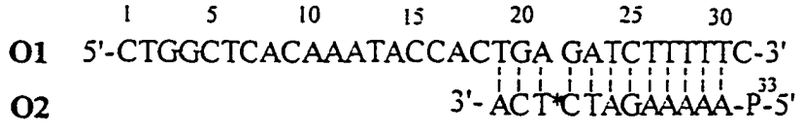


Figure 4.9 Self-cleavage assay of bipyridine modified oligonucleotide **O2**. Lane 1: 5′-end ^{33}P labeled **O2**; lane 2: **O1** (unlabeled) and **O2** (labeled) in the absence of Fe^{2+} and H_2O_2; lane 3: **O2** treated with Fe^{2+} and H_2O_2; lane 4: reaction of **O2** with Fe^{2+} and H_2O_2 in the presence of **O1** (unlabeled).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(section 4.6) but with unlabeled substrate **O1**. After three hours of the reaction at room temperature, a second aliquot of 5′-end ^{33}P labeled **O1** was added to the above reaction system. The same cleavage reaction procedure was carried out again. On analysis of the cleavage yield of the second portion of **O1**, we found the substrate **O1** was cleaved in only 10% yield (Figure 4.10). This result appears to indicate that the **O2** nuclease activity was substantially lost during the first round cleavage reaction, even though its oligonucleotide chain was not fragmented. This observation may imply that either the bipyridine ligand was damaged during the oxidation process, therefore it was unable to complex and position an iron at the target site, or the redox potential of the oxidized-bipyridine iron complex was changed unfavorably to mediate an electron transfer. It has been reported that when the 2,2′-bipyridine has substituents on the aryl rings, its ability to complex transition metal ions and the resulting complex redox potentials do change.[160] However, the nature of the bipyridine-oligonucleotide deactivation has not yet been understood.

To assess the sequence effect of the cleavage ability of a bipyridine tethered oligonucleotide, the bipyridine ligand was positioned in the middle of different sequences of different length. The sequences **O1/A9**; **O1/A13**; **O1/B9**; **O1/B13** are shown in Figure 4.11. In sequence **JC31** the bipyridine was positioned close to the 3′-end of the oligonucleotide (Figure 4.11). The cleavage efficiency of **JC31** was 80% over a three nucleotide cutting span (Figure 4.12). However, when the bipyridine ligand was positioned at the center of a DNA duplex, its nuclease activity was dramatically reduced for all four of the sequences tested. The cleavage yields for **O1/A9**; **O1/A13**;

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

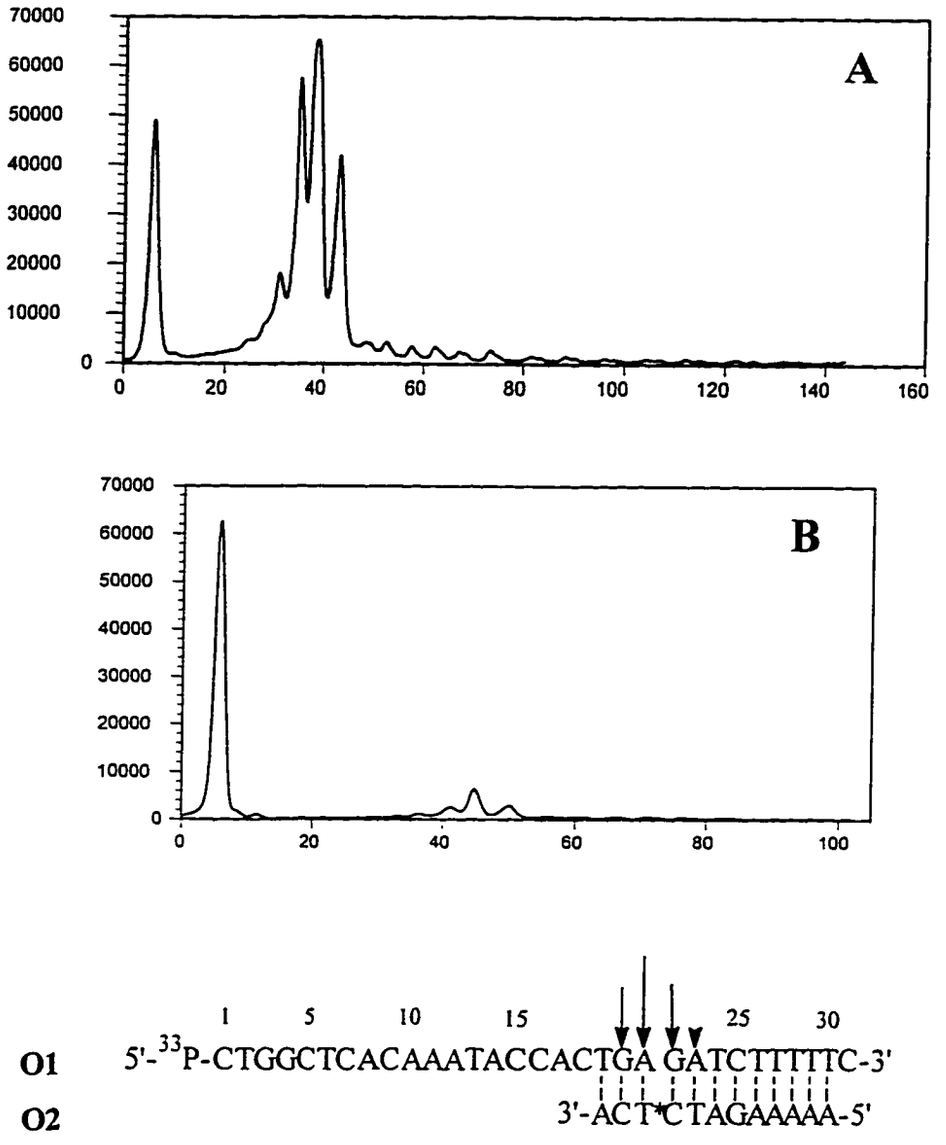


Figure 4.10 **O2** Nuclease activity assay after first round cleavage reaction. Panel A: Cleavage reaction products quantitation. Panel B: Second portion of ^{33}P labeled substrate **O1** cleavage profile after **O2** cleaved a first portion of unlabeled substrate **O1**, all other reaction conditions are the same for both sequential cleavages. **O2** shows reduced nuclease activity after a cleavage reaction.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

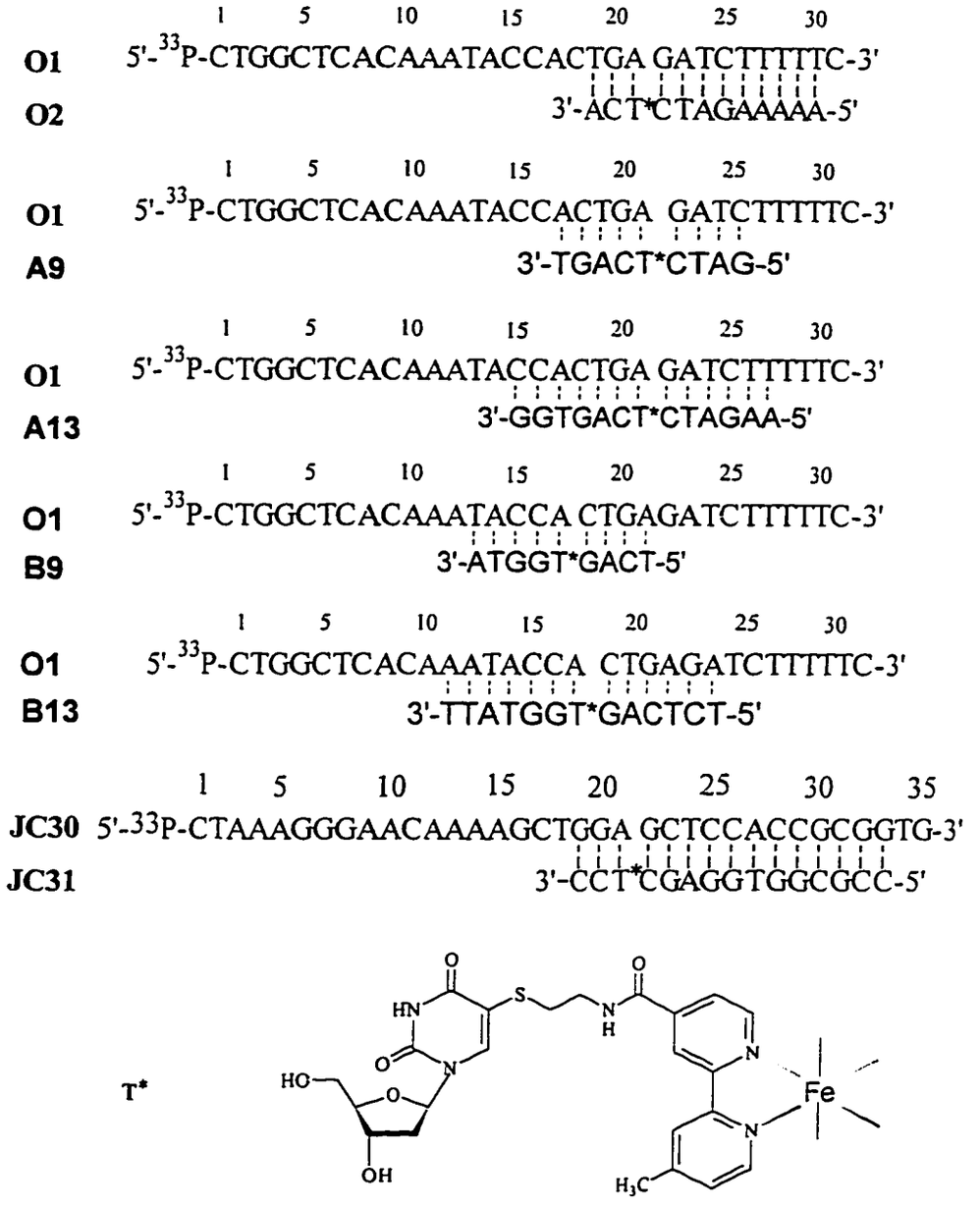


Figure 4.11 Assay of sequence, bipyridine position and targeting oligonucleotide length effect on bipyridine modified oligonucleotide nuclease activity. All sequences are perfectly matched (complementary).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

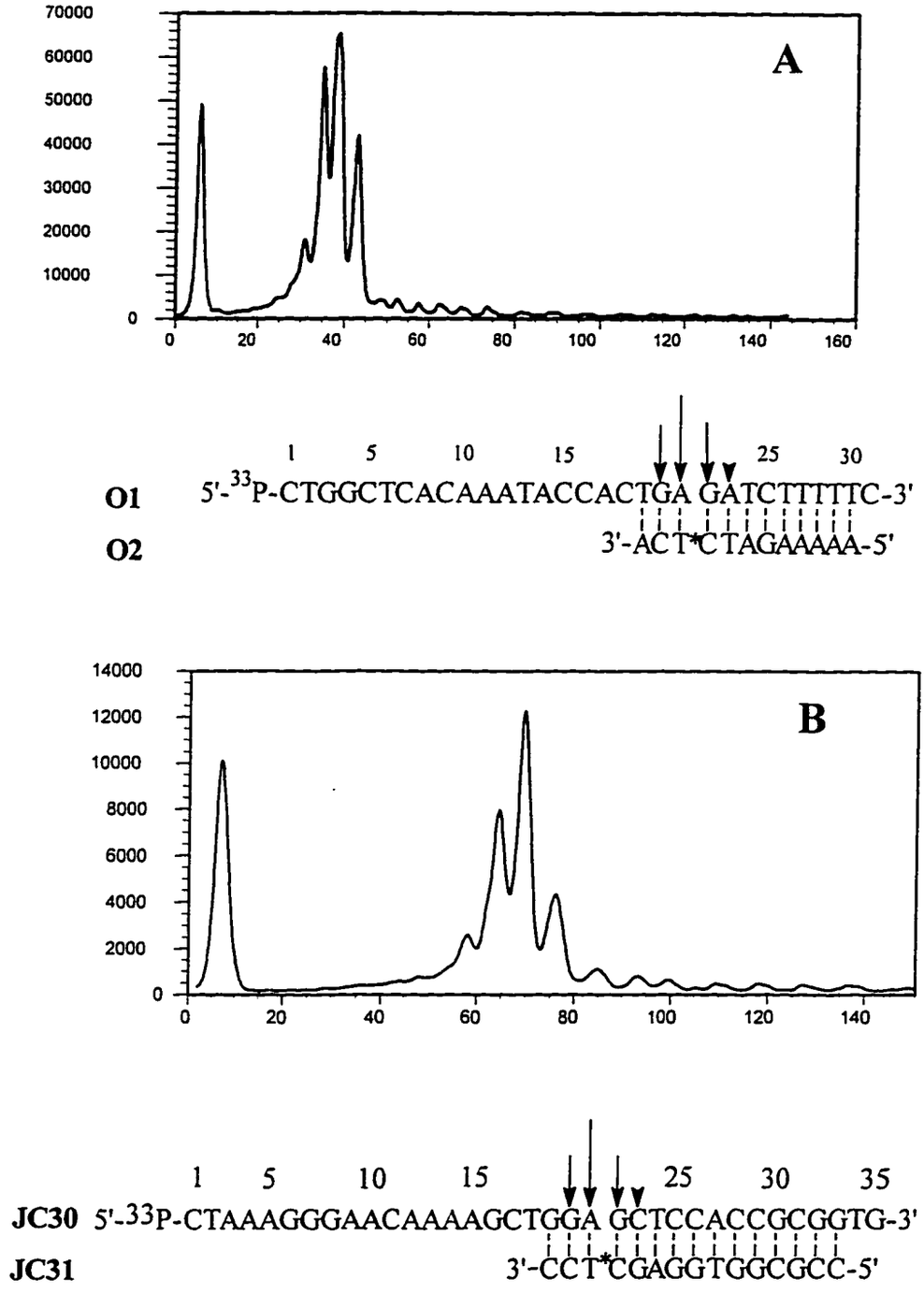


Figure 4.12 Nuclease activity comparison in different sequences. When T* is positioned close to the 3′-end of a targeting oligonucleotide, it can consistently give the same cleavage pattern with 80% cleavage yield regardless of sequence.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

O1/B9; O1/B13 are 21.6%, 26.2%, 12.6% and 44.4%, respectively. Longer sequences with a centered bipyridine (A13 & B13) have relatively higher cleavage yields than the shorter sequences (A9 and B9) (Figure 4.13). It is known that the ends of a DNA duplex are more readily dissociated, so that a bipyridine-Fe complex positioned near the end would have more flexibility than one located at the center of a sequence. The above experimental observation implies that the redox active bipyridine-Fe complex may require flexibility to accommodate additional coordination site(s) and/or cleavage site(s) on the substrate.

The dependence of nuclease activity of the bipyridine-Fe complex on its local structure motivated us to further study the nuclease activity of the bipyridine-Fe complex when it was positioned in a region with structural modifications that perturb local duplex structures. The perturbed DNA duplex structures include: a one base bulge at three different positions; a one four-base-loop; a mismatch and a deletion (Figure 4.14). Each of the six types of non-matched DNA duplexes would be expected to have different local structures. Our experimental results showed that in these six different structural environments the chemical nuclease **A13** had very different activities, with cleavage yield ranging from 0 to 40% and specificity ranging from a two nucleotide span to a four nucleotide span. In all six cases the cleavage yields were far less than those of **O2** and **JC31** (Figure 4.15).

When free ferrous ion or ferrous complex ion and hydrogen peroxide are mixed together, a high potential oxidative species is generated. This process is broadly termed as Fenton chemistry, but the nature or the mechanism of this oxidation is an issue of on-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

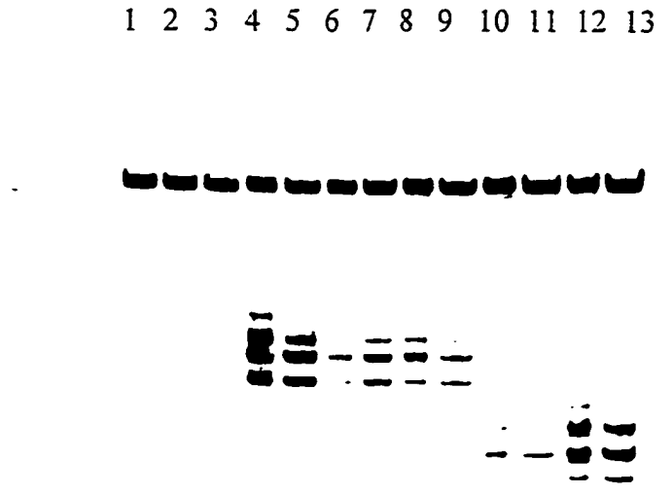


Figure 4.13 Effect of bipyridine position on nuclease activity. Lane 1: Maxam-Gilbert sequencing (A+G); lane 2: Maxam-Gilbert sequencing (T+C); lane 3: labeled substrate **O1** only; lane 4, 6, 8, 10, 12: **O2**, **A9**, **A13**, **B9**, **B13** as targeting oligonucleotides, respectively; lane 5, 7, 9, 11, 13: **O2**, **A9**, **A13**, **B9**, **B13** as targeting oligonucleotides, respectively, but the reaction products were treated by 1M piperidine at 90°C for 30 minutes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

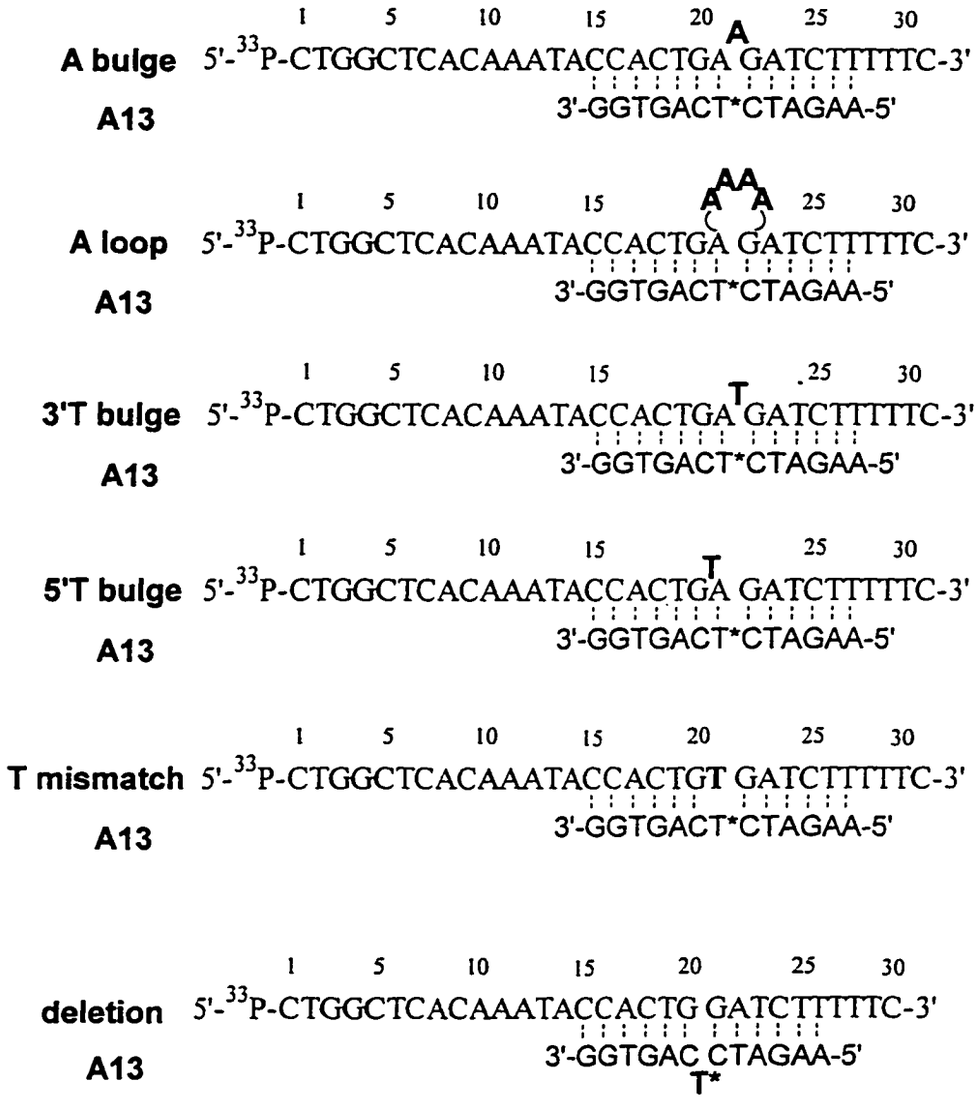


Figure 4.14 Nuclease activity assay for structurally perturbed duplexes. In these six sequences, T* is positioned in a region with A bulge, four A loop, 3′-end T bulge, 5′-end T bulge, mismatch with T and deletion of paired A (T* bulged out).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

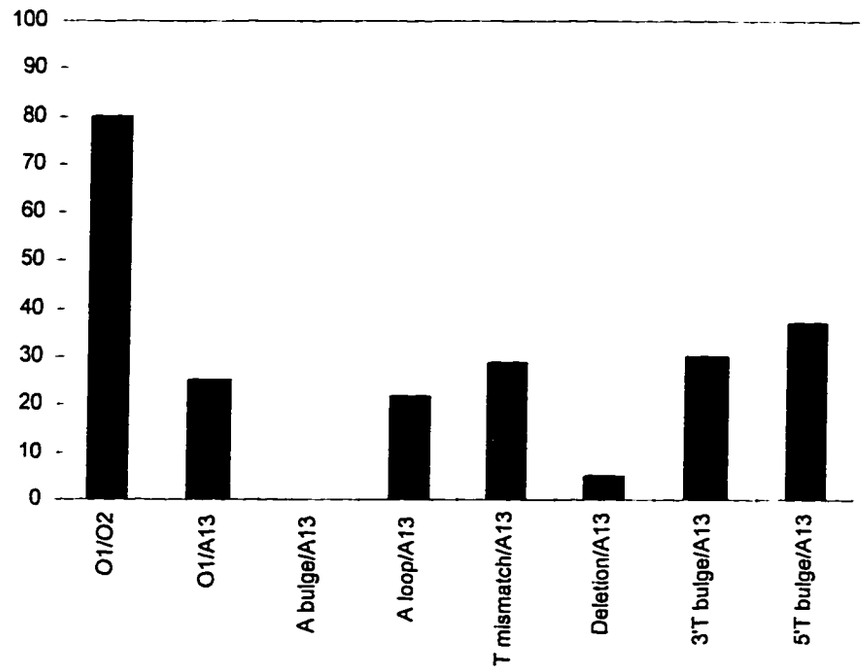


Figure 4.15 Cleavage yield quantitation for non-perfectly matched DNA duplexes. The cleavage yields of the six perturbed sequences were quantitated and compared with perfect matches of **A13** and **O2.** In different local structural environment, T^* showed different nuclease activity. When T^* was positioned close to the 3′-end (**O2**) the nuclease efficiency was best.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

going debate. Figure 4.16 shows the proposed mechanisms involved in Fenton chemistry. A ferrous ion, Fe(II), is first oxidized by hydrogen peroxide to Fe(III), which could be subsequently reduced by a reductant. After the hydrogen peroxide gains one electron, it can be transformed to hydroxyl radical or ferryl-oxo. It is these oxygen species which most likely eventually attack the target(s). In EDTA-Fe(II) and phenanthrolene-Fe(II) systems, a reductant is required for DNA cleavage.[84,127] The reductant is used to reduce the Fe(III) back to Fe(II), thereby propagating the redox cycle. In the bipyridine-Fe complex system, it was found that Fe(III) still had significant nuclease activity (57% cleavage yield) without the need for external reductants (Figure 4.17, Figure 4.18). This observation suggests again that a new DNA cleaving mechanism is involved in the bipyridine modified oligonucleotide.

In our efforts to maximize the substrate cleavage yield, we found that the nuclease activity of the bipyridine modified oligonucleotide is also affected by other variables. The elucidation of these effects may shed some light on the mechanism of the cleavage reaction. As Figure 4.19 shows, the cleavage yield increases as the reaction time increases up to five hours. After ten hours the amount of parent oligonucleotide and the fragmented oligonucleotides begin to reduce. On the basis of this observation, 3 - 5 hours reaction time was chosen as the optimal time for the cleavage reactions. Other investigators have showed that for the EDTA-Fe system a one hour reaction time is sufficient to achieve maximum cleavage.[57] In EDTA-Fe cases the free hydroxyl radical is generally believed to be the active intermediate. The reactive intermediate in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

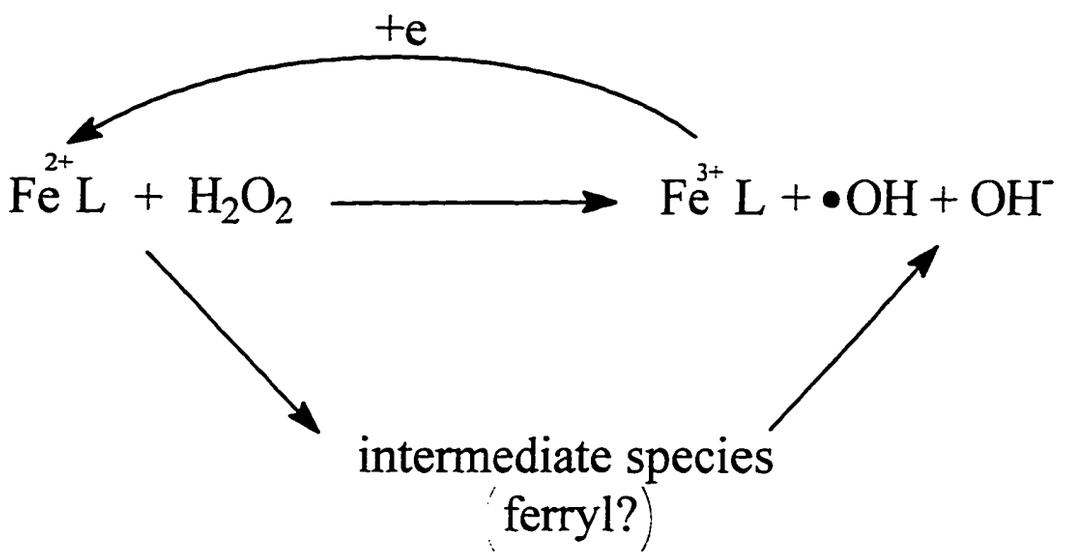
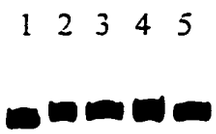
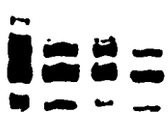


Figure 4.16 Fenton chemistry. The Fenton chemistry definition is broadly used to refer to a mixture which has ferrous or ferrous complex and hydrogen peroxide. The nature of the real oxidative species in this mixture is an issue of debate. Depending on the state of chelation of the iron, including DNA chelation, the structure of the oxygen radicals may vary[161] For the EDTA-Fe system, it is generally accepted that free hydroxyl radical is the real oxidative species. The principal evidence comes from spin-trap, scavenger experiments and aromatic hydroxylation. Some researchers proposed an alternative mechanism in which a tightly bound intermediate, possibly a ferryl radical (FeO^{2+}), is formed first, and then decomposes to a more reactive hydroxyl radical.[162] Externally added reductants could recycle the ferric ion back to ferrous and make the iron ion catalytic. Pre-oxidized ferric ion showed reduced oxidative potential.[163-165]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





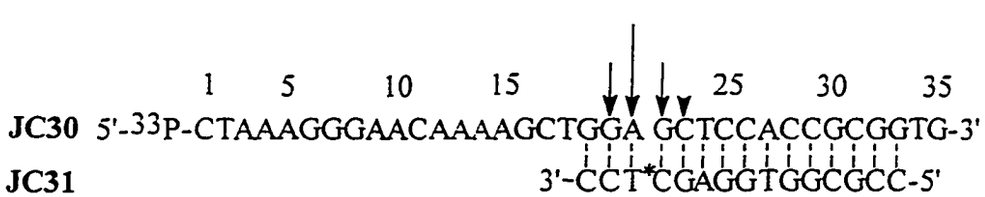


Figure 4.17 Fe(II) and Fe(III) mediated DNA cleavage. Lane 1: control, labeled substrate **JC30**; lane 2: cleavage reaction in the presence of Fe(II), lane 3:as lane 2, but reaction products were treated by 1M piperidine at 90°C for 30 minutes; lane 4: cleavage reaction in the presence of Fe(III); lane 5: same as lane 4, but the reaction products were treated by 1M piperidine at 90°C for 30 minutes.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

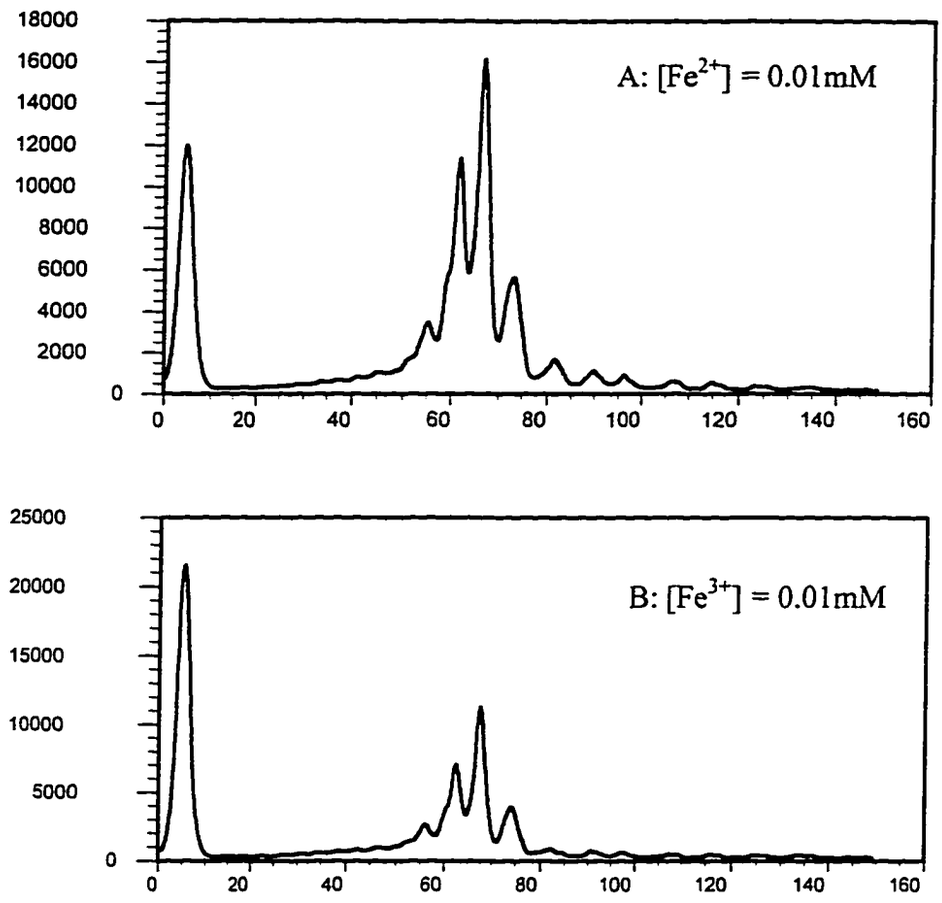
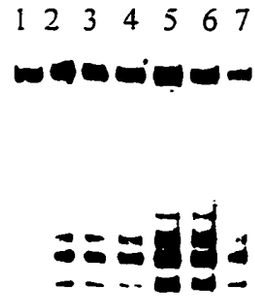


Figure 4.18 Quantitation of Fe(II) & Fe(III) mediated DNA cleavage. Panel A: cleavage is mediated by ferrous ion and hydrogen peroxide; panel B: cleavage is mediated by ferric ion and hydrogen peroxide. Fe(III) has significant nuclease activity without any external reductants.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



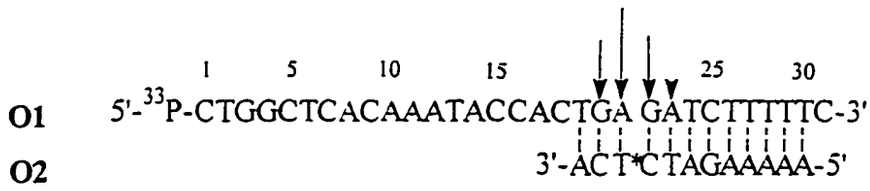


Figure 4.19 Effect of time on the cleavage yield. Cleavage reaction system: 0.9 pmol **O1** substrate, 4.5 pmol **O2**, 0.01 mM Fe(II), 0.4 mM H_2O_2 and room temperature. Lane 1: **O1** control; lane 2: 0.5 hour reaction time; lane 3: 1 hour reaction time; lane 4: 2 hours reaction time; lane 5: 5 hours reaction time; lane 6: 10 hours reaction time; and lane 7: 24 hours reaction time.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the bipyridine-Fe system appears to be less reactive, which would be consistent with a ferryl-oxo as the reactive intermediate.[166]

The effect of salt concentration on DNA duplex stability is known: the higher the salt concentration, the more stable the DNA duplex. This occurs because the negatively charged phosphate-phosphate repulsion between the strands is reduced by the counter ions. The salt effect in bipyridine modified chemical nuclease is shown in Figure 4.20. When the sodium chloride concentration is raised from 0.05M to 0.25M, the cleavage yield is reduced. We also tried the reaction in 1M NaCl, but at this high salt concentration the bands on polyacrylamide gel turned to small dots. There are two possible explanations for the salt effect. First, at higher salt concentration the two strands are held more tightly together, so DNA duplex flexibility is reduced. The resulting motional restriction may be unfavorable to the nuclease activity of the bipyridine-Fe complex, because the bipyridine-Fe complex may require changes in the conformations preferred for duplex DNA in order to achieve coordination to the opposing strand. This explanation is consistent with the previous observation: a centered bipyridine-Fe complex is less efficient than one positioned near the end. Second, when the salt concentrate is high, a few ligands on the bipyridine-iron complex may be replaced by the salt anions. This replacement subsequently modulates the redox potential of the redox active complex, and this redox potential modulation may retard the electron transfer process mediated by the complex. As a result, 0.05 M NaCl was generally used in the DNA cleavage reactions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

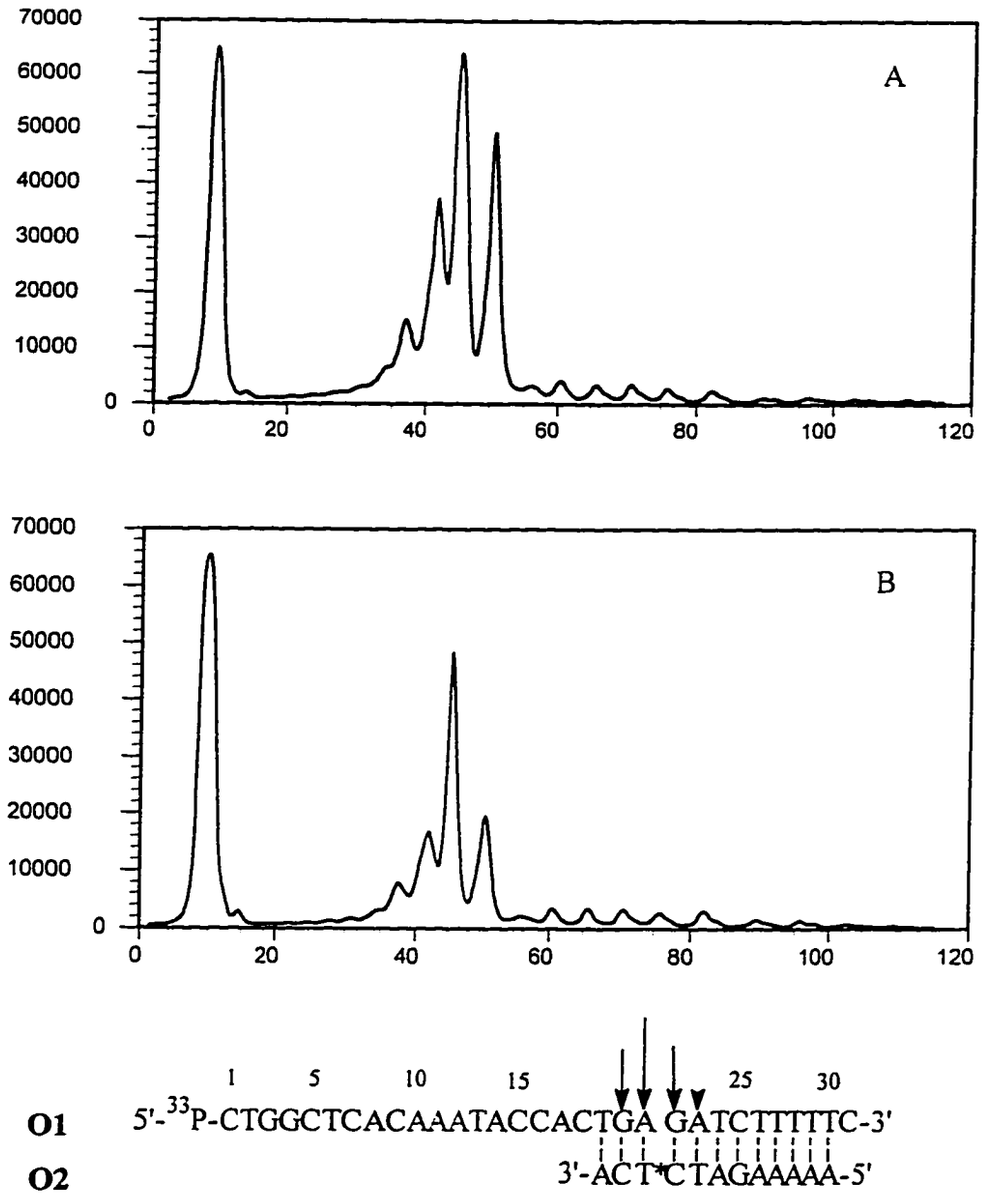


Figure 4.20 Salt concentration effect on DNA cleavage. Panel A: [NaCl]=0.05 M; panel B: [NaCl]=0.25 M. Other reaction conditions are: 0.9 pmol **O1**, 4.5 pmol **O2**, [Fe(II)]=0.01 mM, [H_2O_2]=0.4 mM, three hours reaction time at room temperature.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Temperature may also play a role on the nuclease activity of this chemical nuclease. The temperature dependence was tested for **O1/O2** and the six non-perfectly matched sequences. The results are shown in Figure 4.21, Figure 4.22 and Figure 4.23. For the A bulge, deletion, A loop and T mismatch sequences (Figure 4.14), the temperature had no noticeable effect on cleavage profiles. But on the other hand, for the perfect match **O1/O2**, 3′-end T bulge and 5′-end T bulge sequences, lower cleavage yields resulted at lower temperature. This difference in nuclease activity could be a result of structural variations caused by temperature. A high resolution experimental structure of the chemical nucleases, either from NMR spectroscopy or x-ray crystallography, will be critical to gain a deeper understanding of the complex and its function. Structure elucidation is one of the long range goals of this project.

The bipyridine ligand was designed to be positioned in the major groove of a DNA duplex, with the intention to explore the major groove proton abstractability, specifically H3′. It was of interest to determine whether the observed cleavage originated from H3′ abstraction. If the DNA cleavage is initiated by H3′ abstraction *and* this step is rate limiting in the series of steps which lead to the DNA scission, then we should be able to observe the H/D isotope effect when the H3′ is replaced by ^{2}H3′. It has been reported that DNA cleavage by activated Bleomycin has an isotope effect on 4′-C-H bond in the range from 2 to 4.5,[167] and DNA cleavage by Neocarzinostatin has an isotope effect on 1′-C-H bond in the range from 3.3 to 4.2.[168] It had been proposed that these hydrogen abstraction steps are the rate-limiting steps in DNA lesion by Bleomycin and Neocarzinostatin, which are DNA minor groove cleavers.[167,168]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

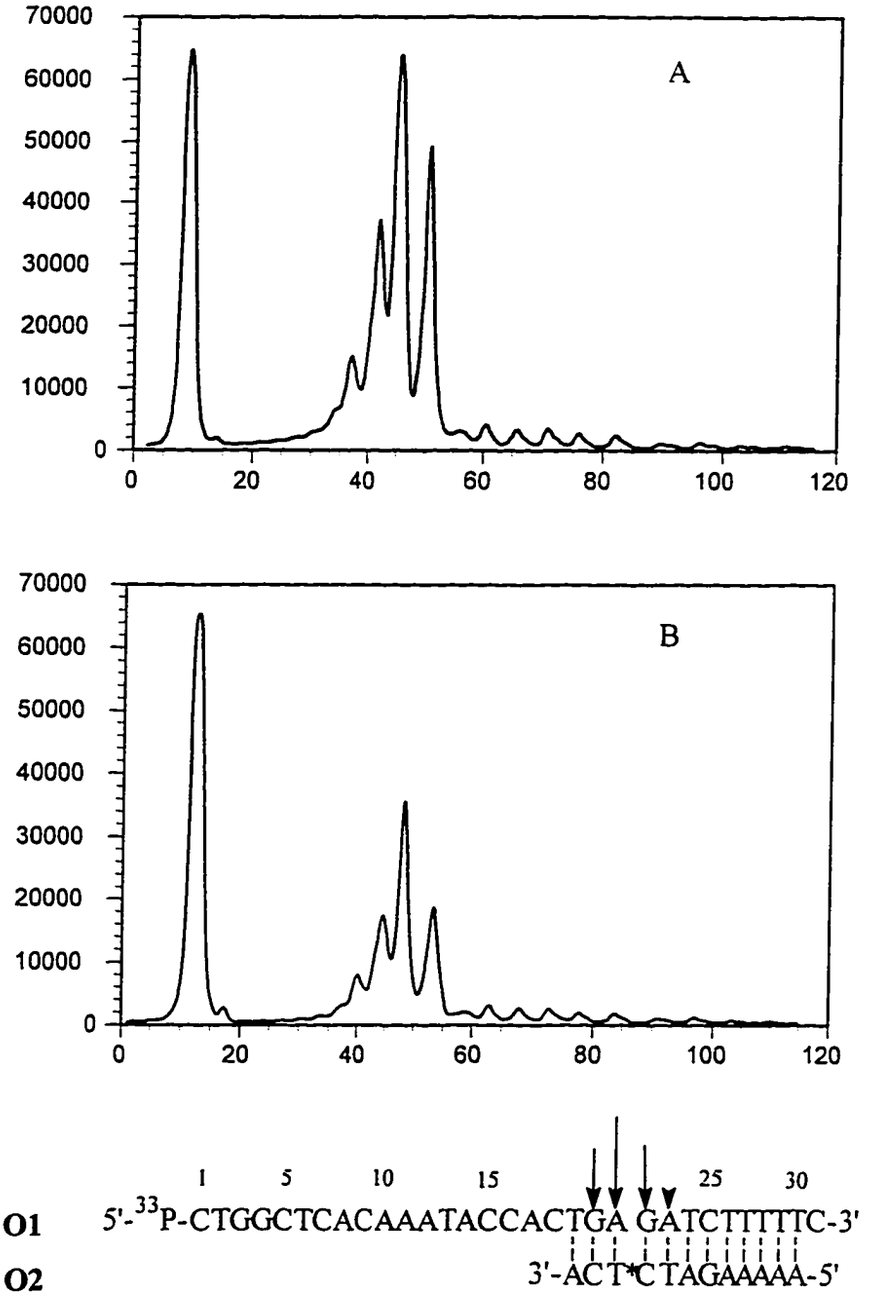


Figure 4.21 Temperature effect on nuclease activity of perfectly matched duplex. This is an **O1/O2** system, with reaction conditions: 0.9 pmol **O1**, 4.5 pmol **O2**, [Fe(II)]=0.01 mM, $[H_2O_2]$=0.4 mM, and three hours reaction time. Panel A: reaction took place at room temperature; panel B: reaction took place at 0°C temperature.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

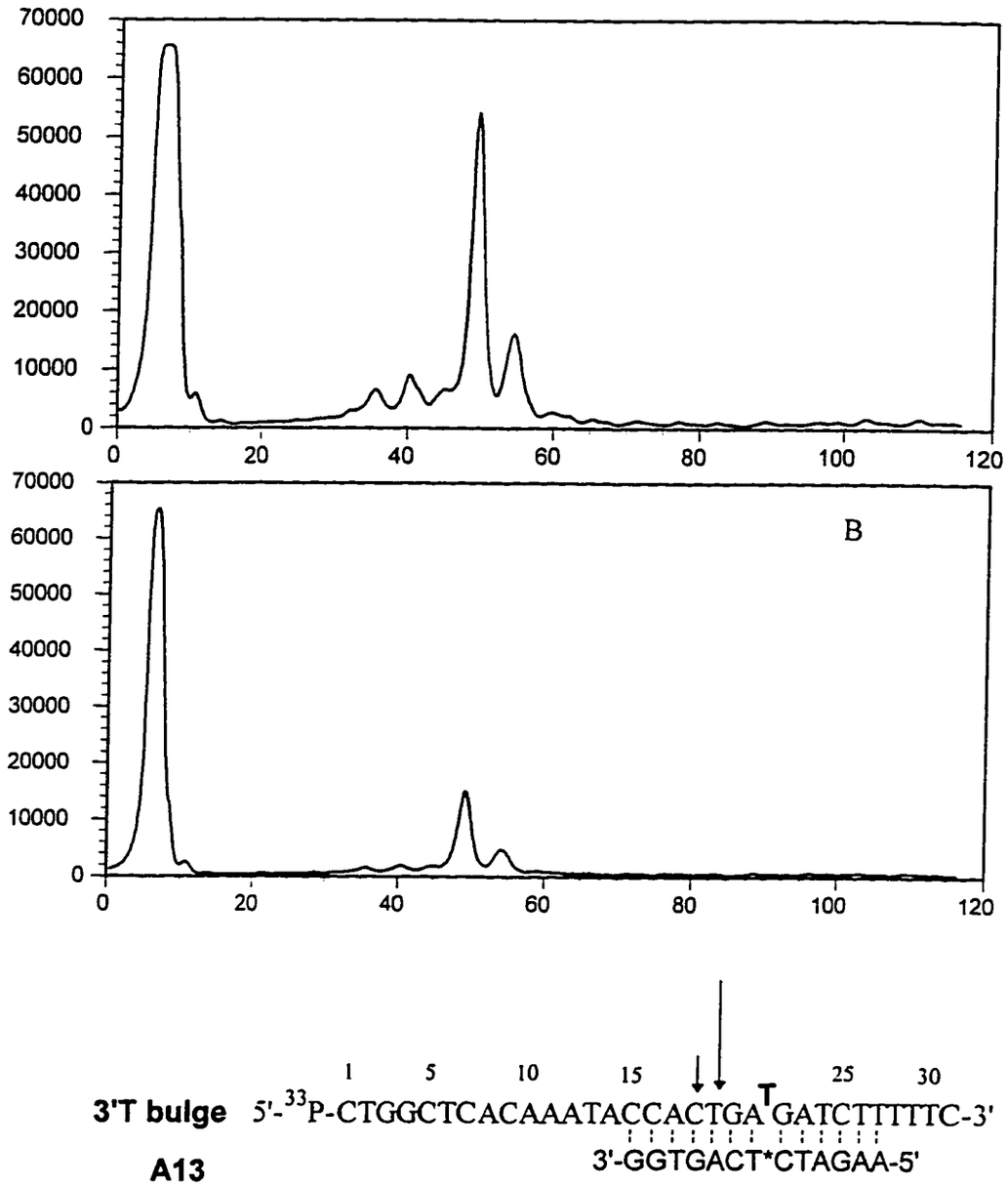


Figure 4.22 Temperature effect on cleavage efficiency of 3′-end T bulge substrate (**3′T bulge**). This is a **3′T bulge/A13** system, with reaction conditions: 0.9 pmol **3′T bulge**, 4.5 pmol **A13**, [Fe(II)]=0.01 mM, [H_2O_2]=0.4 mM, and three hours reaction time. Panel A: reaction at 25°C; panel B: reaction at 0°C.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

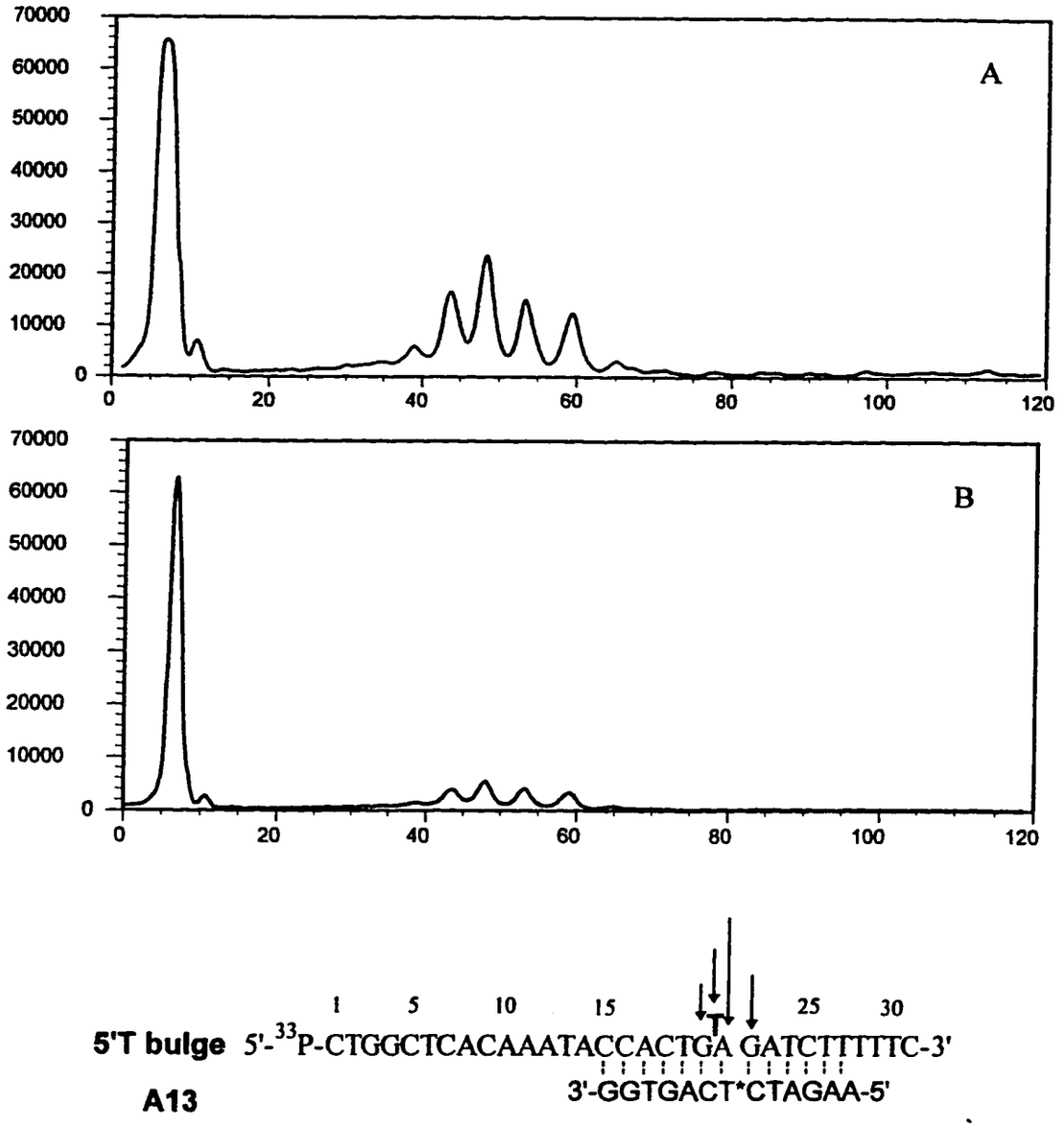


Figure 4.23 Temperature effect on cleavage efficiency of 5′-end T bulge substrate (**5′T bulge**). This is a **5′T bulge/A13** system, with reaction conditions: 0.9 pmol **5′T bulge**, 4.5 pmol **A13**, [Fe(II)]=0.01 mM, [H_2O_2]=0.4 mM, and three hours reaction time. Panel A: reaction at 25°C; panel B: reaction at 0°C.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In the **O1/O2** cleavage systems, there are three main cutting sites on the substrate **O1** (Figure 4.8). If A21 has a ^{2}H3′ and G20/G22 both have a H3′, an isotope effect would be easily observed by using G20/G22 as internal references (Figure 4.24). We synthesized and delivered the ^{2}H3′-A phosphoramidite into substrate **O1** at the A21 site, where it was paired with T*. The cleavage reaction was carried out with the ^{2}H3′-A21 **O1** as substrate and **O2** as the targeting strand. The resulting cleavage profile showed, somewhat unexpectedly, an almost identical cleavage pattern and cleavage yield as that of unlabeled substrate **O1** (Figure 4.25). This result clearly indicates that the H3′ abstraction is not the rate limiting step in the substrate DNA cleavage process, if it is involved in DNA scission at all. The other possible rate-limiting-steps are : 1) bipyridine-Fe complexation to site(s) on the complementary strand that position the iron for hydrogen abstraction; 2) hydrogen abstraction at the only other major groove positioned hydrogen: H2′′(on the same side as H3′ on the deoxyribose ring); 3) a major structural rearrangement in which the DNA assumes an unknown structural form placing the iron in close proximity to deoxyribosyl 5′, 4′ or 1′ protons. A theoretical study has shown that H2′ abstraction is less energetically favorable than H3′ abstraction.[153] However, it has been reported that the H2′/H2′′ protons can be involved in DNA strand scission.[149,169] It has been suggested that it may be possible for the bipyridine-Fe complex to intercalate into the helix with the result that cleavage could be initiated in the minor groove.[143-148,170,171] Only a few examples of bipyridine metal complex intercalating into DNA as known,[170-172] and this intercalation may be only for a square planar metal complex. It is not likely that the octahedral bipyridine-Fe

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

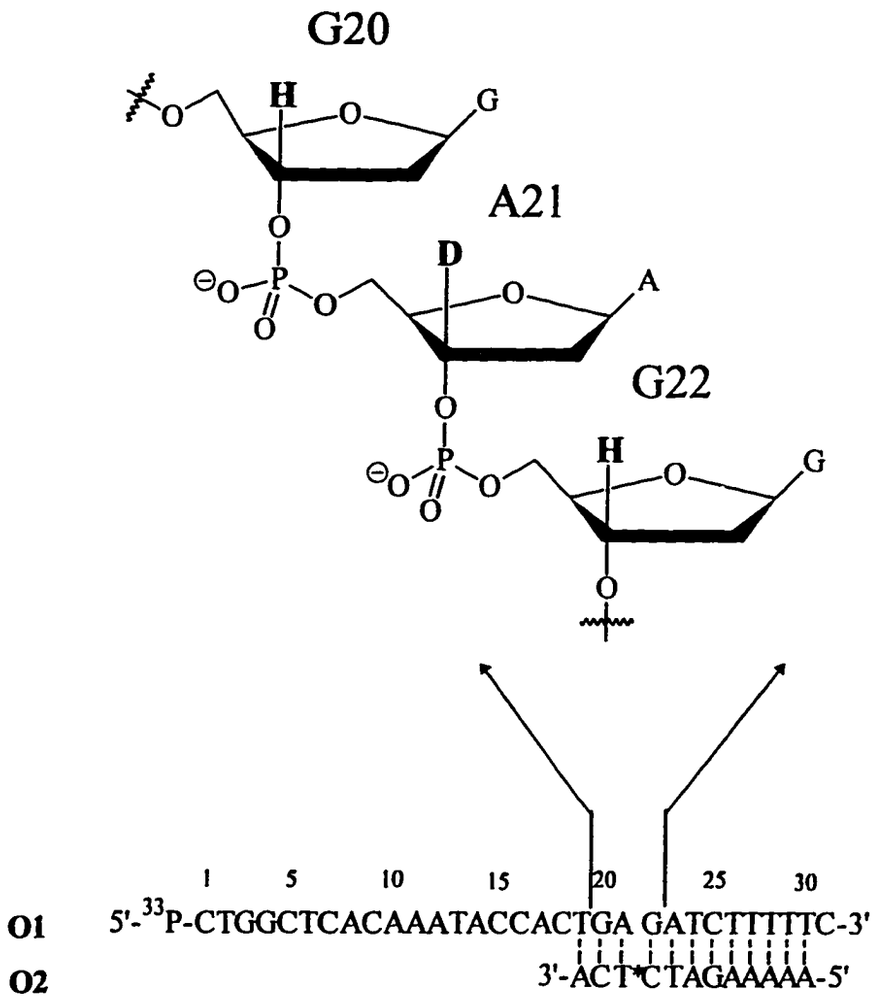


Figure 4.24 Deuterium labeling on substrate. A21 on **O1**, which is paired with T*, is labeled with ^{2}H at C3′ of the deoxyribose. The neighboring cutting sites, G20 and G22, are unlabeled. If there is an isotope effect on the substrate, there would be a change in relative peak intensity.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

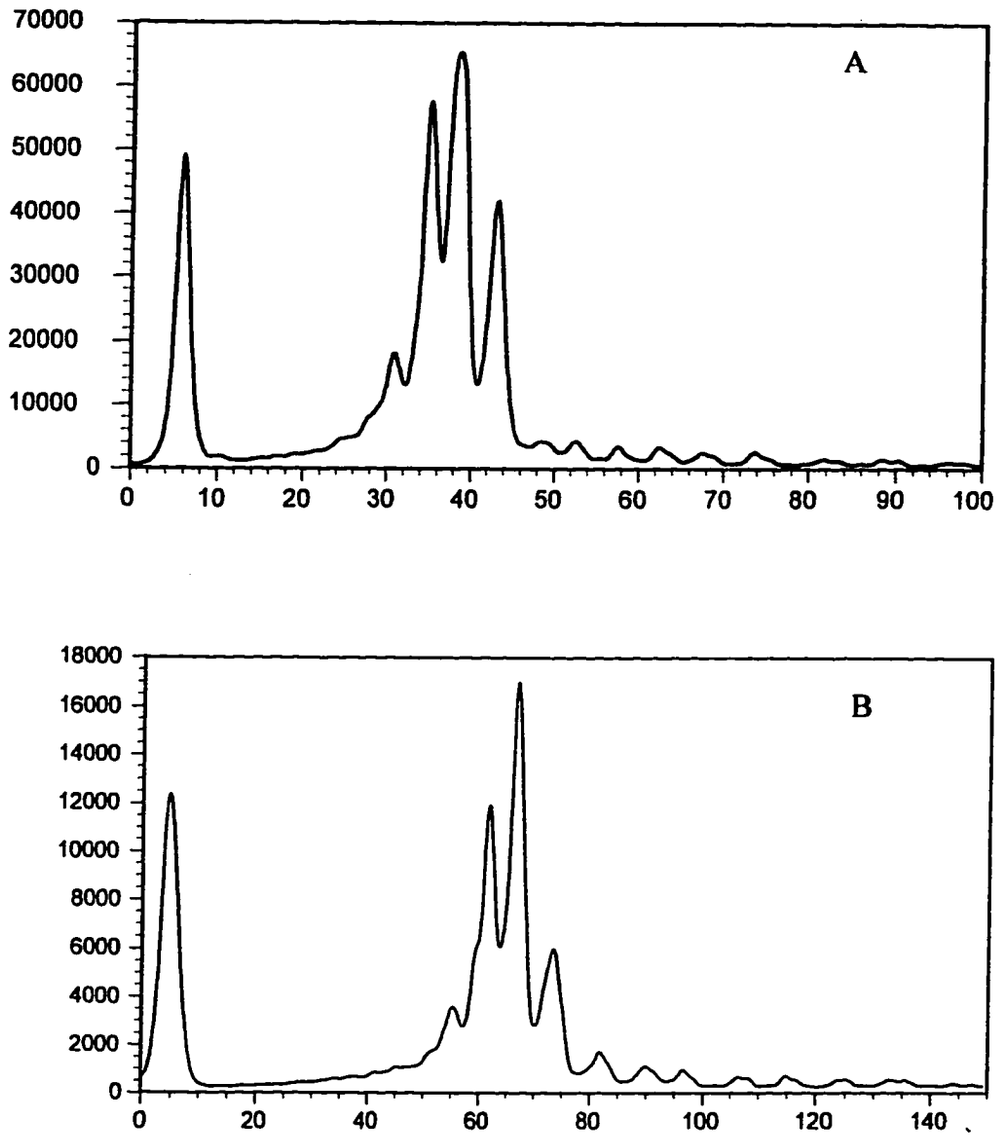


Figure 4.25 Cleavage profile comparison between H3′-A21 **O1** and D3′-A21 **O1.** Panel A: H3′-A21 **O1** substrate; panel B: D3′-A21 **O1** substrate. These results show that the cleavage pattern and cleavage yield are virtually identical in regardless of the deuterium incorporation in the substrate.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

complex would intercalate into DNA. A T_m study of the corresponding DNA duplex may provide some hint on how the bipyridine-Fe complex interacts with the DNA duplex. This work is under way.

4.4.2 Other transition metal mediated single-stranded DNA cleavage

Some other transition metal ions can also mediate redox processes in DNA cleavage. For the bipyridine modified oligonucleotide system, we tested Cu(II), Mn(II) and Ni(II). Cupric ion does not show nuclease activity, when it is alone, with hydrogen peroxide or with a reductant. It was found in this laboratory that the minor groove positioned phenanthrolene-copper has no nuclease activity either, but this may be due to geometry constraints. When a copper complex mediates a redox process, it needs to change geometry between tetrahedral (Cu(I)) and square planar (Cu(II)). In the minor groove of a DNA duplex, this geometry interchange is restricted, so the copper complex is not able to mediate the electron transfer process. Nevertheless, this geometry restriction may not be the reason for the nuclease inactivity of the bipyridine-copper complex when it is positioned in the major groove of DNA duplex. The nuclease inactivity of the bipyridine-copper complex may be a result of a change in redox potentials or inability of the complex to transfer one electron between oxidant and reductant. Other copper complexes with many different ligands have been shown to have oxidative nuclease activity.[55,58,72,75,78,107-109,113]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The bipyridine-Mn(II) and bipyridine-Ni(II) complexes were also tested for nuclease activity. Readable patterns for the DNA bands on polyacrylamide gel could not be obtained with either complex.

4.4.3 Probing the active oxidative species

The high nuclease specificity and lack of self-cleavage of the bipyridine modified oligonucleotides clearly indicates that the active oxidative species in this cleavage system is highly localized. Free diffusible hydroxyl radical can be excluded. The highly localized nature of the active species was further supported by hydroxyl radical scavenging experiments. DMSO is widely used as a free hydroxyl radical scavenger in DNA cleavage/damage experiments.[173,174] In the bipyridine modified oligonucleotide cleavage system, when the DMSO concentration was as high as 0.24M, the substrate **JC30** still cleaved by **JC31** with high efficiency (60% cleavage yield) (Figure 4.26). More interestingly, when the DMSO concentration reached 0.08mM (59% cleavage yield), the DMSO cleavage suppression effect had already leveled off (Figure 4.26). In some transition metal mediated DNA cleavage/damage systems, DMSO can totally quench the substrate DNA cleavage/damage.[173,174] In those systems, the hydroxyl radical is generally believed to be the responsible oxidative species. The observation that bipyridine-Fe(II)/H_2O_2 nuclease activity can't be totally suppressed by DMSO suggests that hydroxyl radical was not the only responsible oxidative species in DNA cleavage, if the hydroxyl radical is involved at all. Ethanol

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

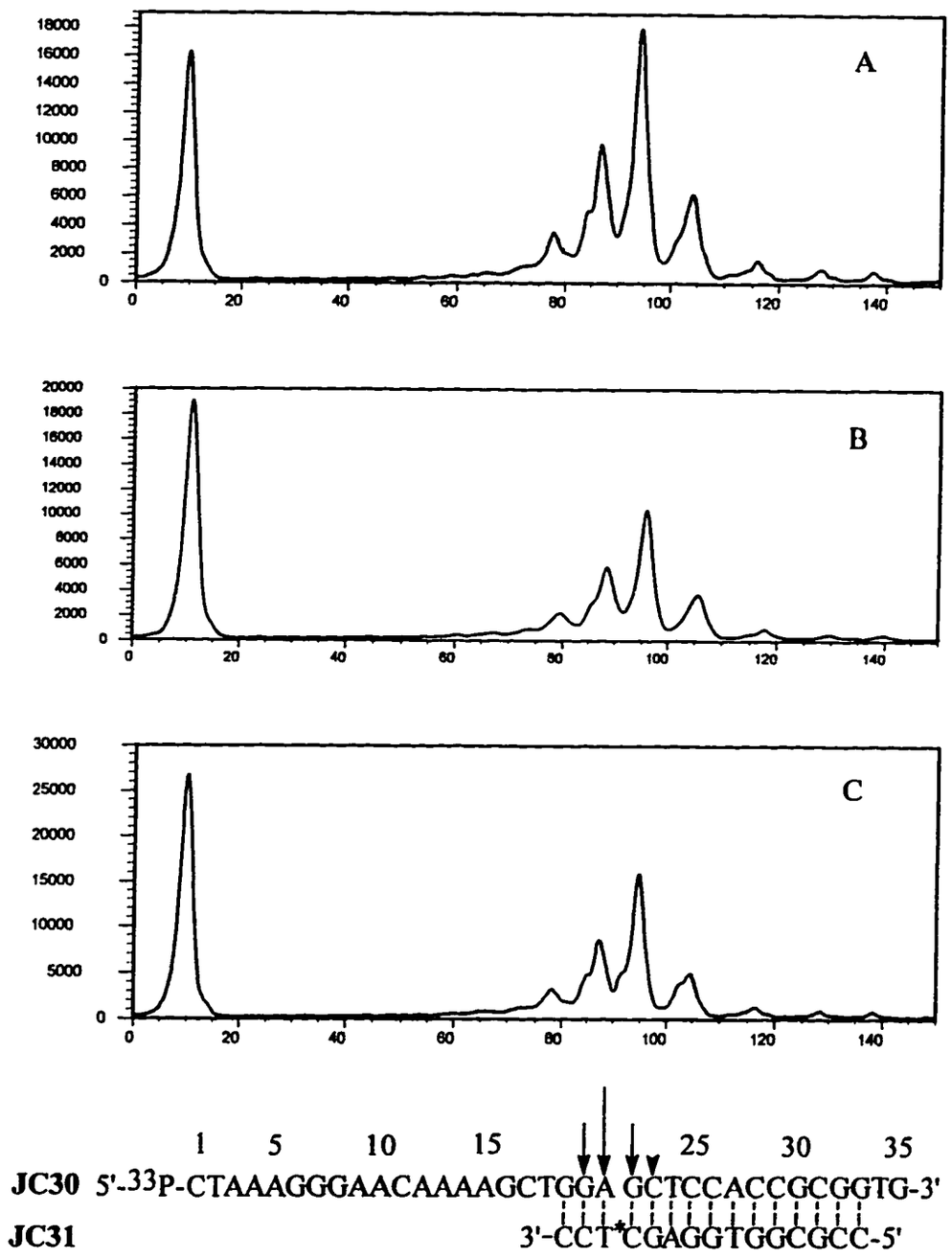


Figure 4.26 Cleavage profiles in the presence of DMSO. Panel A: cleavage reaction in the absence of DMSO; Panel B: cleavage reaction in the presence of [DMSO]=0.08M; panel C: cleavage reaction in the presence of [DMSO]=0.24M. The substrate is **JC30**, and the targeting oligonucleotide is **JC31**.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and azide were not able to quench the nuclease activity of bipyridine modified oligonucleotide either.

There are two possibilities for the nature of the oxidative species in the bipyridine system: 1) metal bound oxygen species; 2) locally trapped hydroxyl radical. These possibilities were supported by recent literature. Wink and co-workers[175] suggested that under conditions in which free hydroxyl radical production would be predicted, iron may in fact form an iron-oxo or iron-peroxy species. Shepherd and co-workers also demonstrated that Fe^{II}(haph) complexes produce a small quantity of freely diffusible hydroxyl radicals in the presence of hydrogen peroxide, but the predominant active species formed in the presence of both hydrogen peroxide and molecular oxygen is iron bound.

4.5 Discussion and conclusion

The results presented in the above sections showed that a bipyridine modified oligonucleotide can chelate a Fe(II) or Fe(III) in the major groove of DNA duplex, and the bipyridine-Fe(II)/bipyridine-Fe(III) complexes can efficiently cleave the complementary DNA strand in the presence of hydrogen peroxide. These results demonstrate that an efficient chemical nuclease which functions through a major groove bound ligand can be designed.

The nuclease specificity of the bipyridine tethered oligonucleotides occurs over a three nucleotide cutting span (Figure 4.8). EDTA-Fe oligonucleotide chemical nucleases[53,54,56,63] usually show a significantly broader cutting span. The specificity of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

bipyridine oligonucleotide is much better than that of EDTA-Fe oligonucleotide. This higher specificity implies that the nature of the oxidative species is a non-diffusible one. The bipyridine-Fe(II) complex in the major groove of DNA duplex is positioned much closer to the substrate strand than to the self strand, because no self strand fragmentation is observed.

The nuclease efficiency of bipyridine modified oligonucleotide is uniquely high: 80% cleavage yield with only five fold excess of the chemical nuclease. One of the reasons for the high efficiency may be the lack of self-cleavage. Self-cleavage is a common problem in redox based chemical nucleases, and this is at least one of the reasons for low chemical nuclease efficiency.

However, we have never achieved 100% substrate cleavage yield, even though the chemical nuclease does not fragment during the cleavage process. According to the nuclease activity loss after one round cleavage reaction, it is likely that the electron rich bipyridine ring was oxidized during the substrate DNA oxidation process. The oxidized bipyridine will reduce its ability to chelate a metal ion and subsequently the nuclease activity. However, the nature of the damaged bipyridine has yet to be determined.

We considered the possibility that two, or even three, bipyridine oligonucleotides could complex with one ferrous ion. This concern can be expelled if the following factors are taken into account. 1) Under the cleavage condition generally employed in this study, the ferrous ion concentration is about 25 times of that of the bipyridine oligo. 2) The first complexation constant of the bipyridine-ferrous complex is higher than that of the second step ($\log\beta_1$=4.36; $\log\beta_2$=3.64; $\log\beta_3$=9.45[176]). 3)Steric

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

hindrance disfavors three bipyridine-oligonucleotides coordinating with one central metal ion, because bipyridine-oligonucleotide is more bulky than a free bipyridine.

When a new metal-ligand complex shows nuclease activity, one of the key questions concerns the nature of the responsible oxidative species. In general, all possible oxidants can be categorized into two groups: non-diffusible (metal bound) and diffusible. The nuclease activity of bipyridine oligonucleotides occurs over a three nucleotide region, there is no self cleavage and neither DMSO, ethanol or azide exert a scavenger effect. Together these results indicate that a non-diffusible oxygen species is likely to be responsible for substrate DNA cleavage. The three nucleotide cutting span is likely to be caused by the local motion of the bipyridine-Fe compiex, which can be easily visualized by molecular dynamics simulation. It has to be pointed out that it may be very difficult to probe the exact nature of the reactive center because of its intrinsic high reactivity and sensitivity to probes and reaction conditions.

The non-specific substrate background cleavage is a concern with redox type chemical nucleases. The bipyridine modified oligonucleotides also showed visible background cleavage ladders, even though this accounted for less than 5% of the substrate DNA. The non-specific background cleavage is due to the non-specific binding between free irons and nucleic bases and/or phosphates. In the presence of hydrogen peroxide, the locally bound iron can initiate oxidation on the substrate, which results in non-specific cleavage. A portion of the non-specific initial damage occurred on the nucleic bases, which appeared only after alkaline treatment.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

We attempted several ways to eliminate the non-specific background cleavage. One way was to include calf thymus DNA (CT DNA) in the cleavage reaction system. We hoped that a large excess of CT DNA would scavenge the non-specific oxidative species and decrease background cleavage of the substrate DNA. A cleavage reaction was carried out in the presence of large excess of CT DNA (mass ratio of CT DNA to substrate oligonucleotide is about 230 and mole ratio of Fe(II) to substrate oligonucleotide is 111; the CT DNA was added either before the addition of Fe(II) or after the addition of Fe(II)), but the amount of background cleavage did not measurably change. DMSO, ethanol and azide were also not able to remove the background cleavage. Possibly the background cleavage could not be eliminated due to the presence of trace iron in the solvents and reagents used in the reaction. If pre-existing iron binds tightly to the nucleic bases and phosphate, then CT DNA may not be able to replace the target DNA strand under the reaction conditions. It seems very difficult to eliminate non-specific background cleavage. However, some DNA damage could be fixed in biological systems by repair enzymes. In the bipyridine modified oligonucleotide system, the non-specific cleavage amounts to less than five percent overall, so it should not prevent us from further exploring potential molecular biology applications for this type of chemical nuclease.

Even though the bipyridine modified oligonucleotides appear to function effectively as chemical nucleases (80% cleavage yield), we still believe that there is room for further improvement in cleavage specificity and efficiency. Before designing the next generation of chemical nuclease, it will be necessary to obtain a high resolution

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

structure of a bipyridine chemical nuclease by either NMR spectroscopy or x-ray crystallography.

Nevertheless, the specificity and efficiency are sufficient to justify the exploration of molecular biology applications. If combined with other DNA recombination methodologies, the lack of precision (cleavage over a three nucleotide span) may not be a significant disadvantage. The details of a preliminary exploration of possible applications are discussed in the next chapter.

4.6 Experimental

Materials: Reagents were purchased from Aldrich Chemical Company, Inc. and were used without further purification unless otherwise noted. Moisture sensitive reagents were kept in a dry box. Anhydrous solvents were purchased from Aldrich Chemical Company, Inc. as Sure/Seal™ packages. Solvents for chromatography were spectroscopic or HPLC grade. Silica gel used for chromatography was Silica Gel 60 (230-400 mesh) purchased from EM Science. Thin layer chromatography plates was purchased from Waterman Inc., as glass-backed silica gel 60 with 1/4mm thickness and containing UV indicator. All TLC plates were visualized by UV (254nm) shadowing or by iodine complexation in an iodine chamber. The 5′-O-(4,4′-dimethoxytrityl) containing compounds were visualized by dipping the plates into a 20% sulfuric acid water solution.

The reagents for polyacrylamide gel electrophoresis were purchased from Life Sciences, Inc.. The electrophoresis apparatus was purchased from Life Sciences, Inc. as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Electrophoresis Model S2. The cleavage reagents and ammonium iron(II) sulfate hexahydrates were purchased from Aldrich Chemical Company, Inc.. Hydrogen peroxide (30% water solution) was purchased from Mallinckrodt Specialty Chemicals Co.. [γ-^{33}P]ATP was purchased from either Amersham Life Sciences or DuPont. The T4 polynucleotide kinase was purchased from New England Biolabs. The DNA transferring Zeta Probe Blotting membrane was purchased from Bio-Rad Laboratories. A speed-vac model RH-20-12 from Savant Instruments, Inc. was used for water evaporation from oligonucleotide samples. A scintillation counter model Liquid Scintillation Analyzer 1900 TR from Packard Instrument Company was used for ^{33}P labeling quantitation.

Equipment and Procedures: All reactions were carried out in standard glass flasks. When necessary, an N_2 atmosphere was applied to the reaction systems. Crude product chromatography separations were performed either by using flash chromatography[177] with the columns and fittings purchased from Ace, Inc., or on centrifugally accelerated radial thin-layer chromatography on a model 7924T chromatotron (Harrison Research) equipped with a model QG150 pump (Fluid Metering Inc.). The Chromatotron TLC plates (1mm, 2mm or 4mm thickness) were prepared with the Silica Gel $60PF_{254}$ containing gypsum (EM Science), and the separations were monitored by UV at 254nm.

^{1}H and ^{31}P nucleic magnetic resonance spectra were recorded on a Varian VXR-500 spectrometer, or Brucker AC250. Chemical Shifts (δ) for ^{1}H spectra were recorded as parts per million downfield from the internal standard tetramethylsilane. Chemical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

shifts (δ) for ^{31}P spectra were reported as parts per million downfield from the external standard 85% phosphoric acid.

Mass spectrometric analysis were performed by the mass spectrometry facilities at the Purdue University Department of Medicinal Chemistry and Molecular Pharmacology. Compounds were analyzed by fast atom bombardment (FAB) in positive ion mode in appropriate matrixes.

Oligonucleotides were purified on preparative polyacrylamide gel, and desalted on a Sep-Pack C-18™ column (Waters Chromatography, Inc.), and quantified on a Cary model 3 Bio UV-Visible spectrophotometer, and stored at -20°C. 5′-End ^{33}P labeling was performed by using [γ-^{33}P]ATP following the manufactures protocol. The labeling efficiency was monitored by a scintillation counter. The oligonucleotide sequencing was performed using the Maxam-Gilbert method.[178,179] After electrophoresis, ^{33}P labeled DNA bands were transferred from polyacrylamide gel to nylon membrane. The bands were visualized by exposing Kodak X-ray film to the membrane, and were quantified on a Bio-Rad Phosphor Imaging™ with Molecular Analyst™ program, PC version 1.3.

The DNA cleavage protocol was varied in different experiments, but a typical procedure used for oligonucleotide substrate cleavage reactions with bipyridine tethered chemical nucleases is outlined in the following steps. Step 1. 20pmol Substrate oligonucleotide was ^{33}P labeled with [γ-^{33}P]ATP (Amersham or DuPont) and T4 kinase (New England Biolabs). Step 2. The ^{33}P labeled oligonucleotide solution was separated into 0.9pmol aliquots, and each aliquot was placed in a 1ml eppendorf tube. The 0.9pmol substrate was mixed with 4.5pmol (5eq) bipyridine modified targeting

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

oligonucleotide. The mixed aqueous solution was then dried on a speed-vac. Step 3. The dried oligonucleotide mixture was redissolved in 8μL cleavage reaction buffer, which was composed of 50mM TRIS and 50mM NaCl at pH 7.4. This reaction mixture was heated at 90°C 5 minutes for denaturing, and then cooled at room temperature for 30 minutes for annealing. In cases where vapor condensed on the eppendorf tube lid, it was necessary to spin the tube to force the condensed solvent down into the main solution. Step 4. 1μL 0.1mL Freshly prepared ammonium iron sulfate aqueous solution and 1μL 4mM freshly prepared hydrogen peroxide aqueous solution were added sequentially into the above DNA solution to initiate the reaction. Step 5. After 30 minutes at room temperature, 1μL 4mM freshly prepared hydrogen peroxide aqueous solution was added to the reaction mixture. Step 6. After 3 hours of total reaction time at room temperature, the reaction was quenched by adding 1μL 10mM EDTA. The quenched reaction was either put on a speed-vac for drying, or 100μL 1M piperidine was added and the mixture heated at 90°C for 30 minutes. After the alkaline treatment, the reaction was dried on a speed-vac. Step 7. Loading buffer (3.5μL, formamide:10TBE, 9:1, v/v) was added to the dried pellet to redissolve the residue. Immediately before the reaction mixture was loaded onto a polyacrylamide gel for analysis, it was denatured at 90°C for 5 minutes and then dipped into ice. Step 8. The cleavage reaction mixture (3μL) was loaded onto a 0.4mm analytical polyacrylamide gel. The gel concentration is determined by the length of the oligonucleotide to be analyzed. 1μL Dye was loaded onto the gel to track the oligonucleotide mobility. The gel was run at 1500~1800 volts and 50 watts for 2~4 hours to reach the desired resolution. Step 9. After electrophoresis, one of the sandwich

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

glass plates was removed and a wetted nylon membrane carefully laid onto the DNA sample region on the gel. Generally, 45 minutes was enough for the oligonucleotides to transfer from polyacrylamide gel onto the nylon membrane. The nylon membrane was then exposed to a Kodak X-ray film after drying. The X-ray film was developed for visualizing the oligonucleotide bands. Step 9. For oligonucleotide quantitation, this radioactive oligonucleotide containing membrane was exposed to a scanning plate , and then the plate scanned by a Phosphor Imaging™ scanner. The digitized image was subjected to Molecular Analyst™ software for analysis, PC version 1.3.

Preparation of 5-(chloromercuri)-2′-deoxyuridine (2): [94,96-101] 2′-Deoxyuridine (14.2gram, 62.2mmol) dissolved in 60ml water, was combined with mercury acetate (20.9gram, 65.6mmol, 1.05eq) dissolved in 90ml water. The solution was diluted by addition of 50ml water and the temperature raised to 50°C. Within a few minutes the cloudy white solution turned clear. After 30 minutes, the solution turned milky and a white precipitate appeared. After three hours reaction, the solution was cooled to 40°C, and 15ml saturated sodium chloride aqueous solution added. The solution was further cooled to room temperature with stirring. The final precipitate was filtered and washed sequentially with 100ml 0.1M sodium chloride solution and 30ml water. The precipitate was collected and dried in air to yield 21.06g product. 73% yield. 5-Chloromercuri-2′-deoxyuridine is not soluble in any solvents and is typically used without characterization. The dried solid 5-chloromercuri-2′-deoxyuridine product was well ground before using in subsequent reactions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Preparation of N,N′-bis(trifluoroacetyl)cystamine (10): To a solution of cystamine dihydrochloride (9.0g, 40mmol) and triethylamine (27.8ml, 200mmol) in 500ml chloroform (filtered through basic alumina) in an ice-bath was added trifluoroacetic anhydride (12.5ml, 88mmol) dropwise. The yellow reaction solution was stirred magnetically overnight. After evaporation of the solvent on a rotary evaporator, the residue was stirred with 100ml pure water for 20 minutes and filtered. The solid was washed with 35ml 1M hydrochloric acid twice, 35ml water twice and dried in air. The final solid product was slightly yellow and weighed 8.56g (62%). R_f=0.75 in ethyl acetate. FAB MS: [MH^+]345.

Preparation of 5-(2-trifluoroacetamido-1-thioethyl)-2′-deoxyuridine (3):[93] A lithium palladium chloride methanol solution (0.1M) was prepared by dissolving lithium chloride (1.696g, 0.04mmol) and palladium chloride (3.546g, 0.02mmol) in 200ml absolute methanol. The solution was stirred overnight to give a clear red-brown solution. 5-Chloromercuri-2′-deoxyuridne **(2)** (4.62g, 10mmol) and N,N′-bis(trifluroacetyl)cystamine **(10)** (8.56g, 25mmol) were added to the 0.1M Li_2PdCl_4 methanol solution (200ml). After 30 minutes, the solution had turned clear red-orange. The reaction mixture was stirred overnight at room temperature. Hydrogen sulfide gas was bubbled into the reaction mixture for one minute, and then the solution was filtered. The filtrate was evaporated on a rotary evaporator, and the residue was chromatographed on silica gel eluting with 20% methanol in chloroform to yield 2.40g compound **3** as a white solid (60% yield). R_f=0.48 in 20% methanol chloroform. ^1H NMR (CD_3OD): δ8.32(s, 1H, H6), 6.27(t, 1H, J=6.6Hz, H1′), 4.43(m, 1H, H3′),

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

3.95(m, 1H, H4′), 3.79(m, 2H, H5′/H5′′), 3.47(t, 2H, NCH_2), 2.87(m, 2H, H2′/H2′′), 2.32(t, 2H, SCH_2). FAB MS: $[MH^+]$400. Data are consistent with those reported in literature.[93]

Preparation of 5-(2-amino-1-thiapropyl)-2′-deoxyuridine (4):[93] Compound **3** (4.60g, 11.5mmol) was dissolved in 200ml methanol and 200ml 30% NH_4OH. The reaction was stirred overnight at room temperature. After evaporation of the solvent, the residue was redissolved in 50ml water, and lyophilized. The crude product was chromatographed on silica gel eluting with 8% Et_3N in methanol to yield final white product 3.310g (94% yield). ^1H NMR (CD_3OD):δ 8.40 (s, 1H, H6); 4.40 (m, 1H, H3′); 3.93 (dd, 1H, J=3Hz, J=6Hz, H4′); 3.82(dd, 1H, J=3Hz, J=12Hz, H5′ or H5′′); 3.73 (dd, 1H, J=3Hz, J=12Hz, H5′′ or H5′); 2.79-2.75 (m, 4H, $SCH_2CH_2NH_2$); 2.32-2.20 (m, 2H, H2′/H2′′). FAB MS: $[MH^+]$304. Data are consistent with those reported in literature.[93]

Preparation of 4′-methyl-2,2′-bipyridine-4-carboxylic acid (8):[94,152]Commercial 4,4′-dimethyl-2,2′-dipyridyl (5.00g, 27.2mmol) and selenium dioxide (4.53g, 40.8mmol, 1.5eq) were added sequentially into CaH_2 distilled pyridine (24ml). The reaction mixture was heated to 115°C and stirred for 20 hours. The color of the solution turned from light yellow to brown, and finally to gray. The reaction mixture was filtered and the solid washed twice with 30ml warm pyridine. The solvent was removed from the red-brown filtrate on a rotary evaporator. For purification, the carboxylic acid crude product was converted into an ester. To the above dried crude product, 200ml absolute methanol and 5ml concentrated sulfuric acid were added. This

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reaction mixture was refluxed for 20 hours. After evaporation of the solvent on a rotary evaporator, the residue was redissolved in 200ml chloroform. The separated organic layer was washed by 30ml saturated $NaHCO_3/H_2O$ twice, and 30ml saturated $NaCl/H_2O$ once. The washed organic solution was dried over anhydrous sodium sulfate, and evaporated on a rotary evaporator to yield a yellow brown residue. The crude product was chromatographed on silica gel (115g) eluting with $EtOAc/C_6H_{14}/CH_3OH$ (4/6/0.5%, v/v/v) to yield 1.30g monoester product as a white solid. The monoester was hydrolyzed by treating with pyridine (12ml) and water (18ml) at reflux temperature for 20 hours. After evaporation of the solvent on a rotary evaporator, the residue was redissolved in 150ml water. The product was recrystallized out of the aqueous solution, and dried in vacuum oven to yield **8** as a white solid (0.775gram, 14% yield). More product could be recovered from the slightly pink colored aqueous supernatant. ^1H NMR (DMSO-d6): δ8.85(dd, J=5Hz, J=1Hz, 1H, H5), 8.80(d, J=1Hz, 1H, H3), 8.56(d, J=5Hz, 1H, H6), 8.25(d, J=0.5Hz, 1H, H3′), 7.85(dd, J=5Hz, J=1.5Hz, H6′), 7.31(dd, J=5Hz, J=0.5Hz, H5′), 2.41(s, 3H, CH_3). FAB MS: $[MH^+]$205. Data are consistent with those reported in literature.[152]

Preparation of 4′-methyl-2,2′-bipyridine-4-succinimidylcarboxylate (9): A solution containing **8** (0.775g, 2.50mmol), N-hydroxylsuccinimide (0.461g, 3.00mmol, 1.2eq) and 18ml DMF was chilled to 0°C. To this solution was added DCC (1.13g, 3.75mmol, 1.5eq) in 5ml DMF. This reaction mixture was stirred in an ice-bath for 30 minutes and then transferred into a freezer for precipitation. The precipitate was filtered, and the filtrate diluted with water. The aqueous solution was evaporated to dryness on a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

lyophilizer, and the crude product chromatographed on silica gel with EtOAc/CH_3OH (98/2, v/v). Compound **9** was obtained as white solid product in 98% yield (1.10g). R_f=0.43 in EtOAc/CH_3OH (98/2, v/v). 1H NMR (DMSO-d_6): δ8.84(d, J=5Hz, 1H, H6 of bipyridine), 8.80(m, 1H, H3 of bipyridine), 8.56(dd, J=5Hz, 0.5Hz, 1H, H6′ of bipyridine), 8.25(m, 1H, H3′ of bipyridine), 7.84(dd, J=5Hz, J=1.5Hz, 1H, H5 of bipyridine), 7.32(dd, J=5Hz, 1.5Hz, 1H, H5′ of bipyridine), 2.58(s, 4H, succinimidyl), 2.42(s, 3H, CH_3 of bipyridine). FAB MS: [MH^+]312.

Preparation of 5-[2-(4′-methyl-2,2′-dipyrid-4-yl-carboxamido)ethylthio]-2′-deoxyuridine (5):[95] Compound **9** (1.08g, 3.47mmol) and compound **4** (1.26g, 4.17mmol, 1.2eq) were dissolved in 15ml DMF. The solution was cooled to 0°C in an ice-bath and triethylamine (1.75ml, 3eq) was added. The reaction was carried out at room temperature for 1 hour, and then the reaction mixture cooled again to 0°C, and 50ml water added to the reaction. The aqueous reaction mixture was evaporated to dryness on a lyophilizer. The solid was washed twice with water, and recrystallized from $CH_3OH/CHCl_3$/EtOAc, and dried in air to yield **5** as a white solid (1.26g, 72% yield). R_f=0.40 in CH_3OH/EtOAc (1/1;v/v). ^1H NMR (DMSO-d_6): δ11.64(s, 1H, CONHCO), 9.06(t, J=6Hz, 1H, CONH), 8.83(d, J=5Hz, 1H, H6 of bipyridine), 8.77(s, 1H, H3 of bipyridine), 8.60(d, J=5Hz, 1H, H6′ of bipyridine), 8.28(s, 1H, H3′ of bipyridine), 8.20(s, 1H, H6), 7.81(dd, J=5Hz, J=1.5Hz, 1H, H5 of bipyridine), 7.34(d, J=5Hz, 1H, H5′ of bipyridine), 6.17(t, J=7Hz, 1H, H1′), 5.30(d, J=4Hz, 1H, 3′OH), 5.14(t, J=4Hz, 1H, 5′OH), 4.30(m, 1H, H3′), 3.83(m, 1H, H4′), 3.58-3.68(m, 2H,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

H5′/H5′′), 3.50(m, 2H, NCH_2), 2.93(m, 2H, SCH_2), 2.44(s, 3H, CH_3 of bipyridine), and 2.15-2.24(m, 2H, H2′/H2′′). FAB MS: $[MH^+]$500.

Preparation of 5-[2-(4′-methyl-2,2′-dipyrid-4-yl-carboxamido)ethylthio]-5′-*O*-dimethoxytrityl-2′-deoxyuridine (6):[87] Compound **5** (0.213g, 0.43mmol) was co-evaporated twice with CaH_2 distilled pyridine. In a dry box the dried **5** was added dimethoxytrityl chloride (0.170g, 0.52mmol, 1.2eq) and 4-dimethylaminopyridine (5mg). The reaction flask was removed from the dry box and 2ml freshly distilled pyridine added. The reaction mixture was stirred at room temperature for seventy minutes, at which time TLC showed that the reaction was complete. Water (1ml) was added to the reaction mixture to quench the unreacted reagents. It was then further stirred for five minutes. The solution was diluted with 10ml methylene chloride, and washed with saturated $NaHCO_3/H_2O$ and $NaCl/H_2O$ solutions, and then dried over anhydrous Na_2SO_4. After evaporation of the solvent, the crude product was chromatographed on silica gel eluting with $EtOAc/C_6H_{14}/CH_3OH/Et_3N$ (4/4/1/0.5%, v/v/v/v) to yield a slightly yellowish solid, which was co-evaporated with methylene chloride a few times to remove the residue solvent to yield 0.290g **6** (85% yield). R_f=0.14 in EtOAc/Hexane/Methanol/Et_3N (4/4/1/0.5%, v/v/v/v). ^1H NMR($CDCl_3$): δ8.77 (d, J=5, 1H, H6 of bipyridine), 8.66(s, 1H, H3 of bipyridine), 8.65(d, J=5Hz, 1H, H6′ of bipyridine), 8.29(d, J=2Hz, 1H, H3′ of bipyridine), 8.20(s, 1H, H6), 7.77(dd, J=5Hz, J=2, 1H, H5′ of bipyridine), 7.40(d, J=7Hz, 2Hz, H3/H5 of phenyl), 7.28-7.33(m, 6H, H4 of phenyl, H2/H6 of methoxyphenyl, H5 of bipyridine), 7.22(d, J=7Hz, 2H, H2/H6 of phenyl), 6.82(d, J=9Hz, 4H, H2/H6 of methoxylphenyl), 6.28(t, J=7Hz,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1H, H1′), 4.61(m, 1H, H3′), 4.06(m, 1H, H4′), 3.79(s, 6, methoxyl), 3.66(m, 2H, H5′), 3.38(m, 2H, NH_2), 2.70(m, 1H, H2′ or H2′′), 2.51(m, 1H, H2′ or H2′′), 2.47(s, 3H, methyl of bipyridine), 2.38(m, 2H, SH_2). MS: [MH^+]802.

Preparation of 5-[2-(4′-methyl-2,2′-dipyrid-4-yl-carboxamido)ethylthio]-5′-*O*-dimethoxytrityl-3′-*O*-β-cyanoethyl-N,N-diisopropylphosphoramidyl-2′-deoxyuridine (7):[87] Compound **6** (0.290g, 0.36mmol) was co-evaporated with 10% pyridine methylene chloride. To the dried **6** was added anhydrous methylene chloride (5ml), CaH_2 distilled diisopropylethylamine (378μl, 4eq) and chloro-N,N-diiopropyl-β-cyanoethoxylphosphine (182μl, 0.54mmol, 1.5eq) in a dry box. The reaction was stirred at room temperature for three hours, whereupon TLC indicated that the reaction was complete. The reaction was stopped by adding 2ml methanol, and further stirred for 10 minutes. The reaction mixture was diluted by 15% Et_3N in EtOAc, and then washed with saturated $NaHCO_3/H_2O$ and $NaCl/H_2O$ solutions. The organic layer was dried over anhydrous Na_2SO_4, and the solvents were removed on a rotary evaporator. The crude product was purified on silica gel (40gram) eluting with $EtOAc/CH_3OH/Et_3N$ (85/10/5, v/v/v) to yield a slightly yellowish solid (0.308g, 85% yield). R_f=0.38 in $EtOAc/CH_3OH/Et_3N$ (85/10/5, v/v/v). ^{31}P NMR (C_6D_6): δ 149.9 and 149.5. High resolution FAB MS: [MH^+]1001.3989 (calculated); 1001.4071 (observed).

Preparation of 6-amino-9-(2,5-bis-O-*t*-butyldimethylsilyl-β-D-ribofuranosyl)-purine (11):[156] Dimethylisobutylsilyl chloride (15.5g, 3eq) was slowly added into adenosine (10.3g, 38.6mmol) dissolved in 90ml anhydrous pyridine. The solution turned clear after addition of dimethylisobutylsilyl chloride. The reaction was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

carried out at room temperature for 48 hours under N_2. The reaction solution turned cloudy white as the reaction proceeded. After evaporation of the solvent, a slightly yellowish solid was formed. The solid was redissolved in ethyl acetate, and the solution was washed sequentially with 1M HCl, $NaHCO_3/H_2O$ and $NaCl/H_2O$. After the ethyl acetate was removed on a rotary evaporator, a white solid was obtained. The crude product **11** was purified on silica gel eluting with ethyl acetate/hexane (7:3, v/v) to yield 9.722g **11** (51% yield). R_f=0.57 in ethyl acetate. ^{1}H NMR ($CDCl_3$): δ8.33 (s, 1, H2 or H8), 8.21 (s, 1, H2 or H8), 6.09 (d, 1, H1), 5.81 (broad s, 2, NH_2), 4.62 (t, 1, H2), 4.30 (t, 1, H3′), 4.20(dd, 1, H4′), 4.00(dd, 1, H5′ or H5′′), 3.80(dd, 1, H5′ or H5′′), 0.94 (s, 9, $C(CH_3)_3$), 0.82 (s, 9, $C(CH_3)_3$), 0.14 (s, 3, CH_3), 0.12 (s, 3, CH_3), -0.07 (s, 3, CH_3), -0.14 (s, 3, CH_3). MS: [MH^+]496.

Preparation of 6-amino-9-(2,5-bis-O-*t*-butyldimethylsilyl-3-keto-β-D-ribofuranosyl)-purine (12):[157] The oxidation solution was prepared in advance. Ac_2O (1.26ml, 3eq) was mixed with pyridine (2.17ml, 6eq), and the mixture kept in an ice-bath. CrO_3 (0.536g, 1.2eq) powder was added into the above cold mixture with magnetic stirring. After the CrO_3 dissolved, 23ml CH_2Cl_2 was added to the cold mixture. Adenosine derivative **11** (2.214g, 4.47mmol) was added into the CrO_3 oxidation solution at 0°C. After the oxidation reaction was initiated at 0°C, the reaction mixture was allowed to stand at room temperature for three hours. The final reaction mixture was transferred into 500ml diethyl ether at room temperature to give a brown precipitate. The precipitate was filtered through celite, and the red-brown filtrate washed with saturated $NaHCO_3/H_2O$ and $NaCl/H_2O$ solutions, respectively. The green

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

aqueous layer was separated from the light yellow ether layer, and the separated ether solution dried over anhydrous Na_2SO_4. After evaporation of the solvent, a brown solid was obtained. The crude product was purified on silica gel eluting with ethyl acetate/hexane (7:3, v/v) to yield white flaky product **12** (1.412g, 64% yield). R_f=0.58 in ethyl acetate. ^1H NMR (DMSO-d_6): δ8.34 (s, 1, H2 or H8), 8.13 (s, 1, H2 or H8), 7.39 (broad s, 2, NH_2), 6.17 (d, 1, H1′), 5.26(d, 1, H2′), 4.46 (m, 1, H4′), 3.88 (dd, 1, H5′ or H5′′), 3.85 (dd, 1, H5′ or H5′′), 0.86 (s, 9, $C(CH_3)_3$), 0.66 (s, 9, $C(CH_3)_3$), 0.04 (s, 3, CH_3), 0.00 (s, 3, CH_3), -0.11 (s, 3, CH_3), -0.28 (s, 3, CH_3). MS: [MH^+]494.

Preparation of 6-amino-9-(2-O-*t*-butyldimethylsilyl-3-keto-β-D-ribofuranosyl)-purine (13):[154] Compound **12** (1.224g) was dissolved in 6ml CF_3COOH/H_2O (9:1, v/v) at 0°C in an ice-bath. The reaction was kept at 0°C, and after 60 minutes all the starting material turned into the product **13**. Since the compound **13** is unstable, the reaction mixture was loaded directly onto silica gel for purification with ethyl acetate as eluant. The collected product **13** was used immediately in the next step reaction. The product **13** is a colorless solid. R_f=0.33 in ethyl acetate. MS: [MH^+]380.

Preparation of 6-amino-9-(2-O-*t*-butyldimethylsilyl-3-deuterio-β-D-ribofuranosyl)-purine (14):[154] Deuterium was introduced by reducing **13** with $NaBD_4$ in CH_3COOH at room temperature. $NaBD_4$ (0.159g, 3eq) was slowly added to 12ml cold CH_3COOH (just above its melting point). After the mixture stopped bubbling, compound **13** from the previous reaction was transferred into the above reducing solution. The reaction was carried out at room temperature for four hours. After evaporation of the solvent, the semi-solid crude product was obtained. This product was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

purified on silica gel with ethyl acetate as eluant. R_f=0.18 in ethyl acetate. ^{1}H NMR ($CDCl_3$): δ8.19 (s, 1, H2 or H8), 7.88 (s, 1, H2 or H8), 6.86 (broad, 2, NH_2), 5.64 (d, 1, H1′), 4.65(d, 1, H2′), 4.29 (t, 1, H4′), 4.00 (dd, 1, H5′ or H5″), 3.93 (dd, 1, H5′ or H5″), 0.78 (s, 9, $C(CH_3)_3$), -0.06 (s, 3, CH_3), -0.19 (s, 3, CH_3). MS: [MH^+]383.

Preparation of 6-amino-9-(3-deuterio-β-D-ribofuranosyl)-purine (15):[154] Compound **14** from the previous reaction was dissolved in 45ml THF, and 5ml 1M $Bu_4N^+F^-$/THF (2eq) added to the stirred solution. This reaction mixture turned from light yellow to orange brown. The reaction was carried out for 2.5 hours at room temperature. After the solvent evaporated, the crude product **15** was purified on silica gel eluting with $CH_3OH/CHCl_3$ (2:5, v/v) to yield 0.492g white powder (90% averaged yield for each step from **12**). Compound **15** has the same mobility on TLC as adenosine. R_f=0.30 in $CH_3OH/CHCl_3$ (2:5, v/v). ^{1}H NMR (DMSO): δ8.33 (s, 1, H2 or H8), 8.12 (s, 1, H2 or H8), 7.33 (broad, 2, NH_2), 5.86 (d, 1, H1′), 5.43(d, 1, 2′OH), 5.40 (q, 1, 5′OH), 5.15 (s, 1, 3′OH), 4.59 (t, 1, H2′), 3.94 (t, 1, H4′), 3.66 (m, 1, H5′ or H5″), 3.53 (m, 1, H5′ or H5″). There is a very small visible peak at 4.13ppm, which is at the position of H3′ in adenosine, due most likely to the $NaBD_4$ which contained only 98% deuterium. FAB-MS: [MH^+]253

Preparation of 6-amino-9-[3,5-(tetraisopropyldisiloxa-1,3-diyl)-3-deuterio-β-D-ribofuranosyl]-purine (16):[159] Compound **15** (0.492g, 1.85mmol) was dissolved in anhydrous pyridine (8ml). The solution was magnetically stirred for 10 minutes, and then tetraisopropyldisiloxyl-1,3-dichloride (1.17ml, 2eq) gradually added to the reaction mixture. After 30 minutes a white precipitate appeared. The reaction was stopped at

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

three hours by adding ethyl acetate (40ml) to the reaction mixture. The resulting solution was washed sequentially by 50ml 0.01M HCl/H_2O, 50ml saturated $NaHCO_3/H_2O$ and 50ml saturated $NaCl/H_2O$. Ethyl acetate was removed on a rotary evaporator until about 1ml of a syrup-like residue left. This mixture was purified on silica gel eluting with ethyl acetate to yield 0.942g **16** as a white solid (85% yield). R_f=0.44 in ethyl acetate. ^1H NMR (DMSO): δ8.20 (s, 1, H2 or H8), 8.06 (s, 1, H2 or H8), 7.32 (broad, 2, NH_2), 5.85 (d, 1, H1′), 5.60(d, 1, 2′OH), 4.50 (d, 1, H2′), 4.04 (d, 1, H5′ or H5′′), 3.98 (m, 1, H4′) 3.91 (dd, 1, H5′ or H5′′), 1.02 (m, 28, $CH(CH_3)_2$). MS: [MH^+]599.

Preparation of 6-amino-9-[2-O-phenoxythiocarbonyl-3,5-(tetraisopropyldisiloxa-1,3-diyl)-3-deuterio-β-D-ribofuranosyl]-purine (17):[155] Compound **16** (0.463g, 0.887mmol) was dissolved in anhydrous CH_3CN (9ml) with three equivalents of 4-dimethylaminopuridine (0.293g). Yellow *O*-phenylchlorothionoformate (188.5μL, 1.5eq) was slowly added into the above solution. The reaction was carried out at room temperature for twenty eight hours. During the reaction a fine yellow precipitate gradually formed. The CH_3CN was removed by evaporation, and the crude product was redissolved in 3ml ethyl acetate. The redissolved solution was loaded onto silica gel for purification eluting with ethyl acetate to yield **17** as a flaky white solid (0.406g,. 70% yield). R_f=0.46 in ethyl acetate. ^1H NMR (DMSO): δ8.27 (s, 1, H2 or H8), 8.03 (s, 1, H2 or H8), 7.48 (t, 2, C_6H_5 *o*-H), 7.40 (broad, 2, NH_2), 7.33 (t, 1, C_6H_5 *p*-H),, 7.15 (d, 2, C_6H_5 *m*-H), 6.48 (s, 1, H2′), 6.32 (s, 1, H1′), 4.02-3.95 (m, 3, H4′, H5′, H5′′), 1.04 (m, 28, $CH(CH_3)_2$). MS: [MH^+]647.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Preparation of 6-amino-9-[2-deoxy-3,5-(tetraisopropyldisiloxa-1,3-diyl)-3-deuterio-β-D-ribofuranosyl]-purine (18):[155] Compound **17** (0.406g, 0.617mmol) was dissolved in 32ml benzene. Newly opened Bu_3SnH (340μL, 2eq) was added into the above solution, followed by adding 5mg AIBN as initiator. The reaction system was heated in an oil-bath to 95°C for eighty minutes, and was stopped by evaporation of the benzene. This reaction residue was redissolved in 2ml ethyl acetate and purified on silica gel eluting with ethyl acetate to yield **18** as a white solid (0.273g, 88% yield). R_f=0.22 in ethyl acetate. ^{1}H NMR (DMSO): δ8.19 (s, 1, H2 or H8), 8.03 (s, 1, H2 or H8), 7.23 (broad, 2, NH_2), 6.22 (dd, 1, H1′), 3.88 (m, 2, H5′, H5′′), 3.78 (m, 1, H4′), 2.82 (dd, 1, H2′ or H2′′), 2.56 (q, 1, H2′ or H2′′), 1.04 (m, 28, $CH(CH_3)_2$). MS: $[MH^+]$495.

Preparation of 6-benzoylamino-9-[3,5-(tetraisopropyldisiloxa-1,3-diyl)-2-deoxy-3-deuterio-β-D-ribofuranosyl]-purine (20):[87] Compound **18** (0.273g, 0.55mmol) was dissolved in anhydrous pyridine (6ml) and the stirred solution cooled in an ice-bath. Benzoyl chloride (320μL, 4eq) was gradually added into the cooled reaction solution. This solution turned cloudy, and a white precipitate appeared. Five minutes later, the reaction mixture was removed from the ice-bath and allowed to warm to room temperature. After two hours at room temperature, the reaction mixture was cooled again in an ice-bath. After 5 minutes cooling, 4ml H_2O was added to quench the remaining benzoyl chloride. Fifteen minutes after the addition of water, 1.5ml concentrated NH_3 was added to the reaction mixture ($[NH_3]\approx$2M). After 40 minutes, solvent was removed from the reaction mixture by evaporation on a rotary evaporator.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The crude product was purified on silica gel eluting with EtOAc/C_6H_{14} (7:3, v/v) to yield **20** as a white solid (0.322g, 97% yield). R_f=0.50 in ethyl acetate. ^{1}H NMR (DMSO): δ11.19 (s, 1, NH), 8.66 (s, 1, H2 or H8), 8.55 (s, 1, H2 or H8), 8.03 (d, 2, C_6H_5 *o*-H), 7.62 (t, 1, C_6H_5 *p*-H), 7.54 (t 2, C_6H_5 *m*-H), 6.41 (dd, 1, H1′), 3.92 (d, 2, H5′, H5′′), 3.88 (t, 1, H4′), 2.96 (d, 1, H2′ or H2′′), 2.62(q, 1, H2′ or H2′′), 1.05 (m, 28, $CH(CH_3)_2$). FAB MS: [MH^+]599.

Preparation of 6-benzoylamino-9-(2-deoxy-3-deuterio-β-D-ribofuranosyl)-purine (21):[87] Compound **20** (0.322g, 0.54mmol) was dissolved in 1 THF (9m). Five minutes later, 1M $Bu_4N^+F^-$/THF (1.182ml) was added whereupon the solution turned light yellow. After thirty minutes at room temperature, the solvent was evaporated on a rotary evaporator and the resulting residue purified on silica gel eluting with EtOAc/MeOH (4:1, v/v) to yield 0.166g **21** (86% yield). R_f=0.34 in EtOAc/MeOH (4:1, v/v). The ^{1}H NMR spectrum was consistent with standard 6-benzoylamino-9-(β-D-ribofuranosyl)-purine, except H3′ peak disappeared and H2′/H4′ peaks simplified. ^{1}H NMR (DMSO): δ11.19 (s, 1, NH), 8.75 (s, 1, H2 or H8), 8.66 (s, 1, H2 or H8), 8.03 (d, 2, C_6H_5 *o*-H), 7.62 (t, 1, C_6H_5 *p*-H), 7.54 (t 2, C_6H_5 *m*-H), 6.45 (t, 1, H1′), 5.34 (s, 1, 3′OH), 4.98 (t, 1, 5′OH), 3.84 (t, 1, H4′), 3.88 (t, 1, H4′), 3.62 (q, 1, H5′ or H5′′), 3.53(q, 1, H5′ or H5′′), 2.78 (q, 1, H2′ or H2′′), 2.35 (q, 1, H2′ or H2′′). MS: [MH^+]357.

Preparation of 6-benzoylamino-9-(2-deoxy-3-deuterio-5-dimethoxyltrityl-β-D-ribofuranosyl)-purine (22):[87] Compound **21** (0.166g) was dissolved in 1ml dry pyridine freshly distilled over CaH_2, and dried on an oven-vacuum system. This co-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

evaporation was repeated twice. This dried compound **21** was redissolved in fleshly distilled pyridine. 4-Dimethylaminopyridine (5mg, 5%eq) and dimethoxytrityl chloride (0.171g, 1.2eq) were added to the above solution at room temperature under dry N_2. The reaction solution turned cloudy orange as the reaction proceeded. After two hours, the reaction mixture was quenched by adding 3ml H_2O. The product was extracted into diethyl ether, and the ether solution evaporated to yield a residue that was loaded onto a 2mm chromatotron plate eluting with EtOAc/CH_3OH/C_6H_{14}/Et_3N (4/1/4/1%, v/v/v/v) to yield **22** as a yellowish solid (0.204g,. 74% yield). R_f=0.18 in EtOAc/CH_3OH/C_6H_{14} (4:1:4, v/v/v), plus 1% Et_3N. ^1H NMR (DMSO): δ11.19 (broad, 1, NH), 8.62 (s, 1, H2 or H8), 8.58 (s, 1, H2 or H8), 8.02 (d, 2, C_6H_5 *o*-H), 7.62 (t, 1, C_6H_5 *p*-H), 7.55 (t 2, C_6H_5 *m*-H), 7.32 (d, 2, trityl H), 7.19 (m, 7, trityl H), 6.79 (q, 4, trityl H), 6.49 (t, 1, H1′), 5.40 (s, 1, 3′OH), 4.02 (m, 3, H4′, H5′, H5′′), 3.69 (d, 6, CH_3), 2.78 (q, 1, H2′ or H2′′), 2.35 (q, 1, H2′ or H2′′). FAB MS: [MH^+]659; [Trityl$^+$]303.

Preparation of 6-benzoylamino-9-(2-deoxy-3-deuterio-3-β-cyanoethoxyl-N,N-diisopropylaminophosphoramidyl-5′-dimethoxyltrityl-β-D-ribofuranosyl)-purine (23):[87] Phosphitylation required that all the reagents and vessels be absolutely dry. Compound **22** (0.104g, 0.158mmol) in a 25ml pear-shaped flask was dissolved in 2ml Aldrich Sure/Seal™ CH_2Cl_2 and 1ml freshly distilled pyridine. The solution was evaporated in a vacuum oven overnight to remove water from the starting material. The newly dried compound **22** was redissolved in 0.8ml Aldrich Sure/Seal™ CH_2Cl_2 via an oven dried glass syringe. This solution was transferred into a dry box (N_2 atmosphere) for addition of reagents. Freshly distilled diisopropylethylamine (110μL, 4eq) was

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

added to the above solution. Newly purchased β-cyanoethyl-N,N-diisopropylchlorophosphoramidite (38.8μL) was gradually added to the reaction solution. The reaction was complete after 60 minutes at room temperature. CH_3OH (0.5ml) was added and the reaction mixture stirred for 5 minutes, and then diluted with 10ml CH_2Cl_2. The resulting solution was sequentially washed with 10% $NaHCO_3/H_2O$ and saturated $NaCl/H_2O$. The CH_2Cl_2 layer was evaporated on a rotary evaporator, and the residue purified on a 1mm chromatotron plate eluting with C_6H_{14}/EtOAc/CH_2Cl/Et_3N (10/4.5/4.5/1, v/v/v/v) to yield 0.130g **N** (96% yield). R_f=0.26 and 0.20 in n-C_6H_{14}/EtOAc/CH_2Cl/Et_3N (10:4.5:4.5:1, v/v/v/v). High resolution FAB-MS: [MH^+]859.3798 (experimental); [MH^+]859.3807 (theoretical).

5′-^{33}P-end labeling of oligonucleotides: The general method for end labeling oligonucleotides was to incubate 20pmol oligonucleotide, 3.5μL T4 PNK buffer (supplied by NewEngland Biolabs), 20-30pmol [γ-^{33}P]ATP, and 15 units T4 PNK in a volume of 35μL for 45 minutes at 37°C. Following incubation, the reaction was quenched by heating to 70°C for 5 minutes, then diluted to 50μL. The labeled oligonucleotide was purified on a ProbeQuant™ G50 micro column (supplied by Pharmacia Biotech) which had been washed with 2 x 200μL of water to remove the storage buffer. The eluent from the column was diluted to 1mL and a 10μL aliquot taken for scintillation counting.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER 5 POTENTIAL APPLICATIONS OF A BIPYRIDINE-TETHERED CHEMICAL NUCLEASE AS A TOOL IN MOLECULAR BIOLOGY

5.1 Introduction

One of our goals in developing chemical nucleases is to utilize them as DNA recombination tools complementary to natural restriction endonucleases. Up to now there have been no reports on the use of chemical nucleases for DNA recombination. This is probably because all chemical nucleases reported up to this time are limited by either low cleavage specificity or low efficiency. As discussed in chapter 4, the major-groove-positioned bipyridine modified oligonucleotide can achieve cleavage yields up to 80% over a three nucleotide site. This level of specificity and efficiency is sufficient to justify exploring potential molecular biology applications. Our preliminary experiments will focus on cleaving large DNA with the bipyridine modified oligonucleotide and then re-ligating the cleaved products enzymatically with ligases.

5.2 Results and discussion

Redox based chemical nucleases cleave substrate DNA into smaller fragments with phosphates on the 3′ and 5′-ends (Figure 3.7, Figure 3.8, Figure 3.9). In order to be used as a ligase substrate, the 3′-phosphate has to be converted into a 3′-hydroxyl.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Uhlenbeck and co-worker[180] reported that T4 polynucleotide kinase (T4 PNK) can remove phosphate from the 3′-end of a polynucleotide. Under optimized conditions, we were able to remove the 3′ phosphate from the redox-based cleavage products, while keep the 5′-phosphate intact (Figure 5.1). This figure shows that the cleavage products after T4 PNK treatment moved slower on a polyacrylamide gel than the untreated oligonucleotide. This indicates that the negatively charged 3′-end phosphate was removed. On the other hand, the intensity of the 5′-^{33}P-labeled parent oligonucleotide band (slowest moving band) is only slightly lightened, but not eliminated. This suggests that under this reaction conditions the T4 PNK did not significantly remove the 5′-phosphate, if it did at all. The reduction in intensity could come from the T4 phosphatase activity which in conjunction with the kinase activity catalyzes phosphate exchange between the 5′-^{33}P and the unlabeled phosphate in the solution.

We have shown in chapter 4 that the oligonucleotide-bipyridine conjugate has high cleavage specificity and efficiency when the substrate is single-stranded DNA. In order to explore applications for this conjugate to engineer double-stranded DNA, it would be necessary to assess this conjugate's nuclease activity toward double-stranded DNA (dsDNA). The pBluescript™ (Stratagene Ltd.) plasmid was selected as the target DNA for oligonucleotide-bipyridine conjugate assay. The advantage in using a circular plasmid for nuclease activity study is that nicked plasmid (form II), linearized plasmid (form III) and intact plasmid (form I) can be easily distinguished on agarose gel due to their different mobilities. Two targeting bipyridine modified oligonucleotides were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1 2

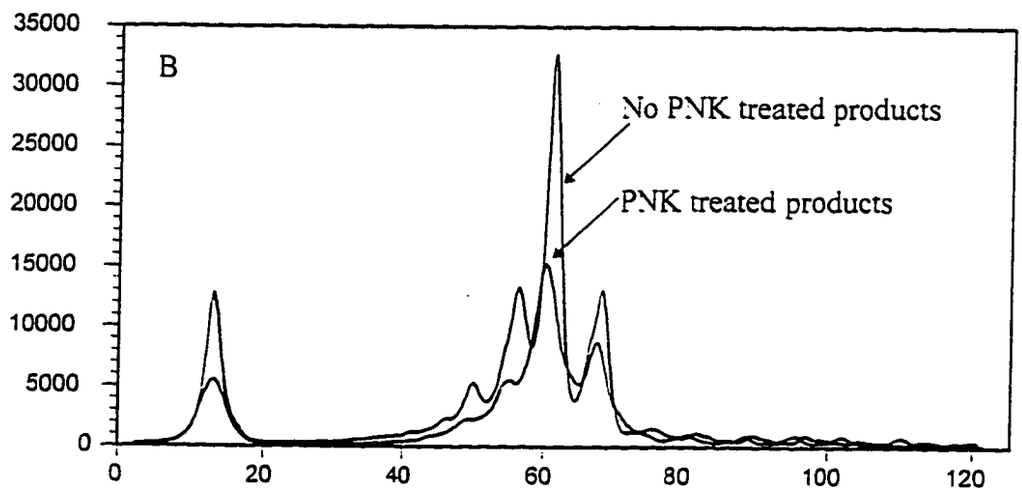


Figure 5.1 T4 kinase activity towards an oligonucleotide 3´ phosphate. Panel A: polyacrylamide gel analysis, lane 1 : products treated by kinase, lane 2: cleavage products directly loaded onto gel, without kinase treatment; panel B: both lanes quantitated by Molecular Analyst™.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

made, 15-mer **JC31** and 17-mer **JC33** (sequences shown in table 4.1). The targeting sequences are located within the gene marker of pBluescript™ plasmid.

The initial cleavage test showed that there was a substantial amount of nicked plasmid in the control lane, which included the cleavage reagents, 50mM TRIS, 50mM NaCl, 10μM $Fe(NH_4)_2(SO_4)_2$ and 0.4mM H_2O_2 at pH 7.4, except for the bipyridine-oligonucleotide conjugate (Figure 5.2). The high background cleavage is probably caused by the cumulate effect of non-specific background cleavage over the 3000 bp dsDNA. The background cleavage needs to be minimized in order to visualize the cleavage resulting from the targeting conjugate. All attempts to minimize the background cleavage by radical scavengers, such as DMSO, propanol, butanol and CT DNA, did not noticeably reduce the background.

In order to circumvent the problem of non-specific cleavage, we decided to cut a fragment out of the plasmid and use the fragment as substrate. The Pvu II restriction enzyme was used to cut at the only two recognition sites of this enzyme on the circular pBluescript™ plasmid to yield two fragments. One of the two fragments had 448 bp, which contained the target sequence. A cleavage reaction was carried out with this relatively short dsDNA substrate (Figure 5.3). As shown on the PAGE gel (Figure 5.3), at a 500 and 5000 ratio of the conjugate to substrate, the cleavage reaction bands all line up with the control band, which indicates that there was virtually no substrate cleaved at all. Two kinds of annealing process for the targeting strand binding to one of the 448 polynucleotide strand were tried. One was to heat the sample to 90°C for ten minutes and let it cool at room temperature, and alternatively to heat the sample to 90°C for ten

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

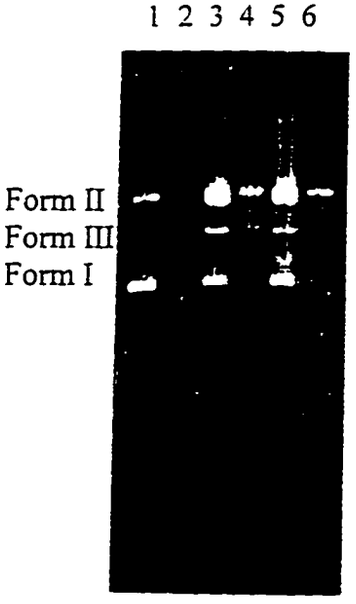


Figure 5.2 Oligonucleotide-bipyridine conjugate targeting pBluescript™ plasmid substrate. Lane 1.: plasmid only; lane 2: NaOH denatured plasmid; lane 3: plasmid treated with cleavage reagents except targeting oligonucleotide; lane 4: as lane 3, but plasmid was NaOH denatured: lane 5: as lane 3, but in the presence of 100 fold targeting oligonucleotide **JC31**; lane 6: as lane 4, but in the presence of 100 fold targeting oligonucleotide **JC31**.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

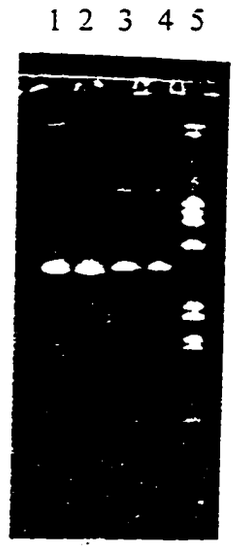


Figure 5.3 Oligonucleotide-bipyridine conjugate targeting a 448 bp Pvu II restriction fragment from pBluescript™ plasmid. Substrate was heat denatured, and annealed at room temperature for 30 minutes. The ratio of targeting strand and substrate strand varied. Lane 1: 5,000 fold targeting oligonucleotide **JC31**; lane 2: 500 fold **JC31**; lane 3: 50 fold **JC31**; lane 4: 448 bp fragment only; lane 5: ØX174 RF DNA Hae III digest markers, 1,353 bp, 1,078 bp, 872 bp, 603 bp; 310 bp; 281 bp; 271 bp; 234 bp; 194 bp; 118 bp and 72 bp.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

minutes and then dip the sample immediately into an ice-bath, and then warm to room temperature for thirty minutes. These two annealing processes gave the same result, no cleavage. The secondary and/or tertiary structure of the substrate strand might be responsible for preventing the short targeting oligonucleotide from binding to the target.

After the second temporary failure, we turned to large circular single-stranded DNA (ssDNA). The selected circular ssDNA is one of the strands of the pBluescript plasmid (plus strand), which contains a sequence complementary to the oligonucleotide-bipyridine conjugate. The ssDNA was obtained using standard molecular biology protocols.[181,182] The nuclease activity assay of the bipyridine conjugate with this ssDNA is still in process and no result is available.

The general scheme for using bipyridine tethered oligonucleotides as chemical nucleases for engineering plasmid sequences is outlined in Figure 5.4, and it is explained briefly as follows. Once the circular plasmid is cleaved by a bipyridine modified oligonucleotide at three sites, the linearized substrate is one nucleotide unit shorter, and has a 3′-phosphate (Figure 3.7, Figure 3.8, Figure 3.9). The 3′-phosphate of the cleaved fragment can be converted into a 3′-hydroxyl by T4 PNK (Figure 5.1). Among the three cleaved fragments, the desired fragment may be selected out with a complementary primer by DNA duplex formation. The nick in this primer/fragment duplex can be sealed by ligase, and at this stage the DNA is re-circularized to a circular DNA containing one less nucleotide. This engineered DNA can be transformed into *Ecoli* for amplification. If the bipyridine modified oligonucleotide targets a site within in a plasmid selection gene region, the engineered DNA will contain a disrupted

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

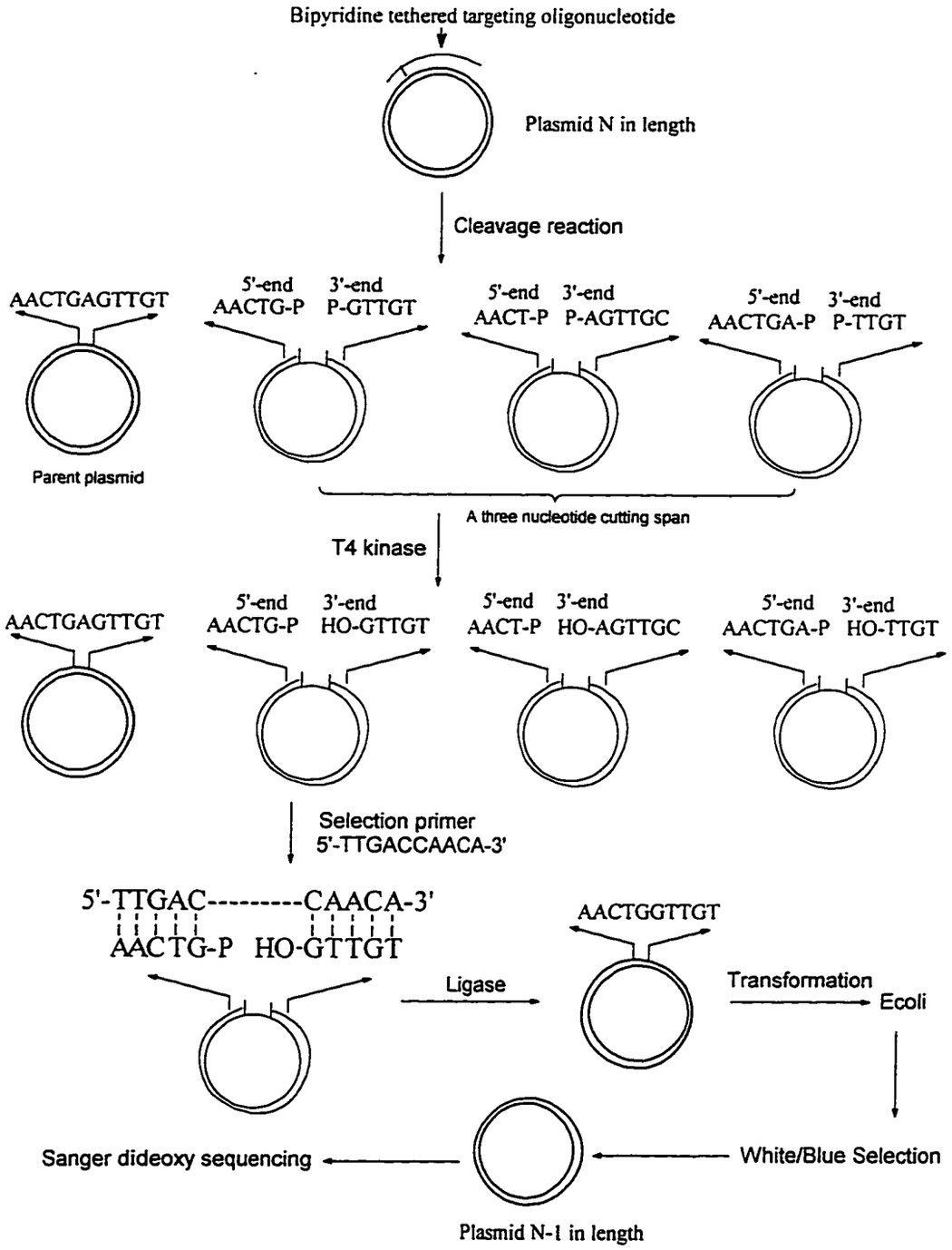


Figure 5.4 Scheme for using bipyridine tethered oligonucleotides as chemical nucleases for engineering plasmid sequences.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

selection gene which can be easily selected. Finally, the sequence of the resulting circular DNA can be confirmed by Sanger dideoxy sequencing.

5.3 Experiments

pBluescript II SK and *Ecoli* strain were purchased from Stratagene Ltd. Restriction enzymes and T4 polynucleotide kinase (T4 PNK) were purchased from New England Biolabs. The manufactures protocols were followed when using these biological reagents. The protocol used for removing the 3′ phosphate with T4 PNK is as follows.

Upon the completion of a cleavage reaction, the reaction mixture (2×0.9 pmol labeled substrate) was divided into two portions. One portion was quenched by EDTA; the second portion was dried on a speed-vac, and the residue was re-dissolved in 10 μL reaction buffer and then 1.5 μL T4 PNK added. The dephosphorylation reaction was carried out at room temperature for one hour and quenched by EDTA. Finally, the reaction solvent was removed on a speed-vac and the residue re-dissolved in loading buffer (formamide:10TBE, 9:1, v/v) and analyzed on a polyacrylamide gel in parallel with the reaction product which had not been treated with kinase. The reaction buffer was prepared as a 5 fold stock solution. Imidazole hydrochloric acid salt (3.40g) and magnesium chloride (0.48g) were dissolved in 100 ml sterilized water, and the pH of the solution adjusted by hydrochloric acid to 5.7. 2-Mercaptoethanol was added into the reaction system just before the dephosphorylation reaction.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

All other reagents, protocols and equipment used are the same as those in chapter 4.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF REFERENCES

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF REFERENCES

1. Bax, A. & Freeman, R. *J. Am. Chem. Soc.* **104**, 1099 (1982).

2. Wuthrich, K. *NMR of Proteins and Nucleic Acids* (Willey, New York, 1986).

3. Ernst, R. R. Bodenhausen, G. & Wokaun, A. *Principle of NMR in One and Two Dimensions* (Clarendon Press, Oxford, 1987).

4. Derome, A. E. *Moderm NMR Techniques for Chemistry Research* (Pergamon Press, Oxford, 1987).

5. Jeener, T. L. *J. Chem. Phys.* **71**, 4546 (1979).

6. Keesler, H., Griesinger, C., Lautx, J., Muller, A., van Gunsteren, W. F. & Berendsen, H. C. J. *J. Am. Chem. Soc.* **110**, 3394 (1988).

7. Wagner, G. *Q. Rev. Biophys.* **16**, 1 (1993).

8. Clore, G. M. & Gronenborn, A. M. *J. Magn. Reson.* **84**, 398 (1989).

9. Yip, P. & Case, D. *J. Magn. Reson.* **83**, 643 (1989).

10. Holak, T. A., Habazetti, J., Oschkinat, H. & Otlewiski, J. *J. Am. Chem. Soc.* **113**, 3196 (1991).

11. Rawn, J. D. *Biochemistry* (Neil Patterson publishers, 1989).

12. Watson, J. & Crick, F. *Nature* **171**, 737 (1953).

13. Watson, J. & Crick, F. *Nature* **171**, 964 (1953).

14. Saenger, W. *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).

15. Sarma, R. H. *Nucleic Acid Geometry and Dynamics* (Pergamon Press, 1980).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

16. Keepers, J. & James, T. *J. Magn. Reson.* **57**, 404 (1984).

17. Bonvin, A. M., Boelens, R. & Kaptein, R. *J. Magn. Reson.* **95**, 626 (1991).

18. Holak, T. A., Berstein, R., Ross, A. & Cieslar, C. *J. Magn. Reson.* **B101**, 185 (1993).

19. Keesler, H., Seip, S. & Saulities, J. *J. Biomol. NMR* **1**, 83 (1991).

20. Habazetti, J. R., Oschkinat, H. & Holak, T. A. *J. Magn. Reson.* **97**, 511 (1992).

21. Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H. & Kaptein, R. *J. Magn. Reson.* **82**, 290 (1989).

22. Zhang, Q., Chen, J., Gozansky, E., Zhu, F., Jackson, P. & Gorenstein, D. *J. Magn. Reson.* **B106**, 164 (1995).

23. Olejniczak, E. T., Gampe, R. T.,Jr. & Fesik, S. W. *J. Magn. Reson.* **67**, 28 (1986).

24. Boelens, R., Koning, T. M. G. & Kaptein, R. *J. Mol. Struct.* **173**, 299 (1988).

25. Borgias, B. A. & James, T. *J. Magn. Reson.* **87**, 475 (1990).

26. Nikonowicz, E., Meadows, R. & Gorenstein, D. *Biochemistry* **29**, 4193 (1990).

27. Nikonowicz, E., Meadows, R. & Gorenstein, D. *Bull. Magn. Reson.* **11**, 226 (1989).

28. Post, C. B., Meadows, R. & Gorenstein, D. *J. Am. Chem. Soc.* **112**, 6796 (1990).

29. Weiner, P. K. & Kollman, P. A. *J. Comput. Chem.* **2**, 287 (1981).

30. Gorenstein, D. Meadows, R. Met, J. T. Nikonowicz, E. & Post, C. B. *Advances in Biophysical Chemistry* (JAI Press, Greenwich, Conn. 1990).

31. Donne, D. G., Gozansky, E. & Gorenstein, D. *J. Magn. Reson.* **106**, 156 (1994).

32. Habazetti, J., Schleicher, J., Oschkinat, H. & Holak, T. A. *J. Mol. Bol.* **228**, 156 (1992).

33. Holak, T. A., Habazetti, J., Oschkinat, H. & Otlewiski, J. *J. Am. Chem. Soc.* **113**, 3196 (1991).

34. Havel, T. F., Kuntz, I. D. & Crippen, G. M. *Bull. Math. Biol.* **45**, 665 (1983).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

35. Braun, W. & Go, N. *J. Mol. Bol.* **186**, 613 (1983).

36. Nilges, M., Gronenborn, A. M., Brunger, A. T. & Clore, G. M. *Protein Eng.* **2**, 27 (1988).

37. Zuiderweg, E., Scheek, R., Boelens, R., Gunsteren, R. v. & Kaptein, R. *Biochemie* **67**, 707 (1985).

38. Macura, S. & Ernst, R. R. *Mol. Phys.* **41**, 95 (1980).

39. Kaluarachchi, K., Meadows, R. & Gorenstein, D. *Biochemistry* **30**, 8785 (1991).

40. Zhang, Q. 1994).(UnPub)

41. Hills, M. (1998). Chemistry Department, Purdue University.

42. Zhu, F. (1995). Chemistry Department, Purdue University. Ph.D.

43. Thomas, P. D., Basus, V. J. & James, T. *Proc. Natl. Acad. Sci. USA* **88**, 1237 (1991).

44. Wuthrich, K. *Science* **243**, 45 (1989).

45. Grzesiek, S., Dobeli, R., Gentz, G., Garotta, G., Labhardt, A. M. & Bax, A. *Biochemistry* **31**, 8180 (1992).

46. Kay, L. E., Clore, G. M. & Gronenborn, A. M. *Science* **249**, 411 (1990).

47. Zuiderweg, E. R. P., Petros, A. M., Fesik, S. W. & Olejniczak, E. T. *J. Am. Chem. Soc.* **113**, 370 (1991).

48. Clore, G. M., Wingfield, P. T. & Gronenborn, A. M. *Biochemistry* **30**, 2315 (1991).

49. Pardi, A., Hare, D. R. & Wang, C. *Proc. Natl. Acad. Sci. USA* **85**, 8785 (1988).

50. Brunger, A. T., Clore, G. M., Gronenborn, A. M., Saffrich, R. & Nilges, M. *Science* **261**, 328 (1993).

51. Bergstrom, D. E. & Gerry, N. P. *J. Am. Chem. Soc.* **116**, 12067 (1994).

52. Boidot-Forget, M., Chassignol, M., Takasugi, M., Thuong, N. T. & Helene, C. *Gene* **72**, 361 (1988).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

53. Boutorin, A. S., Vlassov, V. V., Kazakov, S. A., Kutiavin, I. V. & Podyminogin, M. A. *FEBS Lett.* **172**, 43 (1984).

54. Chen, C.-H. B. & Sigman, D. S. *Proc. Natl. Acad. Sci. USA* **83**, 7147 (1986).

55. Chen, C.-H. B., Gorin, M. B. & Sigman, D. S. *Proc. Natl. Acad. Sci. USA* **90**, 4206 (1993).

56. Chu, B. C. F. & Orgel, L. E. *Proc. Natl. Acad. Sci. USA* **82**, 963 (1985).

57. Dreyer, G. & Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **82**, 968 (1985).

58. Francois, J.-C., Saison-Behmoaras, T., Chassognol, M., Thuong, N. T. & Helene, C. *J. Biol. Chem.* **264**, 5891 (1989).

59. Frolova, E. I., Fedorova, O. S. & Knorre, D. G. *Biochimie* **75**, 5 (1993).

60. Groves, J. T. & Kady, I. O. *Inorg. Chem.* **32**, 3868 (1993).

61. Kady, I. O. & Groves, J. T. *Bioorg. Med. Chem. Lett.* **3**, 1367 (1993).

62. Kuroda, R., Satoh, H., Shinomiya, M., Watanabe, T. & Otsuka, M. *Nucleic Acids Research* **23**, 1524 (1995).

63. Le Doan, T., Perrouault, L., Helene, C., Chassignol, M. & Thuong, N. T. *Biochemistry* **25**, 6736 (1986).

64. Magda, D., Wright, M., Miller, R. A., Sessler, J. L. & Sansom, P. I. *J. Am. Chem. Soc.* **117**, 3629 (1995).

65. Oikawa, S., Kurasaki, M., Kojima, Y. & Kawanish, S. *Biochemistry* **34**, 8763 (1995).

66. Pitie, M., Casas, C., Lacey, C. J., Pratviel, G., Bernadou, J. & Meunier, B. *Angew. Chem. Int. Engl.* **32**, 557 (1993).

67. Sergeyev, D. S., Godovikova, T. S. & Zarytova, V. F. *FEBS Lett.* **280**, 271 (1991).

68. Modak, A. S., Gard, J. K., Merriman, M. C., Winkeler, K. A., Bashkin, J. K. & Stern, M. K. *J. Am. Chem. Soc.* **113**, 283 (1991).

69. Jenkins, L. A., Bashkin, J. K. & Autry, M. E. *J. Am. Chem. Soc.* **118**, 6822 (1996).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

70. Mandal, S. S., Kumar, N. V., Varshney, U. & Bhattacharya, S. *Journal of Inorganic Biochemistry* **63**, 265 (1996).

71. Rammo, J., Hettich, R., Roigk, A. & Schneider, H. J. *Chemical Communications* 105 (1996).

72. Strekas, T. C., Baker, A. D. & Harripersadmorgan, O. *Journal of Coordination Chemistry* **34**, 77 (1995).

73. Sigman, D. S., Mazumder, A. & Perrin, D. *Chem. Rev.* **93**, 2295 (1993).

74. Grover, N., Ciftan, S. A. & Thorp, H. H. *Inorganica Chimica Acta* **240**, 335 (1995).

75. Singh, U. S., Scannell, R. T., An, H. Y., Carter, B. J. & Hecht, S. M. *J. Am. Chem. Soc.* **117**, 12691 (1995).

76. Yashiro, M., Ishikubo, A. & Komiyama, M. *Journal of Biochemistry* **120**, 1067 (1996).

77. Routier, S., Bernier, J. L., Catteau, J. P. & Bailly, C. *Bioorg. Med. Chem. Lett.* **7**, 63 (1997).

78. Hegg, E. L. & Burstyn, J. N. *Inorg. Chem.* **35**, 7474 (1996).

79. Rammo, J. & Schneider, H. J. *Inorganica Chimica Acta* **251**, 125 (1996).

80. Ly, D., Kan, Y. Z., Armitage, B. & Schuster, G. B. *J. Am. Chem. Soc.* **118**, 8747 (1996).

81. Gallagher, J., Chen, C. B., Pan, C. Q., Perrin, D. M., Cho, Y. M. & Sigman, D. S. *Bioconjugate Chem.* **7**, 413 (1996).

82. Husken, D., Goodall, G., Blommers, M. J. J., et al. *Biochemistry* **35**, 16591 (1996).

83. Dowjat, W. K., Kharatishvili, M. & Costa, M. *Biometals* **9**, 327 (1996).

84. Gerry, N. P. Precision Sequence-specific Cleavage of Nucleic Acids by Minor-groove Directed Metal Binding Ligands Linked through N-2 of Deoxyguanasine. (1996). Purdue University. Ph.D.

85. Kim, Y. G., Cha, J. & Chandrasegaran, S. *Proc. Natl. Acad. Sci. USA* **93**, 1156 (1996).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

86. Roberts, R. J. & Macelis, D. *Nucleic Acids Research* **24**, (1):223 (1996).

87. Gait, M. J. *Oligonucleotide Synthesis a Practical Approach* (IRL PRESS, Oxford, 1985).

88. Beaucage, S. L. & Iyer, R. P. *Tetrahedron* **48**, (12):2223 (1992).

89. Beaucage, S. L. & Iyer, R. P. *Tetrahedron* **49**, (10):1925 (1993).

90. Beaucage, S. L. & Iyer, R. P. *Tetrahedron* **49**, (28):6123 (1993).

91. Dale, R. M. K., Livingston, D. C. & Ward, D. C. *Proc. Natl. Acad. Sci. USA* **70**, 2238 (1973).

92. Dale, R. M. K., Martin, E., Livingston, D. C. & Ward, D. C. *Biochemistry* **14**, 2447 (1975).

93. Bergstrom, D. E., Beal, P., Jenson, J. & Lin, X. *J. Org. Chem.* **56**, 5598 (1991).

94. Bergstrom, D. E., Lin, X., Wang, G., et al. *Synlett* 179 (1992).

95. Wang, G. & Bergstrom, D. E. *Tetrahedron Letters* **34**, 6721 (1993).

96. Ruth, J. & Bergstrom, D. E. *J. Org. Chem.* **43**, 2870 (1978).

97. Bergstrom, D. E. & Ruth, J. *J. Am. Chem. Soc.* **98**, (6):1587 (1976).

98. Bergstrom, D. E. & Ogawa, M. K. *J. Am. Chem. Soc.* **100**, 8106 (1978).

99. Bergstrom, D. E., Ruth, J. & Warwick, P. *J. Org. Chem.* **46**, 1432 (1981).

100. Bergstrom, D. E., Beal, P., Hussain, A., Land, R. & Jenson, J. *J. Am. Chem. Soc.* **111**, 374 (1989).

101. Bergstrom, D. E. & Ruth, J. *J. Carbohydrates Nucleosides Nucleotides* **4**, (5):257 (1977).

102. Boutorine, A. S., Brault, D., Takasugi, M., Delgado, O. & Helene, C. *J. Am. Chem. Soc.* **118**, 9469 (1996).

103. Hall, J., Husken, D. & Haner, R. *Nucleic Acids Research* **24**, 3522 (1996).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

104. Bigey, P., Pratviel, G. & Meunier, B. *Nucleic Acids Research* **23**, (19):3894 (1995).

105. Moser, H. E. & Dervan, P. B. *Science* **238**, 645 (1987).

106. Strobel, S. A. & Dervan, P. B. *Science* **249**, 73 (1990).

107. Durackova, Z., Labudova, O., Andrezalova, L., et al. *International Journal of Biochemistry & Cell Biology* **27**, 1341 (1995).

108. Chen, C.-H. B. & Sigman, D. S. *J. Am. Chem. Soc.* **110**, 6570 (1988).

109. Sigman, D. S. *Acc. Chem. Res.* **19**, 180 (1986).

110. Morrow, J. R. & Trogler, W. C. *Inorg. Chem.* **28**, 2330 (1989).

111. Yamamoto, K. & Kawanishi, S. *J. Biol. Chem.* **264**, 15435 (1989).

112. John, D. C. A. & Douglas, K. T. *Transition Metal Chemistry* **21**, 460 (1996).

113. Frey, S. T., Sun, H. H. J., Murthy, N. N. & Karlin, K. D. *Inorganica Chimica Acta* **242**, 329 (1996).

114. Silver, G. C. & Trogler, W. C. *J. Am. Chem. Soc.* **117**, 3983 (1995).

115. Groves, J. T. & Marla, S. S. *J. Am. Chem. Soc.* **117**, 9578 (1995).

116. Goll, J. G. & Thorp, H. H. *Inorganica Chimica Acta* **242**, 219 (1996).

117. Hertzberg, R. P. & Dervan, P. B. *J. Am. Chem. Soc.* **104**, 313 (1982).

118. Pogozelski, W. K., McNeese, T. J. & Tullius, T. D. *J. Am. Chem. Soc.* **117**, 6428 (1995).

119. Cases, C., Lacey, C. J. & Meunier, B. *Bioconjugate Chem.* **4**, 366 (1993).

120. Pitie, M., Pratviel, G., Bernadou, J. & Meunier, B. *Proc. Natl. Acad. Sci. USA* **89**, 3967 (1992).

121. Pratviel, G., Pitte, M., Bernadou, J. & Meunier, B. *Nucleic Acids Research* **19**, 6283 (1991).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

122. Pratviel, G., Duarte, V., Bernadou, J. & Meunier, B. *J. Am. Chem. Soc.* **115**, 7939 (1993).

123. Lown, J. W. & Joshua, A. V. *J. Chem. Soc. Chem. Commun.* 1298 (1982).

124. Lown, J. W., Sondhi, S. M. & Ong, C.-W. *Biochemistry* **25**, (18):5111 (1986).

125. Hashimoto, Y., Iijima, H., Nozaki, Y. & Shudo, K. *Biochemistry* **25**, (18):5103 (1986).

126. Dervan, P. B. *Science* **232**, 464 (1986).

127. Hertzberg, R. P. & Dervan, P. B. *Biochemistry* **23**, 3934 (1984).

128. Iverson, B. L. & Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **85**, 4615 (1988).

129. Schultz, P. G., Taylor, J. S. & Dervan, P. B. *J. Am. Chem. Soc.* **104**, 313 (1982).

130. Taylor, J. S., Schultz, P. G. & Dervan, P. B. *Tetrahedron* **40**, 457 (1984).

131. Schultz, P. G. & Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **80**, 6834 (1983).

132. Schultz, P. G. & Dervan, P. B. *J. Biomol. Struct. Dyn.* **1**, 1133 (1984).

133. Strobel, S. A., Moser, H. E. & Dervan, P. B. *J. Am. Chem. Soc.* **110**, 7927 (1988).

134. D'Aurora, V., Stern, A. M. & Sigman, D. S. *Biochem. Biophys. Res. Commun.* **78**, (1):170 (1977).

135. D'Aurora, V., Stern, A. M. & Sigman, D. S. *Biochem. Biophys. Res. Commun.* **80**, (4):1025 (1978).

136. Sun, J.-S., Francois, J.-C., Lavery, J.-C. & Helene, C. *Biochemistry* **27**, (16):6039 (1988).

137. Thederahn, T. B., Kuwabara, M. D., Larsen, T. A. & Sigman, D. S. *J. Am. Chem. Soc.* **111**, (13):4941 (1989).

138. Le Doan, T., Perrouault, L., Chassignol, M. & Helene, C. *Nucleic Acids Research* **15**, (21):8643 (1987).

139. Boutorin, A. S., Le Doan, T., Battioni, J. P. & Helene, C. *Bioconjugate Chem.* **1**, (5):350 (1990).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

140. Frolova, E. I., Ivanova, E. M. & Zarytova, V. F. *FEBS Lett.* **269**, (1):101 (1990).

141. Bergstrom, D. E. & Chen, J. *Bioorg. Med. Chem. Lett.* **6**, (18):2211 (1996).

142. Stubbe, J. & Kozarich, J. W. *Chem. Rev.* **87**, 1107 (1987).

143. Pratviel, G., Pitie, M., Bernadou, J. & Meunier, B. *Angew. Chem. Int. Engl.* **30**, (6):702 (1991).

144. Goyne, T. E. & Sigman, D. S. *J. Am. Chem. Soc.* **109**, (9):2846 (1987).

145. McCall, G. H., Rabow, L. E., Ahsley, G. W. & et al. *J. Am. Chem. Soc.* **114**, (13):4958 (1992).

146. Burger, R. M., Drlica, K. & Birdsall, B. *J. Biol. Chem.* **269**, (42):25978 (1994).

147. Wu, J. C., Kozarich, J. W. & Stubbe, J. *J. Biol. Chem.* **258**, (8):4694 (1983).

148. Sugiyama, H., Xu, C., Murugesan, N. & et al. *Biochemistry* **27**, (1):58 (1988).

149. Cook, G. P. & Greenberg, M. M. *J. Am. Chem. Soc.* **118**, 10025 (1996).

150. Sitlani, A., Long, E. C., M., P. A. & Barton, J. K. *J. Am. Chem. Soc.* **114**, 2303 (1992).

151. Ahmadian, M. *Results of the thermal studies of oligonucleotides containing C-5 modified deoxyuridines* (1996).(UnPub)

152. Wang, G. & Bergstrom, D. E. *Synlett* 422 (1992).

153. Miaskiewicz, K. & Osman, R. *J. Am. Chem. Soc.* **116**, 232 (1994).

154. Robins, M. R., Samano, V. & Johnson, M. D. *J. Org. Chem.* **55**, 410 (1990).

155. Hansske, F., Wilosn, J. S. & Robins, M. R. *J. Am. Chem. Soc.* **105**, 4059 (1983).

156. Ogilvie, K. K., Beaucage, S. L., Schifman, A. L., Theriault, N. Y. & Sadana, K. L. *Can. J. Chem.* **56**, 2768 (1978).

157. Hansske, F. & Robins, M. R. *Tetrahedron Letters* **24**, 1589 (1983).

158. Hansske, F., Madej, D. & Robins, M. R. *Tetrahedron* **40**, (1):125 (1984).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

159. Markiewicz, W. T. *J. Chem. Res. (M)* 178 (1979).

160. Titze, C. & Kaim, W. *Zeitschrift fur Naturforschung Section B-A Journal of Chemical Sciences* **51**, 981 (1996).

161. Luo, Y. & Henle, E. S. *METHODS IN ENZYMOLOGY Vol. 234* (Academic Press, Inc. 1994).

162. Halliwell, B. & Gutteridge, M. C. *METHODS IN ENZYMOLOGY VOL. 186* (Academic Press, Inc. 1990).

163. Halliwell, B. & Gutteridge, J. M. *FEBS Lett.* **128**, 343 (1981).

164. Harris, D. C. & Aisen, P. *Biochim. Biophys. Acta* **329**, 156 (1973).

165. Grootveld, M. & Halliwell, B. *Free Radical Res. Commun.* **1**, 243 (1986).

166. Koppenol, W. H. & Liebman, J. F. *J. Phys. Chem.* **88**, 99 (1984).

167. Kozarich, J. W., Worth, L. J., Frank, B. L., Christner, D. F., Vanderwall, D. E. & Stubbe, J. *Science* **245**, 1396 (1989).

168. Kappen, L. S., Goldberg, I. H., Wu, S. H., Stubbe, J., Worth, L. J. & Kozarich, J. W. *J. Am. Chem. Soc.* **112**, 2797 (1990).

169. Sugiyama, H., Fujimoto, K. & Saito, I. *Tetrahedron Letters* **37**, (11):1805 (1996).

170. Lippard, S. J. *Acc. Chem. Res.* 211 (1978).

171. Lippard, S. J. *Biochemistry* **15**, 4330 (1976).

172. Lippard, S. J., Bond, P. T., Wu, K. C. & Bauer, W. R. *Science* **194**, 726 (1976).

173. Repine, J. E., Pfenninger, O. W. & Talmage, D. W. *Proc. Natl. Acad. Sci. USA* **78**, 1001 (1981).

174. Shepherd, R. E., Lomis, T. J. & Loepsel, R. R. *J. Chem. Soc. Chem. Commun.* 222 (1992).

175. Wink, D. A., Nims, R. W., Saavedra, J. E., Utermahlen, W. E.,Jr. & Ford, P. C. *Proc. Natl. Acad. Sci. USA* **91**, 6604 (1994).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

176. Dean, J. A. *Lange's Handbook of Chemistry* (McGraw-Hill, 1992).

177. Still, W. C., Kahn, M. & Mitra, A. *J. Org. Chem.* **43**, 2923 (1978).

178. Maxam, A. M. & Gilbert, W. *Proc. Natl. Acad. Sci. USA* **74**, 560 (1977).

179. Maxam, A. M. & Gilbert, W. *Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages* (Academic Press, Inc. 1980).

180. Cameron, V. & Uhlenbeck, O. C. *Biochemistry* **16**, (23):5120 (1977).

181. Sambrook, J. Fritsch, E. F. & Maniatis, T. *Molecular Cloning a Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).

182. Ausubel, F. Brent, R. Kingston, R. E. et al. *Short Protocols in Molecular Biology* (John Willey & Sons, Inc. 1995).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

VITA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

VITA

Jiyan Chen was born on August 18, 1964 in Hangzhou (P. R. China), which is the most beautiful city in China, possibly in the world. He grew up there and educated there. In summer 1987, he got his Bachelor of Science in Chemistry from Hangzhou University. Before he moved to the United Stated in 1991, he had taught chemistry in Hangzhou for four years. He briefly stayed at Indiana University of Pennsylvania for master degree after he came to US, and then he started working on his Ph.D. in Chemistry Department of Purdue University in 1992. Initially his research was supervised by Dr. David G. Gorenstein, but due to Dr. Gorenstein's moving in 1994, he joined Dr. Donald E. Bergstrom's research family in Department of Medicinal Chemistry and Molecular Pharmacology. He was awarded a degree of Doctor of Philosophy in May 1997.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.