

**NMR Studies of Protein-DNA Interactions:
Determinations of DNA Structures Recognized by
Hin Recombinase and Studies of Their Roles in
Protein Binding Interactions.**

Thesis by
Yun Sun

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, California

1992

(Defended January 6, 1992)

To My Parents

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. John H. Richards, for being a reliable and sensible boss, and for his guidance, enthusiasm, patience and encouragement throughout my graduate study. Thanks for providing me with everything, including the freedom of independent research, which is invaluable to me. In my culture, a teacher once is a father forever. I will always carry that same kind of respect for you in the years ahead.

I would like to thank Drs. Frances H. Arnold, William A. Goddard III and Douglas C. Rees for serving in my Ph.D. committee. Special thanks go to Frances who introduced me to 2-D NMR and her continuous enthusiasm inspired and encouraged me to work in this field. I am also grateful to Dr. Peter B. Dervan for introducing me to this project, which I knew at first sight I would work on. My deepest appreciation goes to Dr. Sunney I. Chan for introducing me to Caltech and for his constant advice throughout my graduate study.

I would also like to thank the Richards' group for the scientific interactions as well as indispensable help in language, writing and culture adjustments. Among them, Tom Chang, Dave Long, Diane Hollenbaugh, Danny Casimiro, Marc Labgold and Bill Healey deserve special thanks. I am deeply grateful to Neil Farrow for discussing, proofing and editing my thesis. I thank Josh Kurutz for being working

with me on this project, and brought it fun and excitement. I thank Agnes Lew for letting me use her computer to write the thesis. Thanks also go to Phoebe Ray for keeping everything running smoothly. I also appreciate Juris Germanas and Tom Chang for encouraging me to run the 1991 L. A. Marathon, which was a truly sensational and once-in-a-lifetime experience.

Being a faithful resident of 362 South Catalina Apartment 202, I appreciated the chance to meet roommates of a variety of nationalities, experiences and personalities. I thank especially Mingxiang Li, Petr Pich, Blair Zajac, Howie Choset, Tim Yun Yen and Guangqing Chen for being great roommates and personal friends, and for making the apartment like a home.

I could not leave without mentioning Caltech C (Chinese Association), where I found most of my buddies and comrades in Caltech throughout all these years. Our common background, experiences, language, culture, values and goals have made it a sanctuary during difficult times, a platform of actions and a dance ground at fun times. My student life here in Caltech would have been much less colorful and memorable without you guys. My special thanks goes to Peijun Cong, Yuejin Guo and Youqi Wang for being helpful classmates and naturally-made friends; Dawei Dong for doing the greatest amounts of the work for Caltech C organization without expecting recognition; Guanhua Chen, Jian Zhu, Dali Ding, Qing Jiang, Mingsheng Han and Yixin Liu for being dedicated comrades; Junshan Wen for various help offered; Xiangdong Fang and the Keck House boys and girls, Yibin Cao,

Zhengwei Peng, Yong Guo and Weimin Lu for lots of parties and fun moments; Zhe Wu and Yan Chen for being great tennis partners and friends; Qizhi He for being a trustful personal friend; last but not least, Dr. Yuk L. Yung for being a warm, enthusiastic and caring teacher and advisor. Take care, all my old and new friends; we shall meet in another time at another place in the future.

I will certainly remember my numerous friends and acquaintances in the southland who has helped to make this part of my life unforgettable. Without mentioning their names, they are friends from UCLA, USC, Cal State Univ., Pasadena City College, Citrus College, Woodbury Univ., Northrop Univ. and many other universities. My thanks go to many good lawyers, politicians, businessmen and other important people here including President Everhart who helped us during difficult times. I am also grateful to the friendship and love of many members of the First Evangelical Church of San Gabriel Valley and the Lake Avenue Congregational Church in Pasadena. My special thanks go to Timothy Sun, Sue Ferrier and Enid Miller for being my best friends and for lots of the discussions about love and truth. Thanks for introducing me to a treasure of a lifetime.

Finally, I am most grateful to my family for their understanding, encouragement, love and support throughout all these years leading to this degree, ever since I left home for college at fourteen years old. I was also immensely fortunate to have my aunt and uncle around and enjoy their love as well as food whenever I needed. Thank you very much, auntie.

Abstract

The solution structures of two DNA oligomers were determined by the 2-D NMR method. These 14-base-pair DNA molecules contain the recognition sites for Hin recombinase. In spite of the differences in their sequences, the two structures are remarkably similar. The refined DNA structures possess a significant bend ($25\text{-}32^\circ$) in the middle of the helices. As a result of the bending, the nearby major groove is compressed at almost exactly the position where the recombinase binds. The DNA molecules were also found to have a deepened and narrowed minor groove near the continuous dA tracts, where the minor groove contact happens between the N-terminal residues of the recombinase binding domain and the DNA molecules. Such pre-existing unique features of the free DNA molecules are likely to contribute to the specific interaction of the protein and the DNA tracts. Structure determinations by the NMR method were preceded with the use of complete relaxation matrix analysis and restrained molecular dynamics. A data processing system were developed which allowed us to simulate the NMR spectra and quantifying intensities from an overlapped data set. A complete system for high-resolution structure determinations in solution were set up and evaluated.

The conformation of the Hin 52mer peptide—the binding domain of the Hin recombinase—and its binding interactions with the DNA oligomers are studied by NMR, circular dichroism and chromatographic methods. The conclusion is that the peptide does not have a unique and stable conformation alone as a single monomer in solution. The Hin peptide can be prevented from being aggregated by adjusting to acidic conditions, and it can be folded to a stable tertiary structure in an artificial environment with small amounts of trifluoroethanol. The Hin 52mer peptide conformation is greatly stabilized or induced by the presence of the DNA bearing specific binding sequences. The DNA binding activities of the peptide may be assayed by a chromatographic method. The behavior of the peptide in the binding complex and the characteristic structural features of the DNA molecules suggest the active role of the DNA in protein-DNA interactions providing complementary interactions with the peptide and stabilizing the peptide conformation upon its binding.

Table of Contents

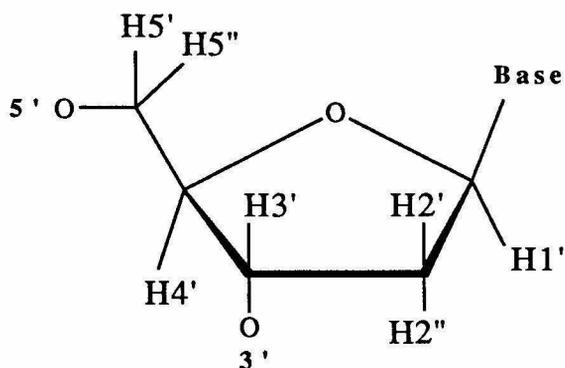
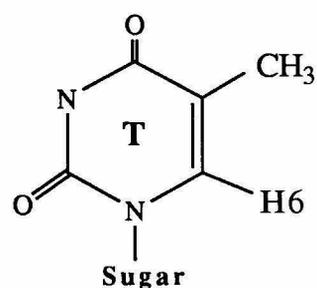
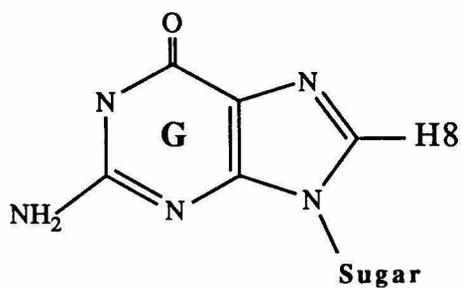
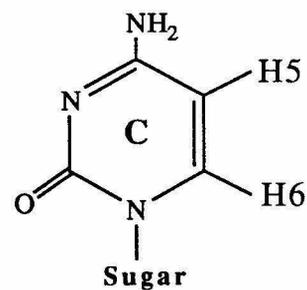
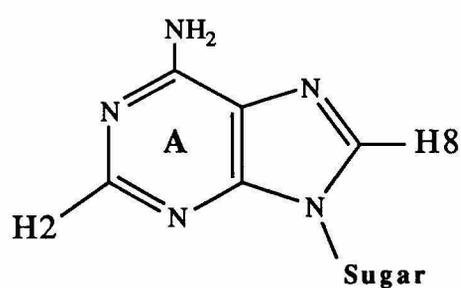
Acknowledgement	iii
Abstract	vi
Table of Contents	viii
Nomenclature and Abbreviations	x
Chapter 1 Introduction	1
References	13
Chapter 2 Solution Structure of the R.HixL DNA	
d(GGTTTTGATAAAG)•d(CTTTATCAAAAACC):	
Use of NMR and Restrained Molecular Dynamics	17
Materials and Methods	24
Results and Discussions	35
References.....	93
Chapter 3 Solution Structure of the L.HixL DNA	
d(GGTTCTTGAAAACC)•d(GGTTTTCAAGAACC):	
Sequence-dependent Structural Variations	
Correlated to DNA-Protein Interactions	97
Materials and Methods	99
Results and Discussions	102

	References	136
Chapter 4	Studies of the Protein-DNA Interactions between the Hin 52mer Peptide and DNA Oligomers	138
	Materials and Methods	139
	Results and Discussions	143
	References	172
Appendix	174

Nomenclature and Abbreviations

2D	two-dimensional
Å	angstrom
A	adenine
BIOGRAF	a molecular simulation and graphics software from BioDesign
bp	base-pair
C	cytosine
CD	circular dichroism
COSY	2D correlated spectroscopy
CROSSPEAK	data processing software created by the author
cross-peak	the off-diagonal peak in a Fourier-transformed 2D NMR spectrum
d-	deoxy-
$d_{\alpha N}$	distance (NOE interaction) between α -proton and amide proton
d_{NN}	distance (NOE interaction) between amide protons
DNA	deoxyribonucleic acid
DNA I	the left half site of the <i>hixL</i> , Figure 3-1
DNA II	the right half site of the <i>hixL</i> , Figure 2-1
DQF	double quantum filter

EDTA	ethylenediamine tetraacetic acid
FT	Fourier transform
FTNMR	NMR software from Hare Research
G	guanine
H-	protons, as illustrated in the following figure

Sugar protons**Base protons**

<i>hixL</i>	the left recombination site of Hin
<i>hixR</i>	the right recombination site of Hin
HPLC	high performance liquid chromatography
MARDIGRAS	software for NOE back-calculations from T. James
MD	molecular dynamics
MWCO	molecular weight cutoff
NMR	nuclear magnetic resonance spectroscopy
NOE	nuclear Overhauser enhancement; H2''(<i>i</i>)-H6/8(<i>j</i>) meaning the NOE interactions between H2'' of nucleotide <i>i</i> and H6 or H8 of nucleotide <i>j</i>
NOESY	2D nuclear Overhauser and exchange spectroscopy
pH*	direct pH meter readings of a D ₂ O solution
Pi	ortho phosphate
ppm	parts per million
propeller twist	the dihedral twisting angle between two base planes within a base-pair
Pur	purine
Pyr	pyrimidine
rMD	restrained molecular dynamics
RMS	root mean square
roll	the dihedral angle between two base-pair planes by which the base-pairs open up toward the minor groove
σ	standard deviation

slide	translation of the base-pairs relative to one another down their long axis
T	thymine
$T1$	spin-lattice relaxation time
$T2$	spin-spin relaxation time
$t1$	the first time domain of a 2D spectrum
$t2$	the second time domain of a 2D spectrum
TFE	3,3,3-trifluoroethanol
τ_c	correlation time
τ_m	mixing time
twist	rotation per base-pair step about an axis perpendicular to the mean base plane
UV	ultraviolet spectroscopy
•	base-pairing, e.g., A•T or G•C
{ x - y }	2D spectral region containing all NOE between spin(s) x and spin(s) y

Chapter 1

Introduction

One of the most fundamental and central questions in life science is how genes are activated, expressed, and controlled. The importance of protein-DNA interactions has been evident from the great amount of effort and enthusiasm put into this field for the past five years to solve crystal structures of protein-DNA binding complexes (Schultz et al., 1991; Luisi et al., 1991; Pavletich & Pabo, 1991; Kissinger et al., 1990; Wolberger et al., 1988; Otwinowski et al., 1988; Anderson et al., 1987; McClarin et al., 1986). However, knowledge about DNA and its role in protein-nucleic acid interactions is still very limited. Our perceptions about DNA structures, in particular, are literarily still at the stage of the idealized double helix of Watson and Crick (1953). Without undermining the monumentary significance of their work, we have known in recent years that DNA structures do exhibit significant variations with sequences (Dickerson & Drew, 1981). More pronounced structural deviations were seen in crystal structures of protein-DNA binding complexes, e.g., DNA-Eco RI endonuclease complex (McClarin et al., 1986) and catabolite gene activator protein (CAP)-DNA complex (Schultz et al., 1991).

In contrast to the abundance of crystal structures of proteins and protein-DNA complexes, relatively few DNA structures have been solved

by the crystallographic method (Nelson et al., 1987; Larsen et al., 1991). There is a knowledge gap, therefore, which hampers our understanding of the protein-DNA interactions. We feel the lack of reference ground for understanding those DNA structural features found in protein-DNA complexes. Part of the difficulties of crystallographic methods in studying DNA is likely that DNA molecules are not readily crystallized in aqueous solutions without certain crystallization-promoting additives. However, it is known that DNA conformations are sensitive to solution conditions (Dickerson et al., 1985) and concerns can be raised about the relevance of the determined structures to the physiological realities. In this regard, the high-resolution nuclear magnetic resonance (NMR) method for solution structure determination appears to be an excellent alternative.

Compared to the solution structure determination of globular proteins by NMR spectroscopy (Wüthrich, 1989; Wright, 1989), DNA structure determination is technically more demanding since the sequence-dependent structural variations are relatively subtle, albeit biologically important changes. In general, higher precision and confidence of the data—as well as a more exact theoretical model used to deduce the structure—are required, as compared to proteins. The traditional method of DNA structure determination can be exemplified by gauging interproton distances directly with the corresponding nuclear Overhauser enhancements (NOE) between them, followed by a distance geometry (DG) computing algorithm to generate the structures that fit the distances (Patel et al., 1987; Nilges et al., 1987). Until very recently it

was not realized that such a direct gauging method may introduce substantial errors in distances (Borgias et al., 1990; Nerdal et al., 1989). Methods or computer programs are now developed for back-calculation of NOE intensities (Banks et al., 1989), complete relaxation matrix analysis (Borgias et al., 1990) or iterative relaxation matrix approach (Koning et al., 1991). These methods provided more exact theoretical treatment of the distance-NOE relationship. Another development in NMR structure determination was the more favored use of restrained molecular dynamics (Gochin & James, 1990; Baleja et al., 1990) relative to the distance geometry method, mainly due to the large radius of refinement convergence of the former (Nilges et al., 1987).

In our research work reported here, we have used complete relaxation matrix analysis to deduce distances from NOE intensities, ensuring a most exact theoretical model in the deduction process. Restrained molecular dynamics was used to search the converged structures in a wide conformational space. We also developed a data processing system which allowed us to simulate the NMR spectra and extract intensities from an overlapped data set. Parallel experiments were performed to rigorously evaluate the consistency of each step in going from the raw NMR spectra to the refined DNA structures (Chapter 2).

As a result of these efforts, we have consistently obtained record numbers of distance constraints for a DNA molecule (440 constraints for DNA I, for instances). The two DNA structures reported here, each with 28 base-pairs, are also the largest DNA molecules having their complete

structures determined by NMR (other large DNA's: Baleja et al., 1990; Nerdal et al., 1989; Banks et al., 1989; Grütter et al., 1988). More significantly, for the first time, we have studied a series of different sequences which have the same biological functions. The two structures here are remarkably similar in spite of the differences in sequences. The unique and common features, i.e., pronounced structural deviations from the idealized B-form DNA, are well-correlated to the protein-DNA binding interactions in the *Hin* recombinase system, in which these DNA sequences reside, as described below.

The *Hin*, in terms of its biological functions at gene level, can be classified into a gene category called *transposons*, or mobile genes (Stryer, 1988). The presence of transposons makes possible large-scale frequent rearrangements of the genome. Found in *Salmonella typhimurium*, the *Hin* recombinase, the expressed form of *Hin*, inverts reversibly a specific 996 base-pair segment of DNA, causing phase variations between two flagellar antigen types (Figure 1-1). More detailed descriptions about the recombination mechanism and functions have been given by Hughes et al. (1988), Silverman & Simon (1980) and Simon et al. (1980). A simplified illustration of the DNA inversions is given in Figure 1-2. The inversion of the DNA loop is accomplished by disconnecting and rejoining the DNA molecule at the recombination sites.

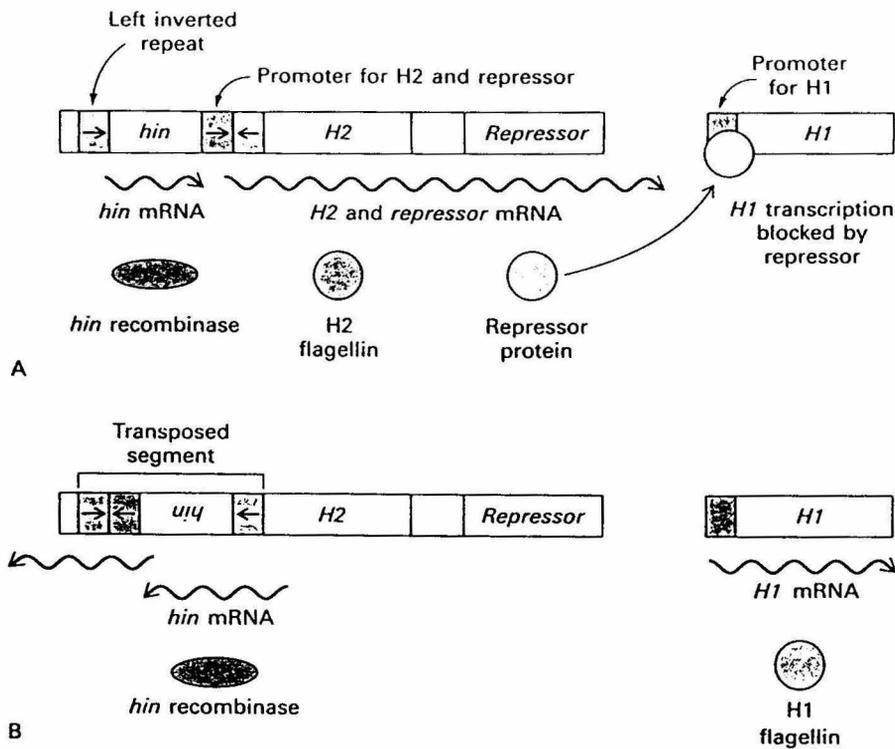


FIGURE 1-1: Phase variation in *Salmonella*. Flagellins H1 and H2 are expressed in a mutually exclusive manner. (A) In phase 2, the H1 gene is silenced by a repressor protein formed along with the H2 protein. (B) In phase 1, inversion of a DNA segment catalyzed by Hin recombinase encoded by it leads to the loss of the promoter for H2 and the repressor. H1 is then expressed. Phase variations are reversible. Adapted from Stryer (1988) *Biochemistry* page 818.

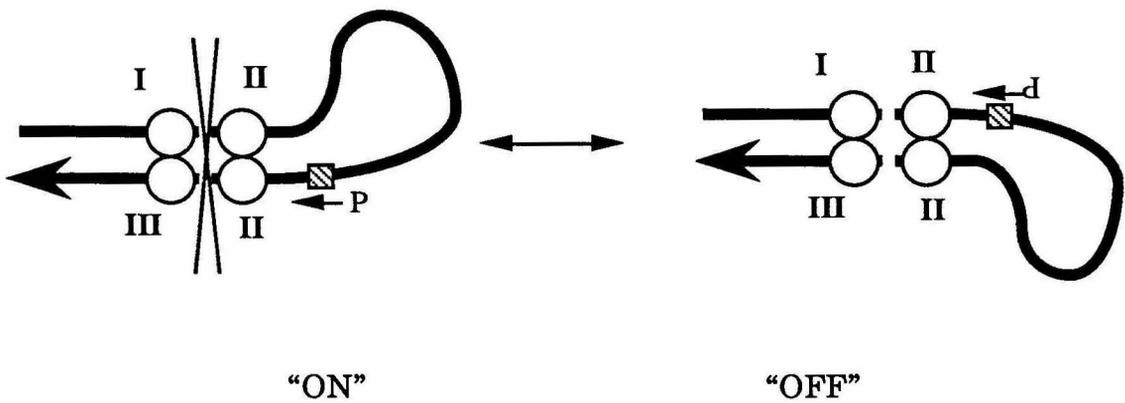


FIGURE 1-2: The inversion of the DNA segment in recombination. The DNA loop containing the promoter is inverted by a cut-and-reconnect process, performed by at least four enzyme molecules. When the promoter is in "on" configuration with the genes downstrand, H2 and H1-repressor are expressed. Two right halves of the DNA recombination sites are labelled the same (II) since they have the same binding sequence for Hin recombinase.

The Sequence of the DNA recombination sites, designated *hixL* and *hixR* (Glasgow et al, 1989; Hughes et al., 1988; Bruist et al., 1987), is shown in Figure 1-3. The Hin recombinase also recognizes a secondary binding site (sequence not shown).

	DNA I (L. <i>HixL</i>)	DNA II (R. <i>HixL</i>)
<i>HixL</i> on	TTATTGG <u>TTCTTGAAA</u> ACCAAGG <u>TTTTTGATAA</u> AAGCAATC	
	AATAAC <u>CAAGAACTTTT</u> GGTTCC <u>AAAACTATT</u> TCGTTAG	
	DNA III (L. <i>HixR</i>)	(R. <i>HixR</i>)
<i>HixR</i> on	CATAAA <u>TTTTCCTTTT</u> GGAAGG <u>TTTTTGATAA</u> ACCAATGT	
	GTATTT <u>AAAAGGAAA</u> ACCTTCC <u>AAAACTATT</u> GGTTACA	

FIGURE 1-3: The sequence of *hixL* and *hixR* with the H2 flagellin promoter in “on” configuration. Numbering reads -20 to +20 (left to right) for *HixL* and +977 to +1015 (right to left) for *HixR*. The bold letters denote recognition sequences for the Hin DNA binding domain (half sites of recombination). Underlined sequences, named DNA I and DNA II, are the DNA oligomers synthesized and studied in this research. DNA III is studied by Kurutz et al. (unpublished results).

The Hin recombinase is a 190 amino acid protein related to the resolvase recombinase of transposon $\gamma\delta$ (Hughes et al., 1988). The DNA binding domain was identified by sequence homology with the resolvase. The carboxyl-terminal 52 amino acid segment of the Hin recombinase (Hin 52mer) was chemically synthesized and demonstrated to have the same binding specificity of the intact enzyme (Bruist et al., 1987). The

binding domain/Hin 52mer peptide is believed to belong to a large family of DNA-binding proteins possessing a "helix-turn-helix" motif (Struhl, 1989; Pabo & Sauer, 1984). Based on the knowledge of other helix-turn-helix protein structures and experimental observations, Sluka (1988) and Plaxco et al. (1989) have proposed a model structure for the Hin recombinase binding domain (Figure 1-4). An interesting feature of this binding domain is the minor groove contact as suggested by the affinity cleavage experiments (Sluka et al., 1990). The Hin binding domain is also similar to the eukaryotic *Drosophila* homeodomains (Affolter et al., 1991).

In our studies, the refined DNA structures possess a significant bend ($25-32^\circ$) in the middle of the helices for both sequences. Structural deviations of this scale have been reported for other sequences in NMR studies (Nilges et al., 1987; Nerdal et al., 1989; Banks et al., 1989). Compared to other more isolated cases, our results seem more reoccurring and convincing. As a result of the bending, the nearby major groove is compressed at almost exactly the position where the recognition helix sits. Such a pre-existing unique feature of the free DNA molecule is likely to contribute to the specific interaction of the protein and the DNA tracts. The DNA is also found to have a deepened and narrowed minor groove near the continuous dA tracts, where the minor groove contact happens by the interactions of A•T hydrogen bond accepters and arginines on the N-terminal of Hin binding domain (Plaxco et al., 1989). Both structural features, as well as the effects of the DNA bending to orientate minor and major grooves to the protein binding domain, are

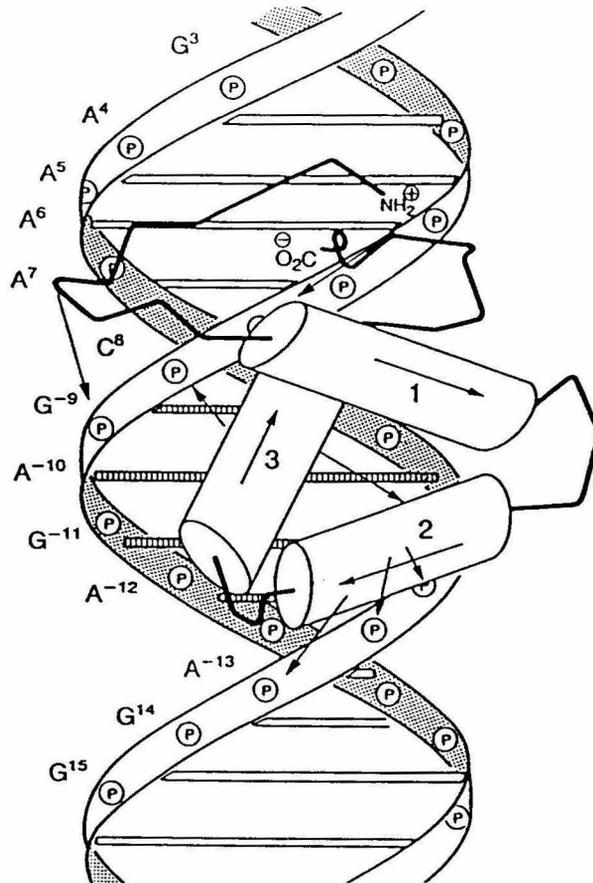


FIGURE 1-4: A model structure of the DNA-binding domain Hin recombinase with a DNA half site. Helix 3 is the recognition helix. Helix 2 and 3 form the "helix-turn-helix" motif. Adapted from Plaxco et al., 1989.

likely reasons that can account for the high affinity of the peptide to the DNA, as compared to other prokaryotic helix-turn-helix binding domains (Affolter et al., 1990). The DNA's role here is, therefore, much more than just being a passive target of recognition without any characteristic feature. Instead, it actively prepares a conformational environment for the protein recognition (Chapter 3).

In Chapter 4, we report the studies on the Hin 52mer peptide and its binding interactions with the DNA binding sequences by NMR, circular dichroism and chromatographic methods. The conclusion is that the peptide does not have a unique and stable conformation alone as a single monomer in solution (for comparison: Wade-Jardetzky et al., 1978). The Hin peptide can be prevented from being aggregated by adjusting to acid buffer, and it can be folded to a stable 3-D conformation in an artificial environment with small amounts of trifluoroethanol. Its structure may be studied in the future and give useful insight about the tertiary structure of the binding domain (for other NMR studies of DNA binding protein domains: Billeter et al., 1990; Qian et al., 1989; Arrowsmith et al. 1990; Hard et al., 1990; Lee et al., 1989; Parraga et al., 1988; Pan & Coleman, 1991). The Hin 52mer peptide conformation is greatly stabilized or induced by the presence of the DNA bearing specific binding sequences. Examination of the peptide-DNA binding complex confirmed the tight binding between the two and indicate a well-defined 3-D structure which may be solved in the future to a good resolution (for comparison with *lac* headpiece complex: Boelens et al., 1987a; Boelens et al., 1987b; Lamerichs et al., 1989). The DNA binding activities of the

peptide may be assayed by a chromatographic method. The behavior of the peptide in the binding complex further suggested the role of the DNA as an active partner of protein-DNA interactions providing complementary interactions with the peptide and thus stabilizing the peptide conformation upon its binding.

REFERENCES

- Anderson, J. E., Ptashne, M., & Harrison, S. C. (1987) *Nature* **326**, 846.
- Affolter, M., Percival-Smith, A., Müller, M., Billeter, M., Qian, Y. Q., Otting, G., Wüthrich, K., & Gehring, W. J. (1991) *Cell* **64**, 879.
- Affolter, M., Percival-Smith, A., Müller, M., Leupin, W., & Gehring, W. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4093.
- Arrowsmith, C. H., Pachter, R., Altman, R. B., Iyer, S. B., & Jardetzky, O. (1990) *Biochemistry* **29**, 6332.
- Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry* **29**, 4828.
- Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry* **28**, 6996.
- Billeter, M., Qian, Y. -Q., Otting, G., Müller, M., Gehring, W. J., Wüthrich, K. (1990) *J. Mol. Biol.* **214**, 183.
- Boelens, R., Scheek, R. M., Lamerichs, R. M. J. N., de Vlieg, J., van Boom, J. H., & Kaptein, R. (1987a) in *DNA-ligand Interactions* (Guschlbauer, W., & Saenger, W., Eds.) pp.191-215, Plenum, New York.
- Boelens, R., Scheek, R. M., van Boom, J. H., Kaptein, R. (1987b) *J. Mol. Biol.* **193**, 213.
- Borgias, B. A., Gochin, M., Kerwood, D. J., & James, T. L. (1990) *Progress in NMR Spectroscopy* **22**, 83.
- Bruist, M. F., Horvath, S. J., Hood, L. E., Steitz, T. A., & Simon, M. I. (1987) *Science* **235**, 777.
- Dickerson, R. E., & Drew, H. P. (1981) *J. Mol. Biol.* **149**, 761.

- Dickerson, R. E., Kopka, M. L., & Pjura, P. (1985) in *Biological Macromolecules and Assemblies*, Vol. 2, *Nucleic Acids and Interactive Proteins* (Jurnak, F. A., & McPherson, A., Eds.) pp. 37-126, 490-493, Wiley-Interscience, New York.
- Glasgow, A. C., Bruist, M. F., & Simon, M. I. (1989) *J. Biol. Chem.* **264**, 10072.
- Gochin, M., & James, T. L. (1990) *Biochemistry* **29**, 11172.
- Grütter, R., Otting, G., Wüthrich, K., & Leupin, W. (1988) *Eur. Biophys. J.* **16**, 279.
- Hard, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, Kaptein, R. (1990) *Science* **249**, 157.
- Hughes, K. T., Youderian, P., & Simon, M. I. (1988) *Genes Dev.* **2**, 937.
- Kissinger, C. R., Liu, B. S., Martingblanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell* **63**, 579.
- Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry* **30**, 3787.
- Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B., & Rüterjans, H. (1989) *Biochemistry* **28**, 2985.
- Larsen, T. A., Kopka, M. L., & Dickerson, R. E. (1991) *Biochemistry* **30**, 4443.
- Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., Wright, P. E. (1989) *Science* **245**, 635.
- Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., & Sigler, P. B. (1991) *Nature* **352**, 497.

- McClarín, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., Rosenberg, J. M. (1986) *Science* **234**, 1526.
- Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature* **330**, 221.
- Nerdal, W., Hare, D. R., & Reid, B. R. (1989) *Biochemistry* **28**, 10008.
- Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., & Nilsson, L. (1987) *Biochemistry* **26**, 3718.
- Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature* **335**, 321.
- Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293.
- Pan, T. & Coleman, J. E. (1991) *Biochemistry* **30**, 4212.
- Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., Kleivit, R. E. (1988) *Science* **241**, 1489.
- Patel, D. J., Shapiro, L., & Hare, D. (1987) *Ann. Rev. Biophys. Biophys. Chem.* **16**, 423.
- Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809.
- Plaxco, K. W., Mathiowetz, A. M., & Goddard III, W. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 984.
- Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1989) *Cell* **59**, 573.
- Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991) *Science* **253**, 1001.
- Silverman, M. & Simon, M. (1980) *Cell* **19**, 845.
- Simon, M., Zieg, J., Silverman, M., Mandel, G., Doolittle, R. (1980) *Science* **209**, 1370.

- Sluka, J. (1988) Ph.D. Thesis (California Institute of Technology, Pasadena, CA).
- Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., & Dervan, P. B. (1990) *Biochemistry* **29**, 6551.
- Struhl, K. (1989) *Trends Biochem. Sci. (TIBS)* **14**, 137.
- Stryer, L. (1988) in *Biochemistry* pp. 698-701, 817-818, W. H. Freeman & Company, New York.
- Wade-Jardetzky, N. et al. (1978) *J. Mol. Biol.* **128**, 1259.
- Watson, J. D. & Crick, F. H. C. (1953) *Nature* **171**, 737.
- Wolberger, C., Dong, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature* **335**, 789.
- Wright, P. E. (1989) *Trends Biochem. Sci. (TIBS)* **14**, 255.
- Wüthrich, K. (1989) *Science* **243**, 45.

Chapter 2

**Solution Structure of the R. *HixL* DNA
d(GGTTTTTGATAAAG)•d(CTTTATCAAAAACC):
Use of NMR and Restrained Molecular Dynamics**

Recent advancement of two-dimensional (2-D) NMR techniques has made structure determinations of biological macromolecules in solution one of the most active research areas of biochemistry (Wüthrich, 1986). The primary attraction of 2-D NMR methods versus the conventional 1-D NMR is that they effectively resolve crowded regions in the spectrum by mapping the spectral information onto two frequency axes (Bax & Lerner, 1986). The use of these techniques potentially enables us to achieve sequence-specific assignments of resonances and to obtain, in a single experiment, hundreds or thousands of structural constraints, which can ultimately lead to a high-resolution structure of the molecule in question.

The 2-D NMR techniques rely, in particular, on the measurement of nuclear Overhauser enhancement (NOE) to determine interproton distances. However, until very recently, the accuracy and reliabilities of distances derived from NOE intensities are complicated by, not only experimental imperfections, but also the approximate computational methods used to derive the distances (Borgias et al., 1990). The distance between two spins is often estimated by assuming inverse proportionality

to the sixth root of the NOE cross-peak intensity. Thus, an unknown distance can be estimated by scaling its intensity to a reference distance:

$$r_{ij} = r_{\text{ref}} (a_{ij}/a_{\text{ref}})^{-1/6} \quad (\text{Eq. 2-1})$$

where r_{ij} and a_{ij} are the distance to be determined and the corresponding NOE intensity, respectively; r_{ref} is a reference distance taken from fixed distances in the molecule, a_{ref} is the reference intensity.

It is known that distances so derived are only approximate since the cross-peak intensity due to direct cross-relaxation between spins i and j is modified by additional cross-relaxation with other spins nearby, a phenomenon known as spin diffusion. In fact, Eq. 2-1 is an approximation of the complete solution (Macura & Ernst, 1980):

$$\mathbf{a}(\tau_m) = \exp(-\mathbf{R}\tau_m) \quad (\text{Eq. 2-2})$$

$$= 1 - \mathbf{R}\tau_m + \mathbf{R}^2\tau_m^2/2 - \dots + (-1)^n \mathbf{R}^n \tau_m^n/n! + \dots \quad (\text{Eq. 2-3})$$

where \mathbf{a} is a matrix consisting of all NOE intensities between all spins in the system. τ_m is the mixing time of the NOE experiment; \mathbf{R} is the matrix describing the complete dipole-dipole relaxation network. Elements in the matrix \mathbf{R} are relaxation rates proportional to r_{ij}^{-6} (Soloman, 1985).

In Eq. 2-3, if the mixing time is short, the higher order terms of the series expansion vanish. The off-diagonal elements of \mathbf{a} , a_{ij} , will be directly proportional to r_{ij}^{-6} and Eq. 2-1 becomes valid. Thus, the traditional way of determining distances is to measure NOE intensities

at very short mixing times ($\tau_m < 50$ ms) or to measure the initial build-up rates of the NOE intensities vs. mixing time. The distances obtained might be used for further structure refinements.

A problem with using only intensities of short mixing times is that a significant fraction of the distances will not be observed due to their small intensities. Furthermore, for many of those observed intensities, the accuracy could be relatively susceptible to noise disturbance, because the early build-up period is far away from the optimum mixing time at maximum cross-peak intensity. This approach seems to compromise the quantity of experimental constraints, as well as the accuracy of the experimental data, both of which are critical for successful structure determination of macromolecules like DNA (Gochin & James, 1990). We feel that further refinements of such traditional approaches are necessary for our structure evaluations.

The study of sequence-dependent structural variations in DNA is a difficult and challenging task for 2-D NMR methodology. Unlike globular proteins, there are rarely long-range NOE interactions available. The structural variations, although significant for many of the DNA functions, are expected to be small. For these reasons, pushing the NMR method to its accuracy limit, as well obtaining the most complete sets of constraints possible, is vitally important for successful structure determinations (Nilges et al., 1987a).

In order to obtain the most accurate structural constraints from the 2-D NMR data, we decided to take the approach of complete relaxation matrix analysis, by using the method developed by Borgias et al. (1990).

This approach utilizes the fact that NOE cross-peak intensities may be back-calculated from a putative structure by diagonalization of the matrix exponentiation in Eq. 2-2 (Keepers & James, 1984; Boelens et al., 1989). The predicted intensities are then matched with experimental intensities and the distances in the putative structure are adjusted iteratively until the two sets of intensities converge. This approach is especially appropriate with our object of study. From a qualitative evaluation of the NOE intensities, we have concluded that the DNA molecule has a conformation very close to the standard B-form structure. The B-DNA, therefore, could serve as an excellent initial structure for the iterative calculations. Such an approach not only reduces the possible systematic errors from an approximate method, but also gives a best mixing time $\tau_m = 200$ ms for measuring distance constraints (Borgias et al., 1990). This value is near the optimum mixing-time mentioned above, under which many more of the distances (up to 5 Å) can be measured to a confident accuracy.

The quantities of available constraints are important for accurate back-calculations and the structural refinement (Gochin & James, 1990; and Chapter 3). To obtain structural constraints as complete as possible, we have adopted experimentally an encompassing approach, where all possible NOE interactions are generated *a priori* and the corresponding intensities extracted. The whole data sets are then processed by an automatic procedure to evaluate, deconvolute and select valid cross-peak intensities. Combining with the full-matrix distance calculation, an integrated data processing system (CROSSPEAK) was constructed to

process the vast amounts of distance constraints in an objective, reliable and automatic fashion. We have also adopted an approach using parallel experiments to evaluate the self-consistency of the data and the methodology of the structure determination.

Having distance constraints available, several algorithms can be used to generate the solution structures in high resolution. Among them are distance geometry (Banks et al., 1989; Pardi et al., 1988; Boelens et al., 1988; Havel et al., 1983), metric matrix distance geometry calculations (Havel & Wüthrich, 1985), iterative relaxation matrix analysis (Koning et al., 1991) and restrained molecular dynamics (Gochin & James, 1990; Baleja et al., 1990; Boelens et al., 1989; Nilges et al., 1987a,b; Nilsson et al., 1986). The method we have chosen is restrained molecular dynamics in which the interproton distances are incorporated into the total energy function of the molecular system in the form of penalty energy for distance violations. Molecular dynamics was the favored choice because of the following reasons. Firstly, the molecule can search a much larger conformational space and move across false energy minima. Secondly, we can take advantage of the molecular forcefield to keep the chemical structure at optimum configurations, and to smooth out some of the missing constraints. Thirdly, the experimental accuracy of the distance data can be reflected in the total energy by properly relating them to the distance violations. In restrained molecular dynamics, distance constraints are an artificial part of the molecular system. Therefore, their energy contributions can be adjusted freely, representing the emphasis on the NMR constraints vs. the empirical energy potential.

Apparently, the stronger the emphasis on NMR constraints, the smaller is the distance violation. It seems only natural to have the distance constraints incorporated in a way that distance violations are approaching the experimental distance uncertainty. The method of restrained molecular dynamics is to sample a sufficiently large conformational space under such NMR constraints with established experimental accuracy. An ensemble of the converged structures will define the most probable structure of the molecule in solution to the extent limited by intramolecular motions and the experimental accuracy of NMR constraints.

Despite the difficulties, evaluations of DNA structural variations by NMR methods also offers some advantages. It goes that the qualitative structures of the DNA are usually known before the quantitative analysis, in our case, as very close to the B-DNA conformation. Starting from this knowledge, B-DNA is used as the starting (putative) structure for NOE back-calculations, generating distance constraints. And it would be reasonable only to start restrained molecular dynamics from the vicinities of B-DNA to search for the converged structures of the DNA molecule. Additionally, it is also well known that the structural determinations by NMR can be performed in physiological solution and other choices of conditions, which are important for DNA conformation.

In this chapter, we present the DNA structure containing the *right* half of the *hixL* recombination site of Hin recombinase (Hughes et al., 1988), designated *R.hixL*. *R.hixL* has an sequence identical to *R.hixR*, the *right* half of the cross-over recombination site (*hixR*) on the

“on” configuration of the DNA, whereas the other two Hin binding-sites, *L.hixL* and *L.hixR* have different sequences (Figure 1-3). In our shorthand, as will be referred to thereafter, the molecule *R.hixL* is named DNA II. This sequence is obviously important as it is present in two of the four natural binding sites of Hin recombinase as named above. It is also an interesting sequence due to its abundance of A•T base-pairs. The sequence contains a penta(dA) tract that has been studied extensively for its role in DNA bending (Koo et al., 1986). The DNA structure is studied in the absence of protein as a basis for comparison to the structure of DNA within the protein-DNA binding complex. This study is part of the effort to determine all DNA structures that the protein binds specifically to, as well as other related DNA structures. In doing so, we may gain understanding of how a DNA structure plays its role in protein-DNA recognitions with Hin recombinase and other systems.

Considerably large sections of this chapter will be devoted to evaluating the methodology, namely, NMR constraint determination and restrained molecular dynamics. Structure comparisons will be put into more focused discussions in the following chapters, where studies of other DNA structures are presented.

MATERIALS AND METHODS

Sample Preparation

The two 14-base oligonucleotides, d(GGTTTTTGATAAAG) and d(GTTTATCAAAAACC) of DNA II, were synthesized via the solid phase phosphoramidite method on a 10 micromole scale. After final

of the solutes. The DNA II was eluted as double-stranded at a gradient of 0.01 M to 0.20 M dipotassium phosphate, adjusted to pH 7.0 by phosphoric acid, in 20:80 acetonitrile/water solutions. The pooled fractions were desalted in a procedure similar to that of the first HPLC purification. The concentration of the DNA duplex was determined again by the absorbance at 260 nm with a molar absorbance of 1.78×10^5 . To the fiber-like final product was added calculated amounts of phosphate, NaCl and EDTA, adjusted to pH* 7.0 in 0.3 mL D₂O, exchanged repeatedly with D₂O at lyophilizations, and finally taken up in 0.3 mL of 99.996% D₂O under N₂. Final concentrations were 5.6 mM DNA II in duplex, 20 mM phosphate, 120 mM NaCl, 0.1 mM EDTA, pH* 7.0. All pH measurements are direct meter readings using a glass electrode uncorrected for the isotope effect.

NMR Spectroscopy

All NMR spectra were recorded on two Bruker AM500 spectrometers with an operating frequency of 500 MHz for protons. Preliminary 2-D Fourier transform and graphic plotting were carried out on the on-line Aspect 3000 computer. NOESY spectra were recorded using the conventional pulse sequence (Jeener et al., 1979).

Phase-sensitive data were collected with 2K complex points by using the time-proportional phase incrementation method (Redfield & Kuntz, 1975; Marion & Wüthrich, 1983). Appropriate phase cycling was used for quadrature detection and to eliminate peaks due to multiple quantum coherence transfer. The number of transients collected or t_1

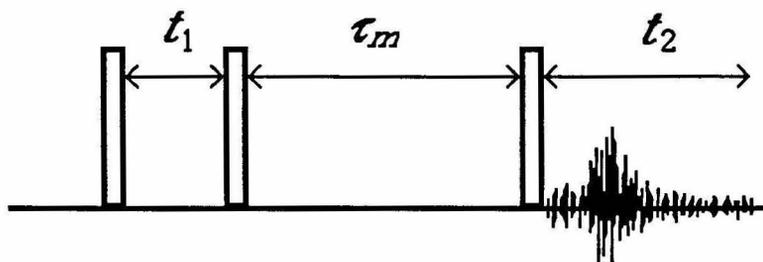


FIGURE 2-2: NOESY $90^\circ - 90^\circ - 90^\circ$ pulse sequence. Time domain is divided into evolution period t_1 , mixing period τ_m and acquisition period t_2 . The three 90° pulses were named in time sequence: excitation, mixing and detection pulses.

increments was 512 or 1K. Each NOESY experiment was limited to about 20 hours by adjusting the number of accumulations (40-128). The delay time between acquisitions was 1 second and the spectral width was set to cover all the resonances of the sample (3.8 KHz). Mixing times of $\tau_m = 50$, and 100 to 800 ms in 100-ms increments were used with a random delay of $\pm 10\%$ or 20 ms incorporated to suppress zero quantum coherence. The residual HOD signal was suppressed by presaturation.

NMR spectra in water were recorded at room temperature (25°C) with moderately strong presaturation of the H_2O resonance. Such suppression was sufficient to perform a NOESY at $\tau_m = 200$ ms. NMR spectra in D_2O were taken for the DNA molecule at a full range of temperatures from 25°C up to the melting of DNA duplexes in 5°C increments. Unless otherwise mentioned, NOESY spectra for quantitative measurement were taken in D_2O at 35°C .

All NMR data were stored on magnetic tape and processed with FTNMR (Hare Research) software and an integrated data processing system CROSSPEAK constructed by ourselves, on two DEC VAXworkstation 3500 computers. The Bruker data format was transformed to FTNMR format by an FTNMR associated executable program BRUKER.EXE. Free induction decays were weighted in each dimension by a window function, the 45° -shifted and squared sine-bell, to enhance spectrum resolution and to avoid truncation effects. Spectra were zero-filled before the Fourier transform so that the final transformed spectra were 1K by 2K real data points. To avoid artifacts, spectra were never symmetrized across the diagonal. To reduce t_1 ridges and correct baselines, the first ten and last ten nominally blank rows were summed along the t_1 dimension, scaled and subtracted from the two-dimensional spectra after Fourier transform (this method is similar but more robust than the one by Otting et al. 1986).

Quantifying Peak Intensities

NOE intensities were obtained by using an integrated data processing system CROSSPEAK (unpublished procedures). The first step is to identify and locate all cross-peaks in the spectra according to an all-cross-peak list generated for the specific DNA sequence, assuming a right-handed DNA conformation (B-form or A-form DNA) and its variations. All cross-peaks were then integrated within certain ellipses, each defined by two radii and the assigned locations, regardless of their intensities or the extent to which they overlap with other peaks. The

coordinates (chemical shifts) of all cross-peaks were checked against each other. Coordinates arising from a same resonance were sorted together and compared. Line width was measured or defined for each proton resonance. The complete data sets including, e.g., 622 cross-peaks for DNA II, were processed by a fully-automatic deconvolution program. The program simulated a virtual 2-D NMR spectra of 622 cross-peaks based on the location, line widths and an initial intensity of each cross-peak. The simulated NMR spectra were integrated with the same parameters as those of the experimental integrations. The difference of the two integrals was used to adjust the initial intensity of each cross-peak. Iterative calculations were performed until the two sets of integrals converged. About 25-30% of the 622 cross-peaks were rejected due to their small intensities or severe overlappings with other peaks. Calculations were carried out at an average rate of 30 min. each. Gaussian line-shape was assumed in the processing. Peak cutoff-range was set at 95% intensity. By this standard, an estimated 45-50% of all cross-peaks were isolated ones in the original spectra. Others were overlapping, to various extent, with neighboring peaks. After deconvolutions, the percentage of useful cross-peaks increased to about 75% (The effect is more dramatic for less resolved 2-D spectra, as for DNA I in the next chapter). NOE interactions involving H5' and H5" were not evaluated.

Comparing to other quantifying methods (Gochin et al., 1990; Denk, 1986; Broido et al., 1984), ours appeared to be the most comprehensive, versatile and practically convenient. A diagram

illustrating the structure of this integrated data processing system (CROSSPEAK), a list of all computer programs and their usages are available in the Appendix.

Distance Constraint Generation

Distance constraints were calculated from intensities with the program MARDIGRAS (Borgias & James, 1990). A correlation time $\tau_c = 2$ ns was obtained for the 14 bp DNA at 35°C through measuring the spin-lattice ($T1$) and spin-spin ($T2$) relaxation times of the base protons, according to the method by Suzuki et al., 1986. As has been noted by Gochin et al. (1990), actual measured correlation times vary among different protons in the molecule. We therefore calculated distances using isotropic correlation times of 0.5, 1, 2, 3 and 4 ns. It appears calculations with $\tau_c = 1$ ns and 2 ns give the best fit to the intensities. The distances calculated from these two differed by 0.2 Å (RMSD), which is within the experimental error. In fact, both sets of distances were used in molecular dynamics and the converged structures are indistinguishable (see RESULTS AND DISCUSSIONS, Table 2-4). The insensitive nature of calculated distances on τ_c seems to further justify the approach of using a single isotropic correlation time for calculating distances in DNA molecules of this size (Reid et al., 1989).

The starting structure was standard B-form DNA built on the structural modeling software BIOGRAF (BioDesign) based on crystal structure data. The coordinate file from BIOGRAF was in Brookhaven (BKV) format, which was converted into Protein Data Bank (PDB)

format by a routine in CROSSPEAK. The correlation time was appended to the PDB coordinate file by CORMA_IN and protons were added by NEWHYD, both were MARDIGRAS associated programs. Noise level was determined as 0.10 in absolute unnormalized value. The fast rotor motion of the thymine methyl group was treated using the Methyl-Jump-3 model in MARDIGRAS. The above-mentioned procedure was embedded in CROSSPEAK data processing system.

Using the starting structure, MARDIGRAS set up a complete relaxation-rate matrix. Diagonalization of the rate matrix is followed by calculation of an NOE intensity matrix. Experimental intensities are substituted into the intensity matrix, which is then back-transformed to a new relaxation matrix. Distances may be derived from the relaxation matrix as the program execution exits, or the matrix can be reset and let go on several cycles of calculations until the calculated and the experimental NOE intensities converge. Each calculation took on average 1.5 hr to complete on a DEC VAXworkstation 3500. Distances corresponding to inconsistent intensities as determined by MARDIGRAS were rejected. Constraint files from MARDIGRAS output were converted to distance constraints in BIOGRAF-acceptable format by a routine in CROSSPEAK. The original force constants, 50 kcal/mol/Å² being maximum and much reduced for longer distances, were scaled down to appropriate strength in the process.

Structure Refinement

The interproton distances derived with MARDIGRAS were not influenced by energetic considerations but solely reflect the simultaneous fittings of all experimental 2-D NOE intensities (Gochin & James, 1990). In restrained molecular dynamics these distances were incorporated into the total energy function of the DNA molecule in the form of constraint energy for distance violations:

$$E_{\text{constr}}^{ij} = 0.5 k_{ij} (r_{ij} - r_{ij}^{\circ})^2 \quad (\text{Eq. 2-4})$$

where r_{ij} and r_{ij}° are the calculated and experimentally-measured distances, respectively. Individual force constants k_{ij} were determined by MARDIGRAS:

$$k_{ij} = k_B T / (\Delta_{ij})^2 \quad (\text{Eq. 2-5})$$

where k_B is the Boltzmann constant, T is the absolute temperature, Δ_{ij} are the distance calculating errors, which were limited to a minimum of 0.1. (In MARDIGRAS, Δ_{ij} were calculated as the differences between the full-matrix-calculated distances and distances calculated via isolated-spin-pair approximation. This might not be an exact evaluation of distance errors. However, it does reduce the force constants of larger distances. The reductions were desirable because they were derived from small, relatively uncertain intensities — communications, Liu, H.)

All force constants k_{ij} were scaled to a maximum of 10 kcal/mol/Å². The distance constraints involving thymine methyl groups were applied to the methyl carbon instead of the geometric center of three protons (the pseudo-atom). Such an approximation for pseudo-atoms can introduce a maximum error of ± 0.33 Å, normally well within ± 0.1 Å. To reduce systematic errors introduced by such an approximation, corrections were made assuming a geometry of B-form DNA. Force constants of these constraints were scaled to half of the original values. An additional 16 hydrogen-bond constraints were added to the total constraints. They are to strengthen certain hydrogen bonds in which the bonding imino or amino protons were observed in NMR spectra in water. These include all imino and amino protons in G-C base-pairs and the imino protons in A-T base-pairs, except the one base-pair at the DNA terminals. The hydrogen-bond constraints for imino protons were set to the maximum force constants while, for amino protons, they were set to half of the maximum. These constraints are expected to contribute negligibly when the hydrogen-bonds are in place. However, they are to help the complementary bases find each other from a very distorted initial DNA structure or to stay together at elevated temperatures in restrained molecular dynamics.

All molecular dynamics and energy minimization calculations were carried out by using the molecular modeling software BIOGRAF (BioDesign). The forcefield used was Dreiding-II (Mayo et al., 1990), which consisted of the usual terms for bonds, bond angles, inversions, improper torsions, van der Waals and electrostatic interactions.

Hydrogen bond potential was Lennard-Jones. Long-range cutoff distance for nonbond interactions was 9.0 Å. All hydrogen atoms were treated explicitly. No solvent molecules or counter ions were added. Integration of Newton's equations of motion was performed by use of a Verlet integration algorithm (Verlet, 1967). The time step of the integrator was 0.001 ps, and the nonbond interactions were updated every 0.1 ps. Initial velocities were assigned at the beginning of each session of molecular dynamics from a Maxwellian distribution at chosen temperatures. Velocities were scaled to the bath temperature whenever the calculated temperature was 50 K off the setting values.

The structural refinements were started with structures Init_I, Init_II and Init_III. Init_I is an idealized B-DNA built with crystallographic data. Init_II was generated by disturbing the B-DNA structure in an MD run of 1.5 ps at 600 K. Init_II is a further disturbed structure in 3.0 ps MD at 600 K.

The restrained molecular dynamics was conducted in the following sequences: (i) four sessions of molecular dynamics, 0.5 ps each at 300 K with the force constants of constraints scaled down to maximum values of 1, 2, 5, and 10 kcal/mol/Å²; (ii) 2 ps of molecular dynamics at 600 K for the molecule to equilibrate and search sufficiently large conformational space; (iii) the molecule was allowed to cool slowly to 300 K over a period of 2 ps, with additional 2 ps to equilibrate at 300 K; (iv) several sessions of straight molecular dynamics were performed at 300 K with time periods of 4 ps each, up to 20 ps total, or until the convergence were reached; (v) the converged structure was subjected to a final session

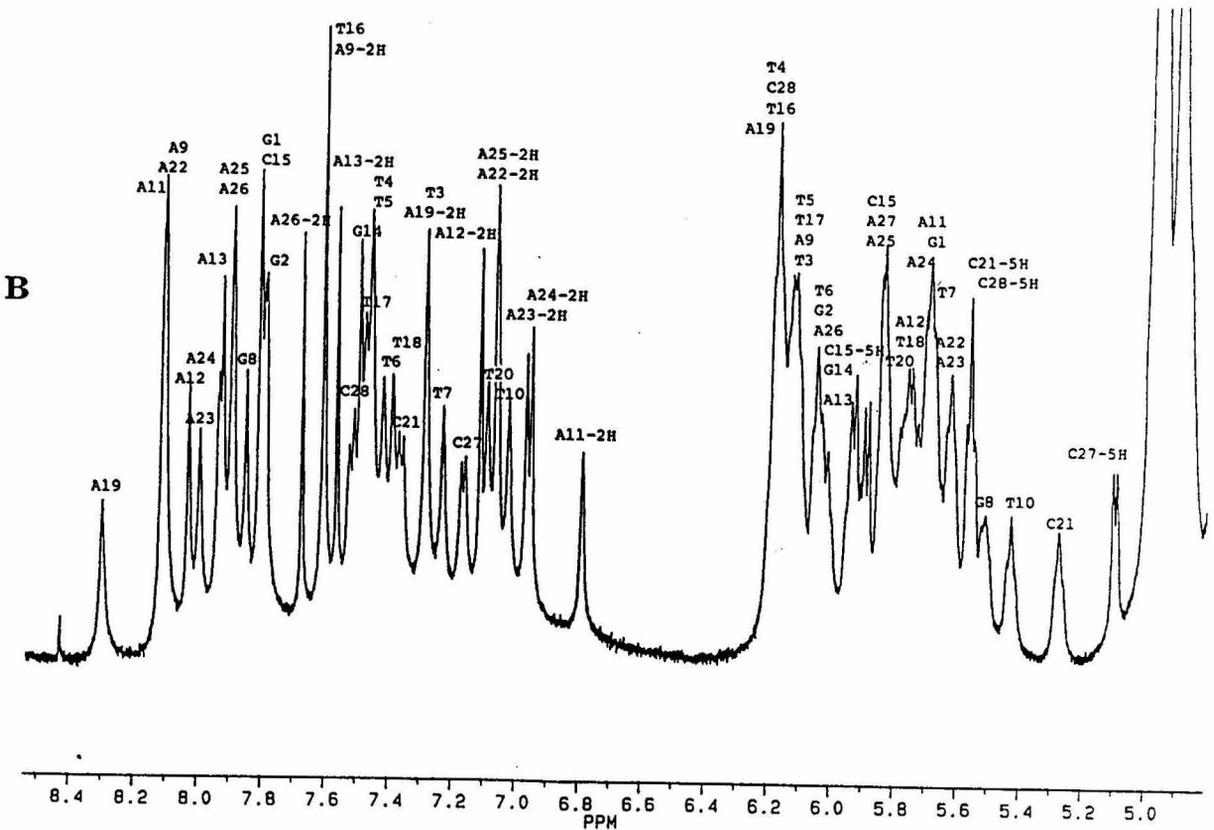
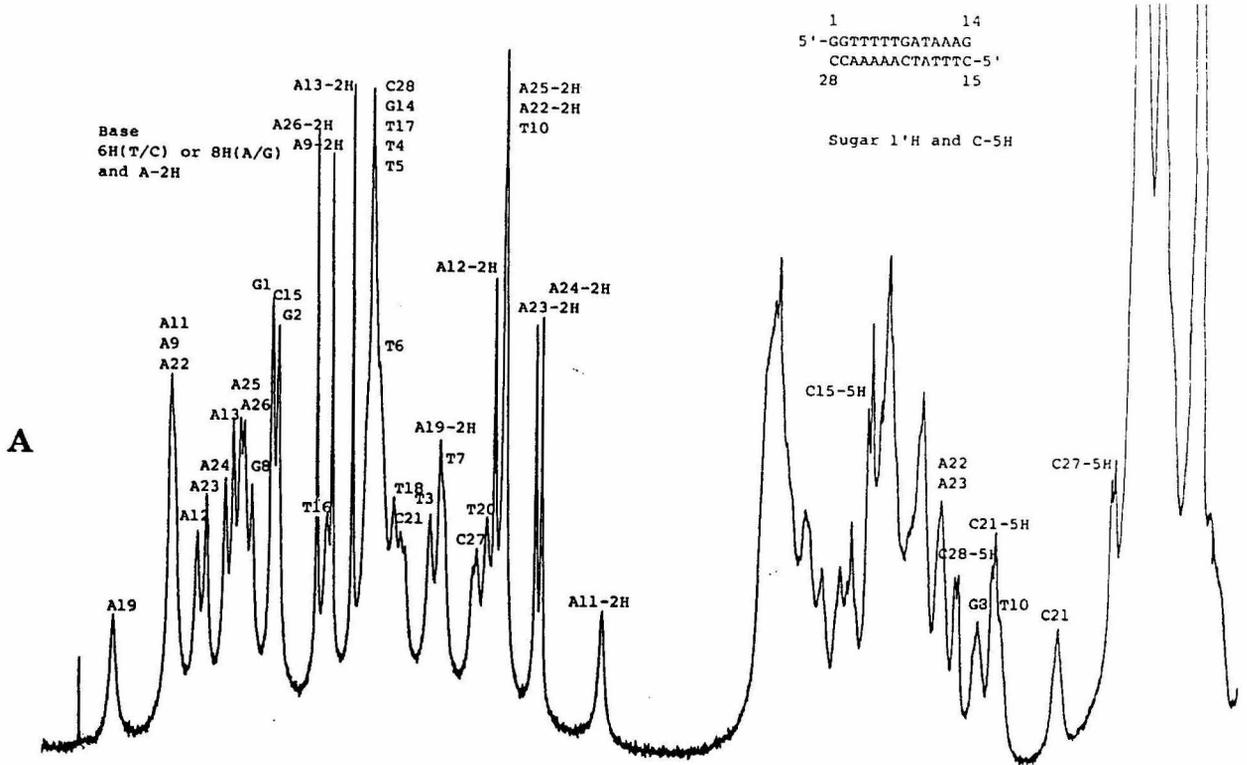
session of energy minimization (500 steps) using a conjugate-gradient algorithm. The convergence were judged by, firstly, RMS deviation of coordinates changes less than 1.0 Å in the last 4 ps molecular dynamics; secondly, RMS deviation of coordinates with the initial structure does not increase, nor decrease significantly (less than 0.5 Å in 4 ps). Unless specified, distance constraints scaled to a maximum of 10 kcal/mol/Å² were applied at all times. Structure refinements were conducted on a dedicated DEC VAXworkstation 3500. Restrained molecular dynamics were calculated at an average speed of 3.5 hr per picosecond of dynamics simulation for a 14-bp DNA. Graphics display is on an Evans & Sutherland PS 390 system.

RESULTS AND DISCUSSIONS

Resonance Assignments

The proton resonances of the DNA II were surprisingly well-resolved, especially at elevated temperatures. Figure 2-3 shows increased spectral resolution with temperatures 25°C, 35°C and 45°C. Spectra at 35°C seem to offer sufficient resolution for full resonance assignment and structural determination without compromising too much on lowered NOE intensities and an increased amount of internal motions.

The complete sequential assignment (Wüthrich, 1986) of ¹H resonances in DNA II was done in a 300-ms NOESY at 35°C, with the aid of spectra at other temperatures to clear out some ambiguities. A long mixing-time was used for two reasons. Firstly, the elevated temperature



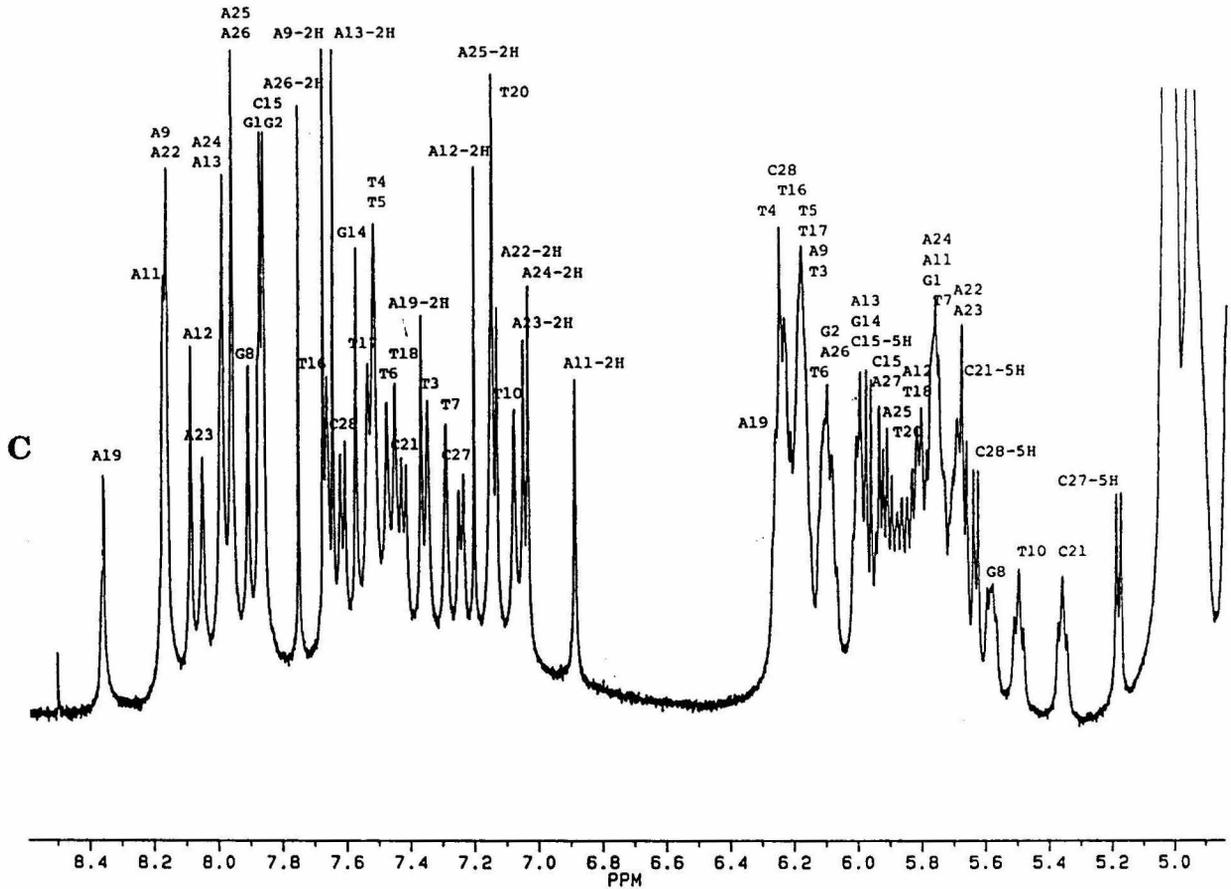


FIGURE 2-3: Conventional 1-D NMR of DNA II. Spectral regions containing base protons and sugar H1' are shown. The following resonances are present: H6, H8, H2, H1' and H5. Spectra obtained on a Bruker AM500 spectrometer. Sample conditions are 5.6 mM DNA II, 20 mM Pi, 120 mM NaCl, 0.1 mM EDTA, pH 7.0. A. 25°C, B. 35°C, C. 45°C.

has reduced NOE build-up rate significantly, possibly by shortening the correlation time; secondly, certain amounts of spin-diffusion were desirable, either to facilitate resonance assignments or to dilute the noise portions of small NOE intensities. The first spectra region to be examined was {H6/H8/H2-H1'/H5}, where pathways $H1'(i-1) \leftrightarrow H6/8(i) \leftrightarrow H1'(i)$ were followed sequentially for both strands (Figure 2-4). The process was greatly facilitated by utilizing characteristically strong and double-double coupled cytosine H6-H5 cross-peaks. Also helpful was simultaneous examination of the {H6/8-CH₃} region, where partial pathways of $H6/8(i-1) \leftrightarrow CH_3(i) \leftrightarrow H6(i)$ were followed. The assignments of base protons H6 of pyrimidines and H8 of purines were further confirmed in regions {H6/8-H2'/H2''} and {H6/8-H3'} where pathways $H2'/H2''(i-1) \leftrightarrow H6/8(i) \leftrightarrow H2'/H2''(i)$ and $H3'(i-1) \leftrightarrow H6/8(i) \leftrightarrow H3'(i)$ were available. Protons H5 of cytosine and CH₃ of thymine were subsequently confirmed. Tentative assignments of H1', H2', H2'' and H3' were checked in regions {H1'/H3''-H2'/H2''}. The chemical shift of H2' was normally larger than H2'' because of its approximation to the base ring, except at the 3' terminal where the difference of the two chemical shifts usually collapses. The distinctions between H2' and H2'' were confirmed from the relative intensities of intranucleotide NOE H1'-H2' and H1'-H2''. The H4' were assigned in region {H1'-H4'} by intranucleotide NOE interactions and further checked in region {H6/8-H4'}. Adenine H2 can be tentatively assigned based on its weak NOE cross-peaks with H1' of the 3'-neighboring bases on the home and the complementary strands. NOE

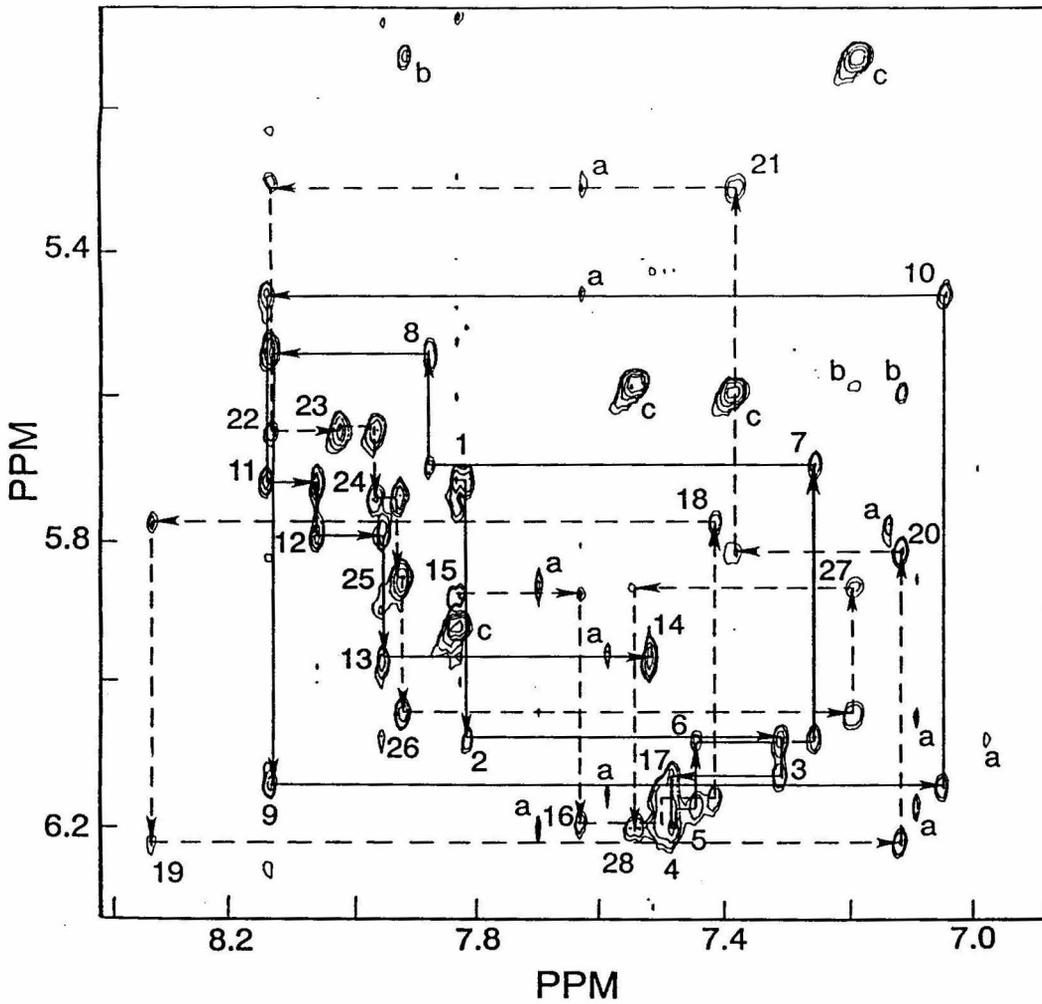


FIGURE 2-4: Assignment of base H6/H8 and sugar H1' of DNA II. The NOESY experiment was done at 35°C with a mixing time $\tau_m = 300$ ms. Spectral resolution is not enhanced. Sequential connectivities between H6/H8 and H1' are shown by the solid line for the strand 1 (residues 1-14) and by the dashed line for strand 2 (residues 15-28). Additional peaks labelled: (a) cross-peaks of adenine H2 to H1', interstrand-sequential or sequential; (b) sequential cross-peaks between the H5 base proton of cytosine and the H6/H8 proton of the preceding base; (c) intranucleotide H5-H6 NOE interactions in cytosine. Resonance assignments are given in Table 2-1.

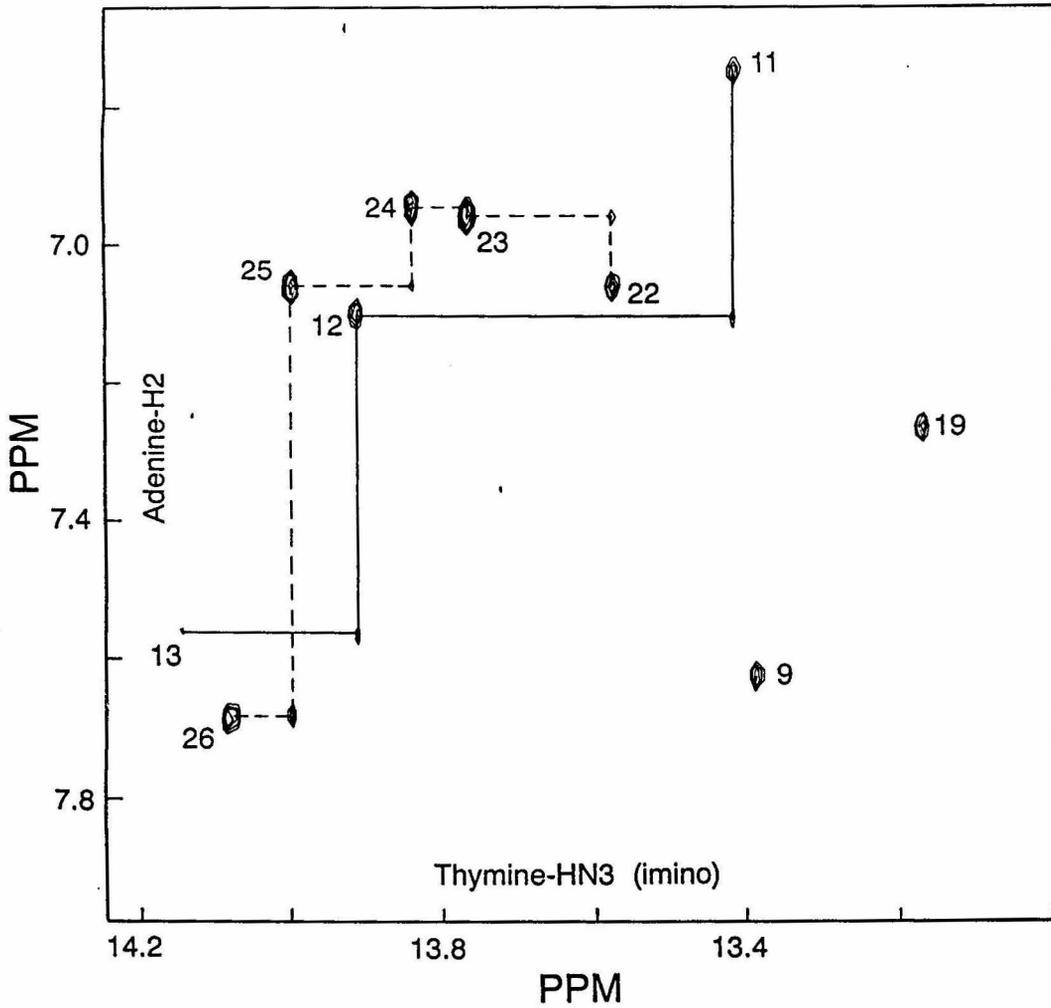


FIGURE 2-5: Assignment of adenine H2 and thymine imino-proton HN3 of DNA II. The NOESY experiment was done at 25°C in 9:1 H₂O/D₂O, $\tau_m = 200$ ms. Numbers in the figure indicate the adenine residue involved in forming hydrogen bond with the correspondent thymine imino-proton. Thus (11) indicates an NOE within a base-pair between H2 of Adenine 11 and HN3 of Thymine 18. The weak cross-peaks linking these numbered cross-peaks are interstrand-sequential NOE, e.g., the cross-peak linking (11) and (12) is arising from HN3 of Thymine 18 and H2 of Adenine 12. Solid and dashed lines indicate the adenines are on the first and second strand, respectively.

interactions $H2(i)$ - $H2(i+1)$ were useful in assigning $H2$, also. The assignments of adenine $H2$ were finalized in a NOESY of DNA II in H_2O by following the pathway: $H2(i) \leftrightarrow HN3$ (imino proton) of thymine $j \leftrightarrow H2(i+1)$, where A_i and T_j are in the same base-pair (Figure 2-5).

Resonance assignments were greatly facilitated by following the chemical shift changes with the temperature and careful bookkeeping of assigned resonances. Constant cross-checking of resonances in various regions has been proven essential for the assignment of some "difficult" resonances.

The coordinates of cross-peaks in different regions were checked against each other for assignment consistency by a routine in CROSSPEAK. This process ensured the correctness and accuracy of each assignment. The average of standard deviations of assignments is 0.003 ppm. At this resolution, all 28 nucleotides are resolved. Nucleotide T4 and T5 were the least resolved, their base protons ($H6$) and most of other protons were almost completely overlapped. Fortunately, their sugar $H1'$ were well-resolved, which sorted out other proton assignments and gave several precious sequential NOE between T4 and T5.

Assignments of nonlabile protons in DNA II are listed in Table 2-1. $H5'$ and $H5''$ were not assigned.

Characteristic NOE Patterns Defining an Overall B-form DNA Structure

Qualitative assessment of the relative NOE cross-peak intensities measured at short mixing times (50 ms, 100 ms) have given patterns that

Table 2-1: Proton Resonance Assignments for DNA II^a

Residue	H8/H6	H5/CH3/H2	H1'	H2'	H2''	H3'	H4'
G 1	7.833		5.712	2.554	2.706	4.808	4.178
G 2	7.815		6.071	2.635	2.844	4.969	4.429
T 3	7.309	1.377	6.123	2.181	2.650	4.886	4.305
T 4	7.485	1.605	6.199	2.248	2.673	4.924	4.241
T 5	7.482	1.644	6.162	2.223	2.661	4.917	4.240
T 6	7.445	1.643	6.072	2.124	2.580	4.906	4.197
T 7	7.254	1.650	5.692	1.995	2.371	4.876	4.096
G 8	7.877		5.540	2.680	2.751	4.994	4.339
A 9	8.135	7.631	6.137	2.544	2.831	4.968	4.405
T 10	7.048	1.385	5.454	1.827	2.192	4.793	4.047
A 11	8.141	6.812	5.715	2.561	2.632	4.996	4.321
A 12	8.058	7.138	5.785	2.565	2.775	5.010	4.383
A 13	7.952	7.589	5.966	2.502	2.789	4.969	4.383
G 14	7.519		5.949	2.248	2.315	4.567	4.148
C 15	7.832	5.914	5.870	2.252	2.598	4.656	4.096
T 16	7.633	1.682	6.189	2.274	2.625	4.914	4.278
T 17	7.503	1.675	6.147	2.200	2.632	4.910	4.224
T 18	7.415	1.730	5.769	2.181	2.533	4.923	4.185
A 19	8.326	7.310	6.214	2.661	2.903	5.011	4.427
T 20	7.116	1.345	5.806	1.947	2.345	4.797	4.130
C 21	7.386	5.589	5.304	1.818	2.159	4.754	4.002
A 22	8.134	7.081	5.648	2.643	2.733	4.977	4.342
A 23	8.023	6.992	5.647	2.527	2.707	4.980	4.291
A 24	7.965	6.975	5.737	2.505	2.785	4.984	4.366
A 25	7.925	7.089	5.849	2.492	2.830	4.988	4.395
A 26	7.919	7.700	6.036	2.472	2.785	4.940	4.403
C 27	7.192	5.122	5.857	1.976	2.382	4.719	4.133
C 28	7.548	5.580	6.196	2.222	2.222	4.517	4.010

^aChemical shifts (ppm) are at 35°C, relative to 3-(trimethylsilyl)propionic acid (external standard). Assignments have been checked in all cross-peaks to which they apply. Standard deviations are averaged at 0.003 ppm.

are indicative of a right-handed B-type structure. The handedness was ascertained from consistent observations of the following NOE: H6/8(*i*-1)-H5/CH₃(*i*), H1'/H2'/H2''(*i*-1)-H5/CH₃(*i*) and (imino H of *i*)-H2(*i*+1) (Figure 2-5). The glycosidic bonds were in anti torsion angle and sugar puckers were in C2'-endo conformation owing to the following relative NOE intensities: H6/8-H2' >> H6/8-H1' > H6/8-H3'. The overall B-type conformation was established from the observation of the following characteristic intensity patterns: H2''(*i*-1)-H6/8(*i*) > H2'(*i*-1)-H6/8(*i*), H2'(*i*)-H6/8(*i*) > H2'(*i*-1)-H6/8(*i*) (Wüthrich, 1986).

It is important to establish that the DNA is in B-type conformation, a basic assumption which is the foundation for further structure refinement. We will use standard B-DNA as the starting coordinates for distance calculations. Similar B-type structures will also serve as the initial conformations for restrained molecular dynamics since the refined structures shall be in the vicinity of the B-DNA conformation.

Unusual features of the NMR data

An obvious feature in Figure 2-4 is the presence of many relatively strong NOE cross-peaks arising from adenine(*i*) H2 and H1' of the 3'-neighboring nucleotide on the same strand, H2(*i*)-H1'(*i*+1), i.e., the sequential H2-H1', and across the strand to the 3' neighbor of the complementary residue *j*, H2(*i*)-H1'(*j*+1), i.e., the interstrand-sequential H2-H1'. Such cross-peaks have been observed in other 2-D NMR studies of B-DNA (e.g., Wemmer et al., 1984; Weiss et al., 1984). They are usually very weak except in the oligo(dA) tract (Kintanar et al., 1987; Katahira et

al., 1990) and poly(dA)•poly(dT) (Behling & Kearns, 1986). DNA II has a sequence of abundant A•T base-pairs, including two tracts of oligo(dA). In fact, all adenine H2 in DNA II have been observed to give interstrand-sequential and (intrastrand) sequential H2-H1' in both NOESY spectra acquired with $\tau_m = 300$ ms (Figure 2-9) and 200 ms (Figure 2-10). Adenine H2 has also been observed, in some cases, to give NOE with the H1' of its complementary bases (interstrand intra-base-pair NOE), and occasionally with its own H1' (intranucleotide NOE). NOE of interstrand-sequential H2-H1' is almost always stronger than sequential H2-H1'.

It is known that from the coordinates of classical B-form DNA, one would not expect to see these NOE since there are no H1' protons or any other non-labile protons within 5Å of adenine H2 (Kintanar et al., 1987). Observations of these NOE cross-peaks suggest that there are substantial structural deviations from classical B-DNA structure. As has been observed and defined by R. Dickerson et al. (1981) from crystal structures, DNA molecules exist with extensive local deviations from classical B-DNA. The deviations involve variations of helical twist, roll angle, slide and propeller twist, etc. In oligo(dA) tracts, propeller twists have been especially pronounced, as seen in crystal structure (Nelson et al., 1987) and proposed from NMR studies of such tracts (Kintanar et al., 1987; Katahira et al., 1990). Strong propeller twists can presumably lead to the compression of the minor groove in B-form DNA. A narrowed minor groove may give an explanation of the NOE interactions between interstrand H2 and H1'. Katahira et al. (1990) further proposed a gradual compression mechanism to explain the asymmetric DNA-

bending phenomenon at 3' and 5' junctions of oligo(dA). Interesting enough, the measured interstrand distances H2-H1' of the two oligo(dA) tracts in DNA II (Table 2-2C and Table 2-3C) coincide well with the gradual compression scheme, that is, the interstrand distance decreases toward the 3' of the oligo(dA) tracts. However, those marker distances, characteristic as they are, are not enough to define a quantitative overall structure to molecular detail, as demonstrated later in the refined structure of DNA II.

Another unusual aspect of the NMR spectra was observed in the intensities of the sequential NOE cross-peaks between base protons H6/8 and sugar protons H1', H2' and H2". In idealized B-DNA, the intensities of cross-peaks arising from similar sets of protons are expected to be independent of the sequences, with H6/8-H2" being characteristically strong (distance = 2.1 Å). In the NOESY spectra of the DNA II, variations in these NOE intensities were observed, indicating some sequence-dependent structural deviations from idealized B-DNA. Relatively weak NOE intensities occur primarily at the pyrimidine-purine junctions, e.g. T7-G8 and C21-A22 (Figure 2-6). The anomalies are not as obvious as in DNA I (Chapter 3). In other NMR studies (Patel et al., 1987; Gochin & James, 1990), similar observations have been reported and attributed structurally to an unwinding of DNA helices at pyrimidine-purine junctions.

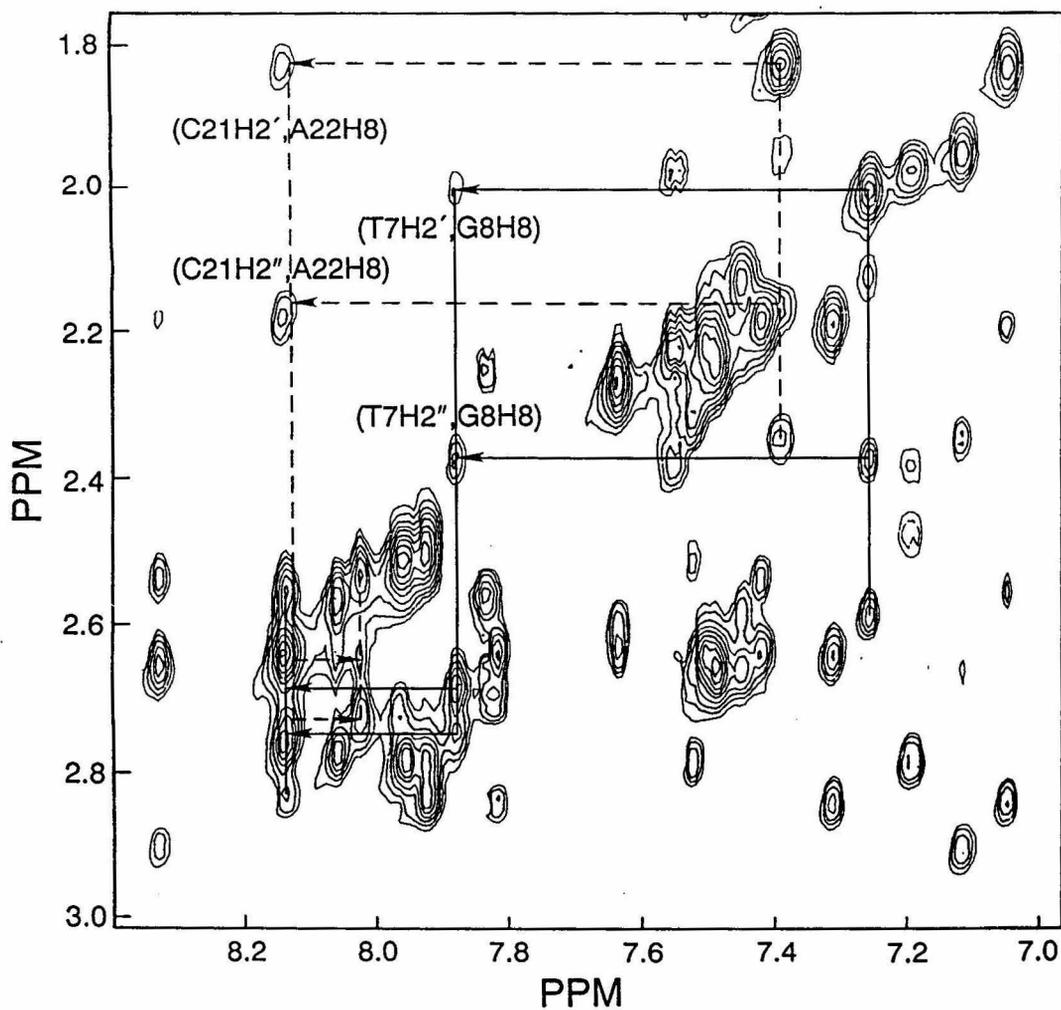


FIGURE 2-6: The 2-D NMR region showing NOE interactions between H2'/H2'' and H6/8 of DNA II. NOESY experiment was done with a mixing time $\tau_m = 100$ ms. Resolution of the 2-D spectrum was not enhanced. Sequential distances H2'-H8 and H2''-H8 are expected to be 3.8 and 2.1 Å, respectively. Much greater differentiations in intensities are, therefore, expected for the labelled cross-peaks. The anomalies, however, are not as pronounced as in DNA I (Figure 3-3).

NOE Cross-peak Intensities

NOE intensities were determined individually by volume integrating cross-peaks followed by an automatic processing procedure in which cross-peaks are checked with each other for overlapping. The overlapping fractions of the intensities were corrected by the procedure as described in MATERIALS AND METHODS. Numbers of valid intensities obtained from 300-ms and 200-ms NOESY experiments of the DNA II were 472 and 433, respectively. This constitutes about 76% and 70% of all observable interproton distances ($<6\text{\AA}$) within variations of a right-handed double-helical DNA II.

To check the self-consistency of the original intensity data, we integrated and compared intensities from upper and lower diagonal regions of a NOESY spectrum, which is, in theory, symmetric across the diagonal. The intensities and the derived distances are plotted against each other in Figure 2-7 and Figure 2-8, respectively. The root mean square difference (RMSD) of intensities between two measurements was 15% and RMSD of distances was, consequently, 0.23 Å. The reason for the skewing data points in Figure 2-7 is likely the limited digital data-points along the t_1 dimension, causing nonequivalent linewidths and signal-sampling of a resonance in two time domains. Normally, intensities were taken from one side of the diagonal (upper diagonal regions), data consistency can be expected higher. On the other hand, some of the weak intensities like H2-H1' were not evaluated here because they were simply

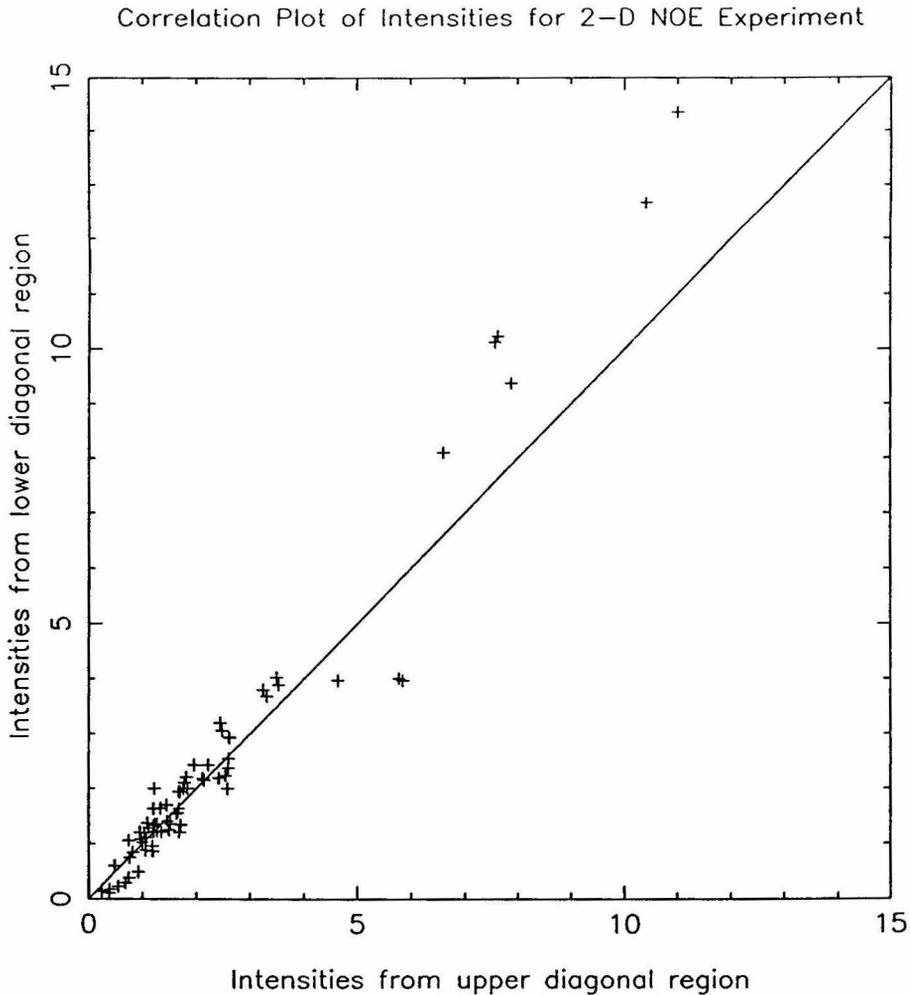


FIGURE 2-7: Consistency of cross-peak intensities from NOESY experiment. 2-D NOE experiment was done at 35°C with $\tau_m = 300$ ms. Intensities are direct integrations of cross-peak volumes in base H6/8-sugar H1' region without correction for intensity overlap. Each point in the figure is plotted using upper-diagonal intensity as the abscissa, lower-diagonal intensity as the ordinate. A perfect correlation implies all points residing on the diagonal line. Average differences of intensities is -0.124, root mean square (RMS) differences of intensities is 0.732, RMS relative differences of intensities is 15.2%.

Correlation Plot of Distances from 2-D NOE Experiment

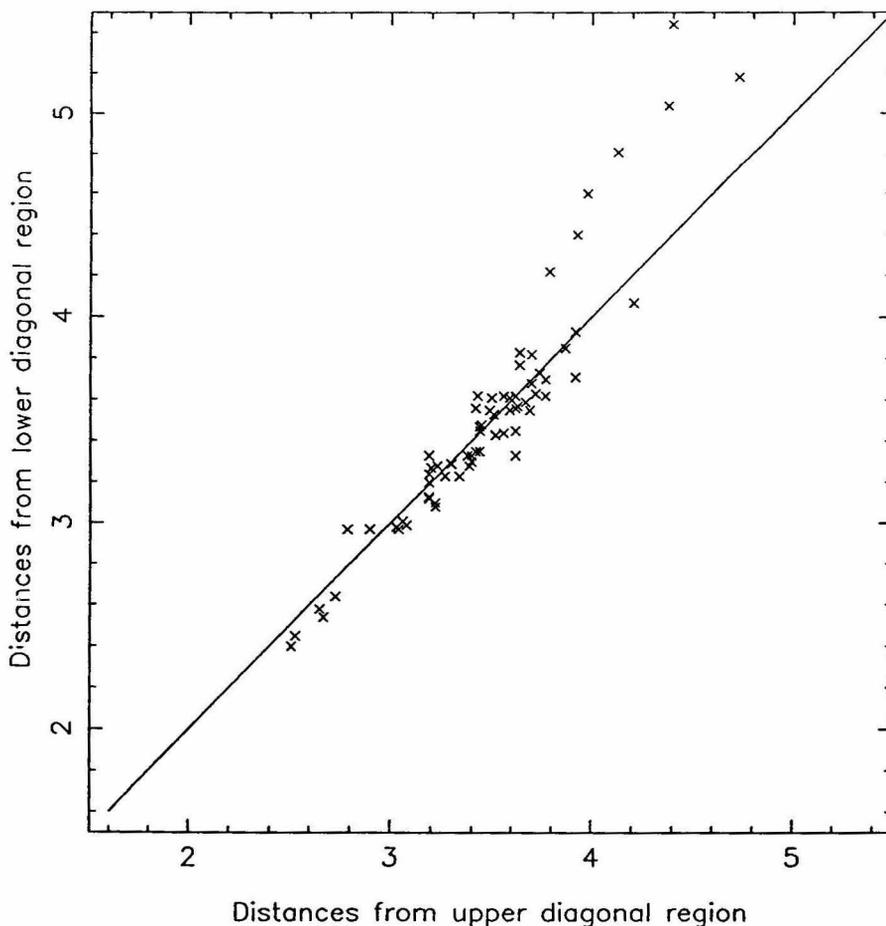


FIGURE 2-8: Consistency of distances derived from cross-peak intensities across the diagonal. The intensities were obtained as specified in Figure 2-7. Distances were derived using Eq.2-1 with the cytosine H5-H6 fixed distance (2.45 Å) as the reference. The figure is a correlation plot as described in Figure 2-7. Average differences of distances is -0.03 Å, RMS differences of distances is 0.226Å, RMS relative differences of distances is 2.67%.

not seen in the lower-diagonal region. For more crowded spectral regions, the consistency is expected to be worse. Overall, we think the estimation is a reasonable measure of our original data consistency.

Distance Constraints

Intensity sets of DNA II from two NOESY experiments with $\tau_m = 300$ and 200 ms were submitted to distance calculations ($\tau_c = 2$ ns), resulting in 424 and 371 distance constraints, respectively. This represents about 68% and 60% of all interproton distances possibly shorter than 5 Å in the molecule. Excluded from the constraint lists were all distances greater than 5 Å, fixed distances in the molecule (H6/8-H5/CH₃/H2 and H2'-H2"), and distances corresponding to inconsistent intensities. Figure 2-9 and Figure 2-10 summarize the statistics of distance constraints, categorized into intranucleotide, sequential (internucleotide) and interstrand distances. Two complete sets of distances from NOESY experiments $\tau_m = 300$ ms (Constraint #1) and 200 ms (Constraint #2) are listed in Table 2-2 and Table 2-3.

A brief survey of the constraint statistics reveals that distances are fairly evenly-distributed. There are at least two sequential distances in between each base step in all cases except for step A22-A23 in Constraint #2, where only one sequential distance is available. In general, steps leading to a pyrimidine have more sequential constraints due to additional constraints involving H5/CH₃. Sequential distance constraints are important because they impose greater restraints on conformation and thus they are structurally more informative.

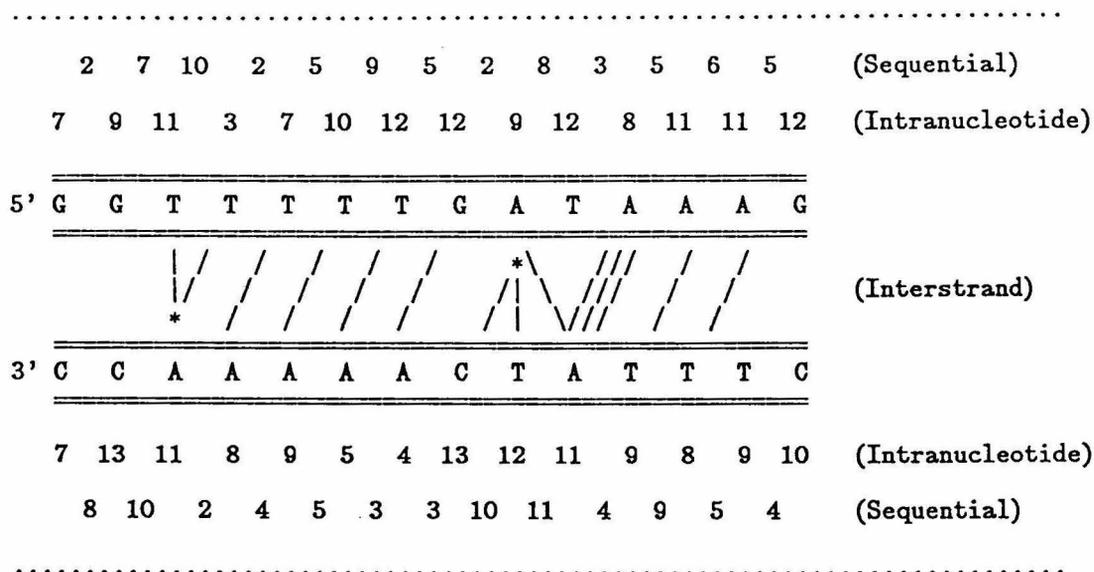


FIGURE 2-9: Statistics of distance Constraint #1. NOESY experiment was performed with $\tau_m = 300$ ms. Distances were calculated using the full spin matrix analysis as described in the MATERIALS AND METHODS. Number of intranucleotide distances for each residue is indicated near the residue, sequential distances placed in between two residues. Interstrand distances are indicated by connection lines between the strands. Total number of intranucleotide distances is 263, sequential 147, interstrand 14. Total distance constraints, 424.

Table 2-2: Distance Constraint Set #1^a

A. Intranucleotide distances

Strand #1: (1-14)

i	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
H1'	H8/H6	3.1	3.4	3.0		3.3	3.2	3.0	3.1	3.0	3.0	3.2	3.2	3.0	3.1
H2'	H8/H6			2.6		2.8	2.6	2.4	2.6		2.4			2.3	3.0
H2 ⁿ	H8/H6		2.7				2.8	2.9	2.7	3.1	2.7		2.3		2.4
H3'	H8/H6	3.2	3.4	3.0				3.0	3.3		3.0			3.0	3.0
H4'	H8/H6		4.6	3.8				3.8	4.2	4.0	4.0	4.3	3.9	3.4	4.5
H2	H1'									4.3			4.1		
H1'	H2'		2.8	3.2		2.8	2.8	3.1	3.0	2.7	2.7		2.5	2.7	2.2
H1'	H2 ⁿ	2.1	2.2	2.5		2.3	2.4	2.3	2.5	2.3	2.3	2.6	2.2	2.2	2.9
H1'	H3'		3.3	3.2	3.2	3.1	3.7	3.2	4.5	3.4	3.4	3.1	4.0	3.4	3.4
H1'	H4'	2.7	2.9	2.8	2.5	2.6	2.7	2.7	3.0	3.1	2.7	3.4	2.7	2.8	2.8
H2'	H3'	2.3		3.0	2.7	2.5	2.8	2.8	2.6		2.7	2.7	2.5	2.8	2.6
H2 ⁿ	H3'	2.5		2.7			2.9	2.8	2.2		2.7	2.6	2.3	3.6	2.5
H3'	H4'	2.5	2.5	2.6			2.6	2.6	2.5	2.9	2.6	2.5	2.5	2.5	2.6

Strand #2: (15-28)

i	i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
H1'	H8/H6	3.2	3.4	2.9	3.1	3.6	3.0	3.0	3.4		3.1		2.9	3.0	3.1
H2'	H8/H6			2.4		2.9	2.5	2.3		2.5				2.4	
H2 ⁿ	H8/H6				2.8	2.9	2.7	2.6			2.8			3.1	
H3'	H8/H6	3.0	3.2		3.2	3.5	3.1	3.0					3.2	3.0	2.7
H4'	H8/H6	3.2	4.6	3.7		4.7	4.2	3.9			4.6	3.6	3.4	4.0	3.5
H2'	H5/CH3							3.0						3.3	3.6
H2	H1'									4.2	4.0	3.9	4.2		
H1'	H2'	2.7	2.8	2.6	2.8		2.6	2.8	3.0	2.5		2.5	2.5	2.6	
H1'	H2 ⁿ	1.7	2.2	2.3	2.2	2.4	2.4	2.2			2.1	2.3	2.3	2.0	
H1'	H3'	3.7	2.9	3.1	4.2	3.3	3.4	3.6			3.6	4.5	3.4	3.8	3.8
H1'	H4'	2.5	2.5	2.5	2.4	3.1	3.1	2.8	2.9	3.3	2.7	2.8	2.9	2.6	2.8
H2'	H3'	2.5	2.5			2.5	2.8	2.7					2.7	2.7	
H2 ⁿ	H3'	2.9			2.9	2.7	2.9	2.7			2.4	2.7	2.8	2.9	
H3'	H4'	2.7	2.4	2.5	2.3	2.6	2.6	2.5	2.9	2.5	2.7	2.5	2.7	2.6	2.6

 B. Sequential distances

Strand #1: (1-14)

i	i+1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14
H1'	H8/H6	2.9	2.9	3.0		3.2	3.0	3.5	2.7	3.0	3.1	2.8	2.9	3.0
H2'	H8/H6			2.8		2.8	3.0	3.6		3.1			3.1	3.3
H2 ⁿ	H8/H6		2.6			2.6	2.9	2.8		2.7	3.0	2.9		2.6
H3'	H8/H6	3.5	4.3	3.7			3.8	3.7		3.7	3.3		4.2	3.3
H1'	H5/CH3		3.9	3.3	3.6	3.5	3.7			3.6				
H2'	H5/CH3		3.0	3.0			3.1			3.5				
H2 ⁿ	H5/CH3			3.2			3.4							
H3'	H5/CH3			4.2			4.4							
H2	H1'									3.6		4.2	3.8	
H8/H6	H8/H6		3.7	3.3		3.5	3.4	3.6	4.2			3.4	4.0	3.6
H8/H6	H5/CH3		3.2	3.2						3.3				
H5/CH3	H5/CH3			4.2	4.5									
H2	H2											3.4	3.4	

Strand #2: (15-28)

i	i+1	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24	24-25	25-26	26-27	27-28
H1'	H8/H6		3.0	3.1	3.3	3.0	3.1	3.4		2.7	2.8		2.8	
H2'	H8/H6		2.7		4.4	3.3	3.0		3.2				2.7	3.1
H2 ⁿ	H8/H6			2.7	3.3	2.7	2.8	3.1		2.8			2.6	2.5
H3'	H8/H6	3.1		3.1		4.0	4.1	4.4					3.4	3.8
H1'	H5/CH3		3.3	4.4		3.7	3.8						4.5	
H2'	H5/CH3			2.8		3.2	2.9						3.0	3.0
H2 ⁿ	H5/CH3			3.4		3.6	2.9						3.0	4.9
H3'	H5/CH3	3.0	3.6	3.7		4.5								
H2	H1'					4.1				4.3	4.0	3.5	3.2	
H8/H6	H8/H6		3.6	3.2	3.7	3.6	3.4		3.8	3.4	3.4		3.5	3.4
H8/H6	H5/CH3	3.1		3.2		3.4	3.3						3.1	3.4
H5/CH3	H5/CH3	3.7					3.7							3.6
H2	H2								3.5	3.3	3.4	3.3		

C. Interstrand distances

Strand #1: (1-14)

i	j dpi	1-	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	14-
		28	27	26	25	24	23	22	21	20	19	18	17	16	15
H2	H1'	3.8													

i	j dps(-)	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	14-	
		28	27	26	25	24	23	22	21	20	19	18	17	16	
H2	H1'	3.5													
H2	H2	4.8 3.5 3.6 3.2													

i	j dps(+)	1-	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	
		27	26	25	24	23	22	21	20	19	18	17	16	15	
H2	H2	3.8													

Strand #2: (15-28)

i	j dpi	15-	16-	17-	18-	19-	20-	21-	22-	23-	24-	25-	26-	27-	28-
		14	13	12	11	10	9	8	7	6	5	4	3	2	1
H2	H1'	4.5													

i	j dps(-)	16-	17-	18-	19-	20-	21-	22-	23-	24-	25-	26-	27-	28-	
		14	13	12	11	10	9	8	7	6	5	4	3	2	
H2	H1'	4.4													
		5.0 4.0 3.5 3.4 3.5													

^aNOESY experiment was performed with $\tau_m = 300$ ms. Distances (\AA) were calculated using the full spin matrix analysis as described in the MATERIALS AND METHODS. In table C, referring to interstrand distances, d_{pi} denotes distances within a base-pair; d_{ps} denotes interstrand-sequential distances, where (-) indicates a distance to 3'-neighboring nucleotide of the complementary base, and (+), to 5' (shorthand by Wüthrich, 1986). A statistics of the distances is available in Figure 2-9.

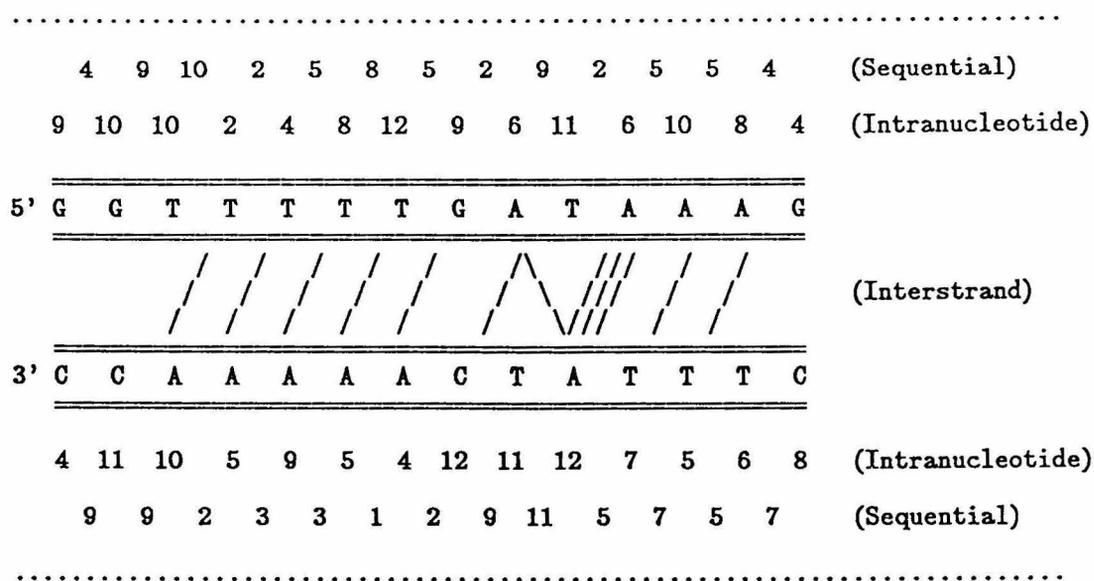


FIGURE 2-10: Statistics of distance Constraint #2. NOESY experiment was done with $\tau_m = 200$ ms. Total number of intranucleotide distances is 218, sequential 143, interstrand 10. Total number of all distance constraints, 371.

Table 2-3: Distance Constraint Set #2^a

A. Intranucleotide distances

Strand #1: (1-14)

i	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
H1'	H8/H6	2.9	3.0	2.9		3.0	3.1	3.2	3.0	2.7	2.8	2.8	3.4	2.9	
H2'	H8/H6	2.4	2.5	2.3		2.5	2.4	2.1	2.5		2.2			2.1	
H2 ⁿ	H8/H6	3.3	2.6				2.6	2.5	2.8		2.7		2.2		
H3'	H8/H6	3.4	3.4	2.8				2.9	3.5		3.0			3.5	3.3
H4'	H8/H6		4.2	3.8				4.6		4.3		3.8	4.8	3.3	
H2	H1'									3.7					
H1'	H2'		2.9				2.5	2.4	3.1	2.6	2.5		2.7	2.6	
H1'	H2 ⁿ	2.1	2.1	2.5		2.3	2.3	2.1	2.2	2.2	2.1	2.9	2.1	2.1	
H1'	H3'	4.1	3.4	3.2	2.8			3.5			3.1		3.0	3.4	3.1
H1'	H4'	2.6	2.7	2.7	2.5	2.5	2.9	2.5	2.8	3.0	2.5	3.2	2.6	2.7	3.0
H2'	H3'	2.2		2.6			2.4	2.3	2.3		2.4	2.4	2.4		
H2 ⁿ	H3'	2.5		4.1			2.5	2.6	2.2		2.5	2.3	2.3		
H3'	H4'		2.5	2.3				2.3			2.3		2.3		2.5

Strand #2: (15-28)

i	i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
H1'	H8/H6		3.1	2.9	3.2	3.2	3.3	2.8	3.3		3.2		3.2	3.1	2.6
H2'	H8/H6			2.2		2.4	2.3	2.0		2.3	2.4				2.1
H2 ⁿ	H8/H6		2.8		2.6	2.7	2.6	2.4			2.8				2.9
H3'	H8/H6	2.2	2.9			3.3	3.3	2.8					3.6	3.4	2.6
H4'	H8/H6			4.2		4.9		3.8	4.7		3.5				
H2'	H5/CH3							3.0							3.5
H2	H1'					4.4				4.7	4.2		3.7		
H1'	H2'	2.2			2.5		2.5	2.6	2.7	2.8	3.3	3.8	2.5	3.2	
H1'	H2 ⁿ	1.9	2.0	2.1	2.0	2.2	2.2	2.0			2.0	2.3	2.2	1.9	
H1'	H3'	3.7	2.8			3.5	3.3	3.4			3.3	4.2	3.1		
H1'	H4'	2.4	2.5	2.5	2.3	2.7	2.9	2.7	2.7	3.3	2.9	2.8	2.8	2.4	2.6
H2'	H3'	2.1			2.3	2.6	2.5	2.4					2.4	2.5	
H2 ⁿ	H3'	3.3			2.8	2.5	2.6	2.5				2.9	3.3	2.8	
H3'	H4'	2.4				2.4	2.3			2.3			2.5	2.5	2.4

 B. Sequential distances

Strand #1: (1-14)

i	i+1	1- 2	2- 3	3- 4	4- 5	5- 6	6- 7	7- 8	8- 9	9- 10	10- 11	11- 12	12- 13	13- 14
H1'	H8/H6	2.6	2.6	2.8		2.8	2.9	3.1	2.5	2.8	2.8	2.7	2.6	
H2'	H8/H6	2.9		3.1		2.9	2.7	3.0		3.2			3.8	3.0
H2 ⁿ	H8/H6	2.8	2.4			2.7	2.5	2.7		2.5	2.9	3.5		2.6
H3'	H8/H6	3.9	4.7	3.3				4.5		4.1			3.7	3.4
H1'	H5/CH3		3.6	3.2	3.5	3.7	4.6			3.7				
H2'	H5/CH3		3.0	3.3	2.8	4.1	4.0							
H2 ⁿ	H5/CH3		3.6	3.3			4.1			3.7				
H3'	H5/CH3		3.7	3.7			3.8			4.0				
H2	H1'									3.2		3.7	3.6	
H8/H6	H8/H6		4.0	4.2			3.9	3.7	4.5			3.9		3.5
H8/H6	H5/CH3		3.2	3.0						3.1				
H5/CH3	H5/CH3			5.0										
H2	H2											3.8	3.6	

Strand #2: (15-28)

i	i+1	15- 16	16- 17	17- 18	18- 19	19- 20	20- 21	21- 22	22- 23	23- 24	24- 25	25- 26	26- 27	27- 28
H1'	H8/H6	3.4	2.6	2.8	3.0	2.8	2.7	4.0		2.7	2.8		2.6	3.2
H2'	H8/H6		2.3		3.6	3.0	2.7		4.1	2.7			2.6	2.3
H2 ⁿ	H8/H6	2.3		2.5	2.6	2.4	2.5	2.7		3.1			2.4	2.5
H3'	H8/H6	2.8			4.5	4.0	4.4						3.8	3.4
H1'	H5/CH3		3.3	3.8		3.7								4.1
H2'	H5/CH3			3.3		3.3	3.0						3.3	2.9
H2 ⁿ	H5/CH3	4.3	3.3	3.3		4.0	2.9						3.0	3.1
H3'	H5/CH3	3.1	3.3	3.5		4.7								
H2	H1'					3.6					3.9	3.4	2.9	
H8/H6	H8/H6				4.1	3.7	4.1						4.8	
H8/H6	H5/CH3	2.7		3.0		3.2	2.9						2.9	3.0
H5/CH3	H5/CH3	4.6				4.4								3.8
H2	H2										3.8	3.5		

C. Interstrand distances

Strand #1: (1-14)

i	j	dps(-)	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	14-
H2	H1'		28	27	26	25	24	23	22	21	20	19	18	17	16
H2	H2									4.1		4.4	3.2	3.3	
												3.4			

i	j	dps(+)	1-	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-
H2	H2		27	26	25	24	23	22	21	20	19	18	17	16	15
										4.7					

Strand #2: (15-28)

i	j	dps(-)	16-	17-	18-	19-	20-	21-	22-	23-	24-	25-	26-	27-	28-
H2	H1'		14	13	12	11	10	9	8	7	6	5	4	3	2
											3.9	3.4	3.4	3.3	

^aNOESY experiment was performed with $\tau_m = 200$ ms. A statistics of the distances is available in Figure 2-10.

The two NMR experiments referred to here were done independently with an NMR sample of DNA II. Other than the difference in mixing time, one was collected with 1024 transients in the t_1 domain ($\tau_m = 300$ ms), the other with 512 transients (200 ms). Therefore, the two experiments produce 2-D spectra that have different linewidths and slightly-offset chemical shifts. In fact, cross-peak integrations and linewidth measurements have to be done independently. These two sets of constraints contain approximately the same numbers of sequential and interstrand distances. What are missing in the Constraint #2 are mostly, intranucleotide distances, specifically, sugar-sugar distances. This is probably due to slightly broadened resonances making certain intensities unresolvable in the deconvolution process. Such broadening affects sugar-sugar intensities more since they usually appear in the more crowded spectral regions.

We have planned these parallel experiments so they will give us a rigorous measure of how consistent our procedure is in producing distance constraints from NMR experiments. This consistency measure is evaluating not only the experimental accuracy of NOE intensities, but also the deconvolution procedure, the distance calculation method and every other step in the process. Knowing the distance consistency also give us a quantitative estimate of the distance accuracy, which is important in restrained molecular dynamics for determining the constraint force constants. In other words, the distance accuracy impose such an up-limit within which these NMR distances should be utilized and emphasized.

In Figure 2-11, distance deviations were plotted for all distance pairs between two constraint sets. Despite the noisy appearance, the data shows a distance RMSD of 0.37 Å and a constraint energy-weighted distance RMSD (E-weighted RMSD) of 0.32 Å. The later is a better measure of the distance accuracy, but to keep the convention, we will still use the unweighted RMSD. Comparing to the original cross-peak data consistency (0.23 Å), we conclude that the experiments and data processing system are working with sufficient reliability. The distance accuracy is given as the standard deviation of distances:

$$\sigma = (\text{distance RMSD})/\sqrt{2} = 0.26 \text{ \AA} \quad (\text{Eq. 2-6})$$

In Figure 2-12, two sets of distances are displayed onto two axes forming a correlation plot. As expected, much greater distance deviations occur at longer distances (3.5-5Å). Those long distance constraints were therefore applied to molecular dynamics with substantially-reduced force constants. The figure also indicates that short distances (2-3Å) appear longer from NOESY with a longer mixing-time (Constraint #1, $\tau_m = 300$ ms), likely an artifact not fully corrected by the calculations and originated from a reduction of these strong NOE's due to spin diffusions. This suggests there could be some fine-tuning in our distance calculations. Such a correction is expected to be quite minor.

It is known that the accuracy of the distances obtained with back-calculation methods increases with (a) fractions of experimental cross-

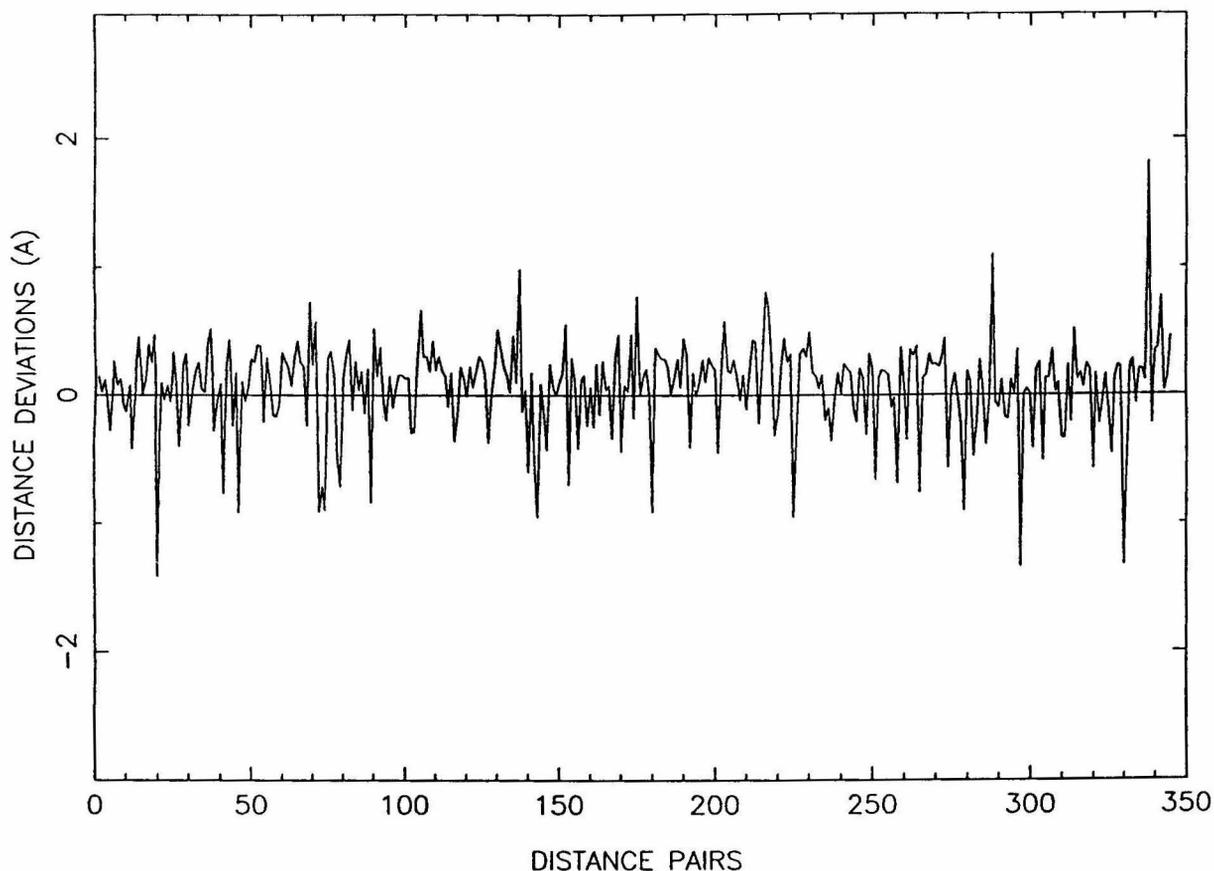


FIGURE 2-11: Distance Consistency: Deviation Plot. The difference between distances in a pair (distance in Constraint #1 minus distance in Constraint #2) is plotted against the pair-number. The distance RMSD is 0.37 \AA , implying a standard deviation $\sigma = 0.26 \text{ \AA}$. The constraint energy-weighted distance RMSD is 0.32 \AA , calculated by weighting each square difference with k_{ij}/k_{max} , before the summation step in RMSD calculations, where k_{ij} is the individual constraint force-constant and k_{max} is the maximum force constant ($10 \text{ kcal/mol/\AA}^2$).

CORRELATION PLOT OF DISTANCES FROM TWO NMR EXPERIMENTS

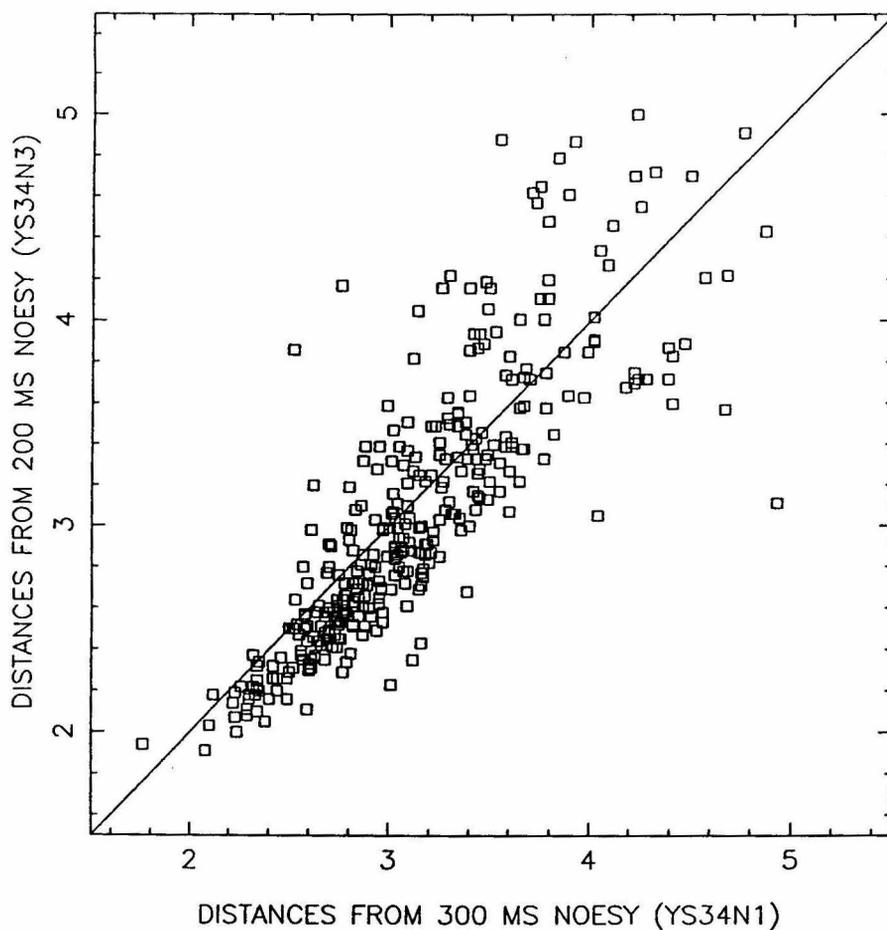


FIGURE 2-12: Distance Consistency: Correlation Plot. The distances in Constraint #2 (y-axis) are plotted against distances in Constraint #1 (x-axis). Greater deviations at longer distances have been expected (see text for more discussions).

peaks observed, (b) improved signal to noise, and (c) accuracy of initial model structure. In our studies, intensities were available for about 70-76% of all distances within 5 Å in the molecule; the cross-peak intensities were measured in good confidence as we collected intensities from NOESY experiments with sufficiently long mixing times such that they have grown much above the noise level; and the DNA molecule II has been defined, by qualitative analysis of NOE patterns, as to be in conformation very close to a B-form DNA, which could thus serve as an excellent starting structure. Therefore, the distances so derived should pertain good accuracy and define the DNA structure with sufficient confidence.

A future approach to utilize these constraint data is to convert multiple sets of distances to an average set of distances with individual standard deviations determined for all distances. The average distances are then supplied to molecular dynamics with the standard deviations converted to constraint force constants. In the present work, we decided to use in parallel two sets of distance constraints for structural refinements. This will be an even more rigorous test of the NMR structural determination method, as we will observe the differences in the refined DNA structures arising from independent NMR experiments.

When Constraint #1 was applied to an idealized B-form structure of DNA II, one of the initial structures for refinements, the initial distance violation was found to be 0.84 Å in distance RMS deviations (Figure 2-13), which is about 3.3σ . This is to say, the distance set is

INITIAL DISTANCE VIOLATIONS

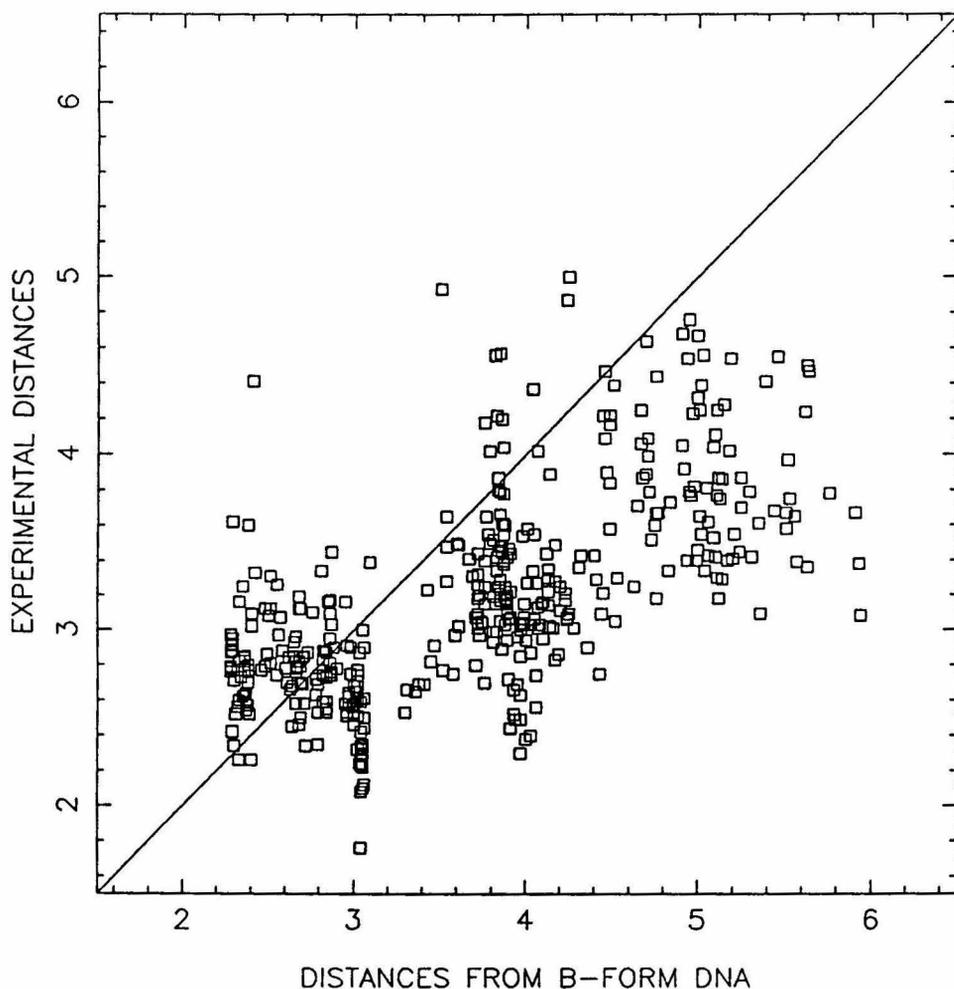


FIGURE 2-13: Initial distance violations of a structural refinement. The distances in Constraint #1 are plotted against correspondent distances in a B-form structure of DNA II, which will serve as one of initial structures for refinement and have been optimized by energy-minimization in the absence of distance constraints. The RMSD between two sets of distances, or the RMS distance violations, is 0.84 Å.

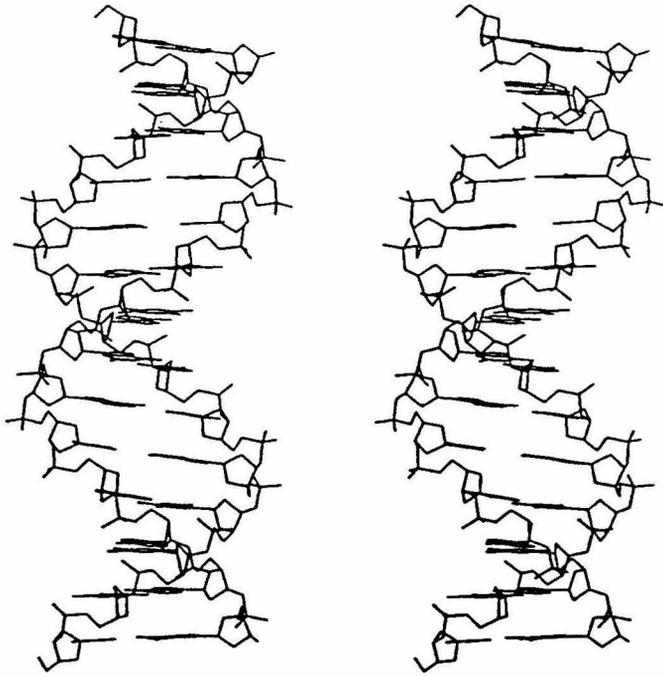
defining a structure quantitatively distinct from the idealized B-DNA and the uncertainties of the distance constraints are sufficiently small to give meaningful refinement of an initial structure such as idealized B-DNA.

Structural Refinement Using Restrained Molecular Dynamics (rMD)

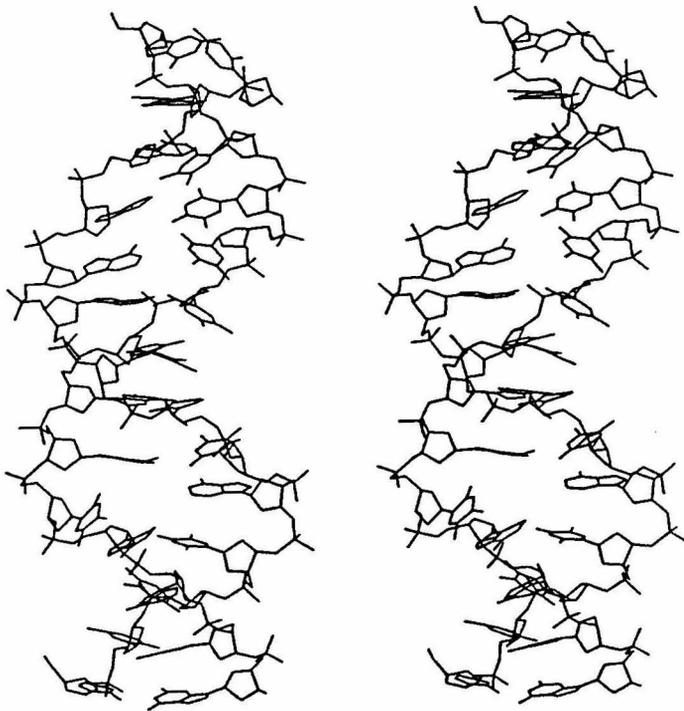
The structural refinements using restrained molecular dynamics were started with structures Init_I (B-DNA), Init_II and Init_III (disturbed B-DNA) (Figure 2-14), having an average atomic RMS difference of 2.8 Å. The rMD runs with Constraint #1 were C1_I, C1_II and C1_III; with Constraint #2, C2_I, C2_II and C2_III. An overall force constant of 10 kcal/mol/Å² was used for applying distance constraints, with reduced values for larger distances (>3.5Å); in C1_II, though, 20 kcal/mol/Å² was used. The atomic RMS differences between all pairs of structures are given in Table 2-4. Total potential energy and energy decompositions of all structures are listed in Table 2-5. The distance violation, defined as RMS differences between the calculated and experimental interproton distances, are presented in the first column of Table 2-5.

The progress of restrained molecular dynamics were monitored by the change in atomic coordinate RMSD. Figure 2-15 and Figure 2-17 shows the coordinate RMSD reach plateaus in the later stage of rMD, in reference to the initial structures. This was taken as an indication of convergence. The thermal fluctuation of the coordinates under the conditions of molecular dynamics is expected to be less than 1 Å. It

A



B



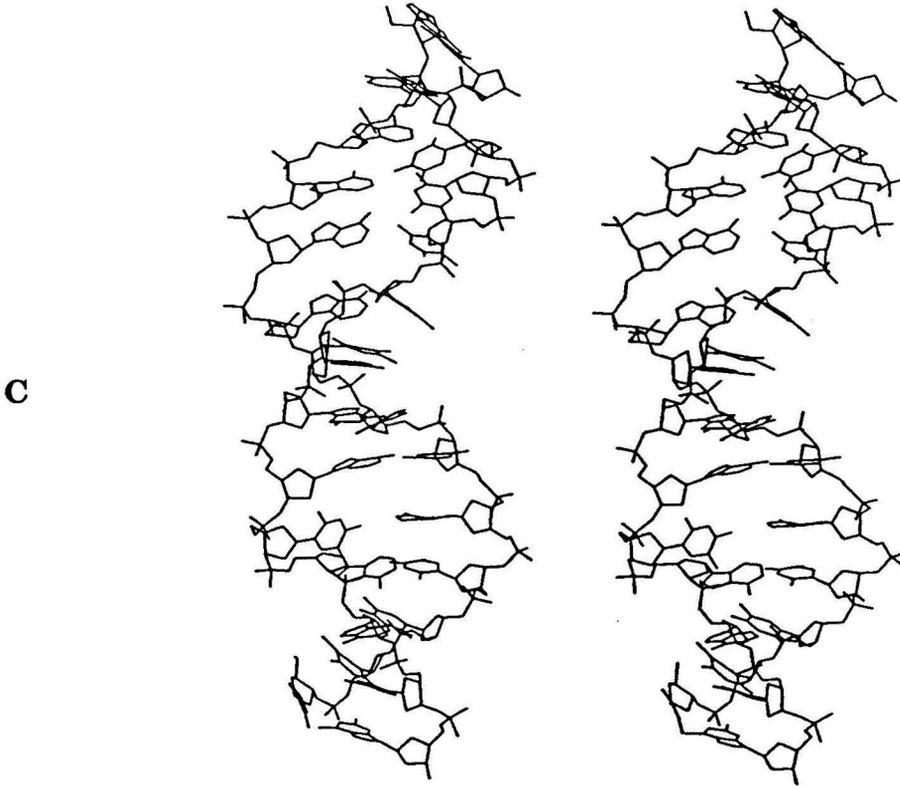


FIGURE 2-14: Initial structures for refinements of DNA II. **A.** Init_I, a classical B-DNA; **B.** Init_II, a disturbed B-DNA (atomic RMSD with Init_I, 2.24 Å); **C.** Init_III, a further disturbed B-DNA (atomic RMSD, 3.64 Å). Init_II and Init_III were generated by free molecular dynamics heating a B-DNA at 600 K for 1.5 ps and 3.0 ps, respectively.

Table 2-4: Atomic RMS difference (Å)^a

	Init_II	Init_III	C1_I	C1_II	C1_III	C2_I	C2_II	C2_III	R1	R2	R1/2	Final_II
Init_I	2.24	3.64	3.45	3.37	3.00	3.82	4.62	4.82	2.87	4.36	3.66	3.59
Init_II	-	2.49	3.74	3.48	3.14	4.11	4.83	4.98	3.36	4.57	3.87	3.85
Init_III		-	4.48	4.06	3.85	4.80	5.47	5.44	4.06	5.17	4.52	4.59
C1_I			-	1.43	1.66	1.20	1.89	2.28	0.96	1.62	0.80	0.77
C1_II ^b				-	1.15	2.15	2.56	2.96	0.67	2.42	1.42	1.48
C1_III					-	2.39	2.83	3.15	0.82	2.66	1.66	1.67
C2_I ^c						-	1.39	1.74	1.80	0.92	0.96	1.15
C2_II							-	1.52	2.31	0.77	1.36	1.57
C2_III								-	2.70	0.98	1.73	2.07
R1									-	2.12	1.06	1.08
R2										-	1.06	1.38
R1/2											-	0.64

^aRefinements were started with Init_I, Init_II or Init_III, using Constraint #1 or #2. C1_I, e.g., was started with Init_I and using Constraint #1. R1 is a refined structure from averaging C1_I, C1_II and C1_III; R2 is from C2_I, C2_II and C2_III. R1/2 is from averaging R1 and R2. Final_II is from energy minimization of R1/2 under Constraint #1. ^bUsing force constant 20 kcal/mol/Å². ^cConstraint distances were calculated using $\tau_c = 1$ ns.

Table 2-5: Constraint Violations and Energies^a

	RMS Violation	Total Energy	Bonds	Angles	Torsions	VDW	Electrost.	Hbond	Constraint
Init_I	0.839	8219	1003	2309	310	3990	-394	-243	1239
Init_II	1.288	5731	730	1790	392	600	-472	-101	2732
Init_III	1.320	5824	749	1712	399	622	-497	-115	2885
C1_I	0.664	1934	126	1164	238	522	-587	-235	702
C1_II ^b	0.607	1978	129	1235	251	618	-602	-234	575
C1_III	0.677	1952	123	1192	246	528	-621	-234	714
C2_I ^c	0.664	1959	127	1161	238	544	-586	-237	704
C2_II ^c	0.669	1916	127	1176	235	527	-617	-234	697
C2_III ^c	0.669	1955	127	1186	235	538	-599	-237	703
R1	0.630	8840	5034	2654	271	964	-474	-236	624
R2 ^c	0.649	11100	5262	2122	277	3474	-463	-239	664
R1/2	0.637	17873	8534	3806	274	5232	-393	-240	657
Final_II	0.666	1948	121	1162	233	522	-569	-233	708

^aConstraint violations are in Å, all other terms (energies) are in kcal. Inversion energy is not listed (<5kcal). VDW = van der Waals energy. ^bCalculations based on 10 kcal/mol/Å² for easy comparison. ^cViolations based on Constraint #2. Others are based on Constraint #1.

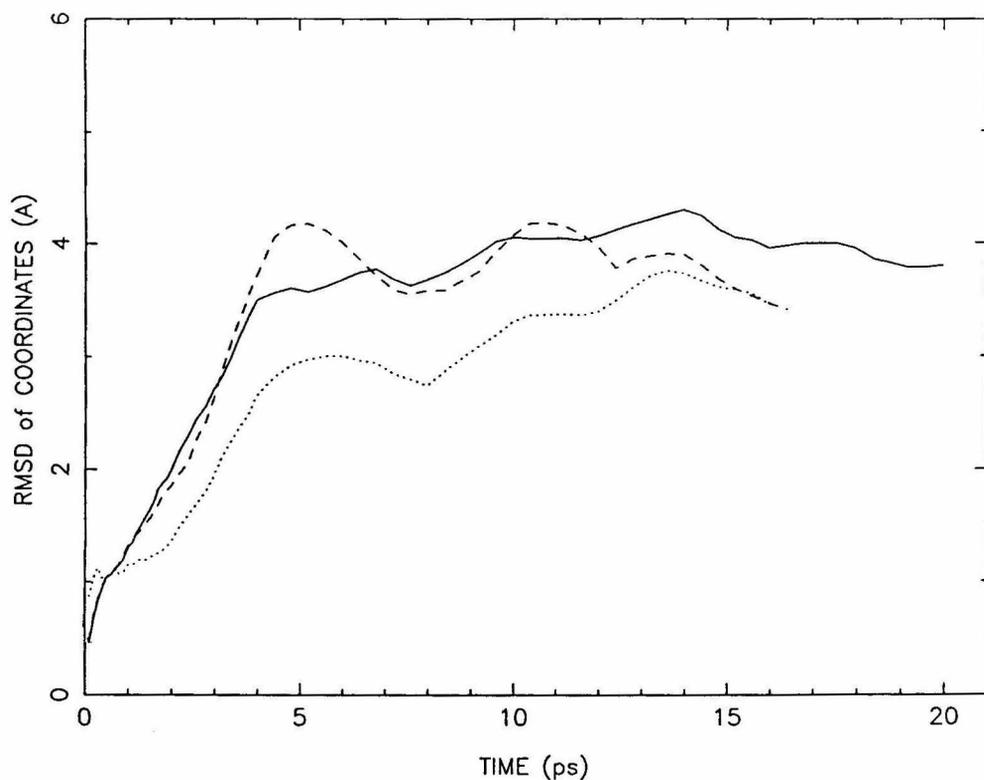


FIGURE 2-15: Coordinate changes in the course of the restrained molecular dynamics. For each curve, the atomic RMS differences between the structures in an rMD simulation and the initial structure are plotted against the time course. The *dotted* line: refinement C1_I starting with Init_I; *dashed* line: C1_II, starting Init_II; *solid* line: C1_III, starting Init_III. All refinements were done with the Constraint Set #1 (NOESY $\tau_m = 300$ ms) applied. Structures were sampled 0.1 ps for 0-2 ps period, 0.2 ps for 2-4 ps, and 0.4 ps for 4 ps-and-after. RMS differences reach plateaus at the later stages of rMD, implying equilibrating of structures. Final setting RMS differences of the curves have no bearings on each other, though.

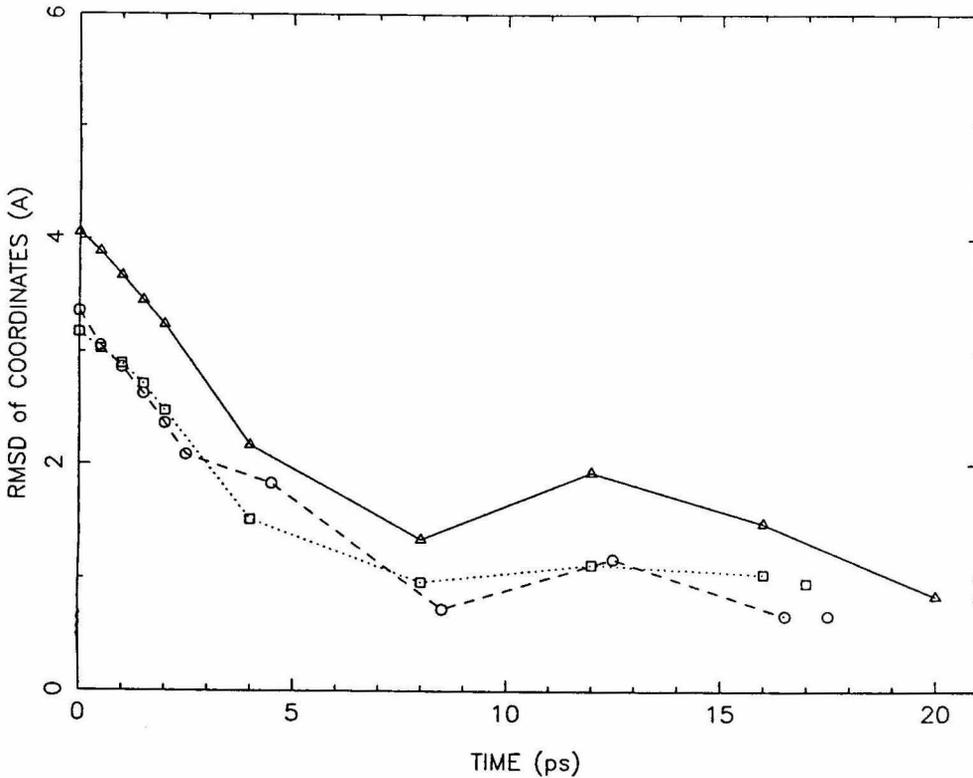


FIGURE 2-16: Atomic coordinate RMS differences in the course of the restrained molecular dynamics, relative to the average converged structure R1. Refinements were with Constraint #1. Structures were sampled at the end of each MD session. Data points were as shown. The *squares* and *dotted* line: refinement C1_I; the *circles* and *dashed* line: C1_II; the *triangles* and *solid* line: C1_III. The figure shows the atomic RMS differences converged to a radius of 1.7 Å.

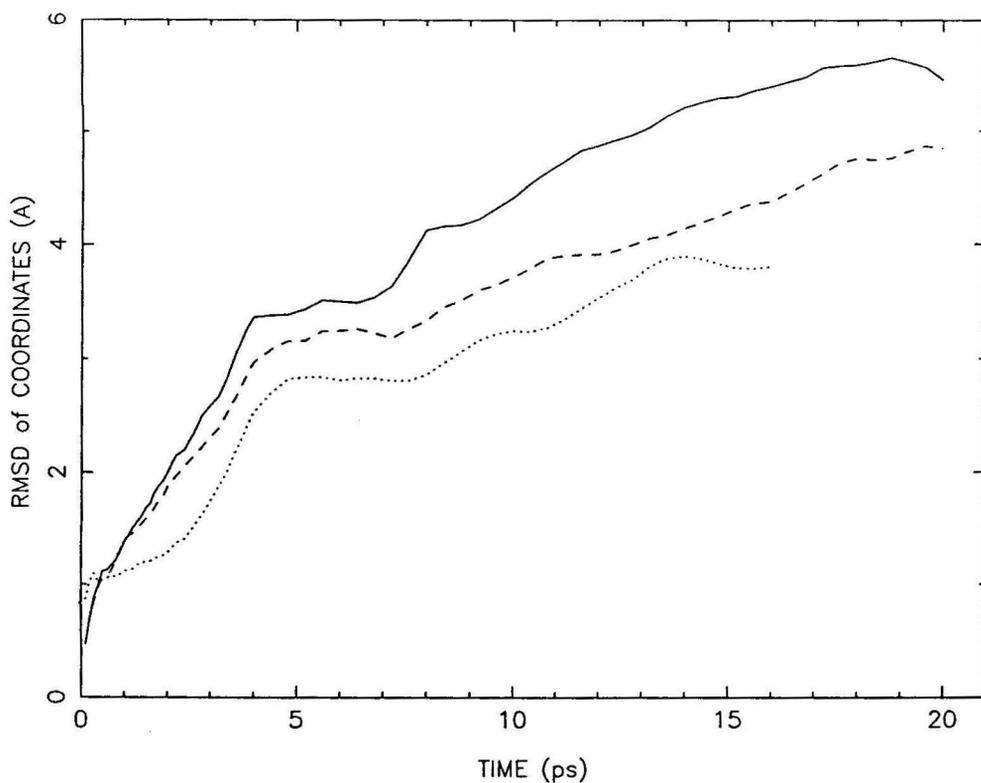


FIGURE 2-17: Coordinate changes in the course of the restrained molecular dynamics relative to the initial structures. The *dotted* line: refinement C2_I starting with Init_I; *dashed* line: C2_II, starting Init_II; *solid* line: C2_III, starting Init_III. Refinements were with Constraint Set #2 (NOESY $\tau_m = 200$ ms).

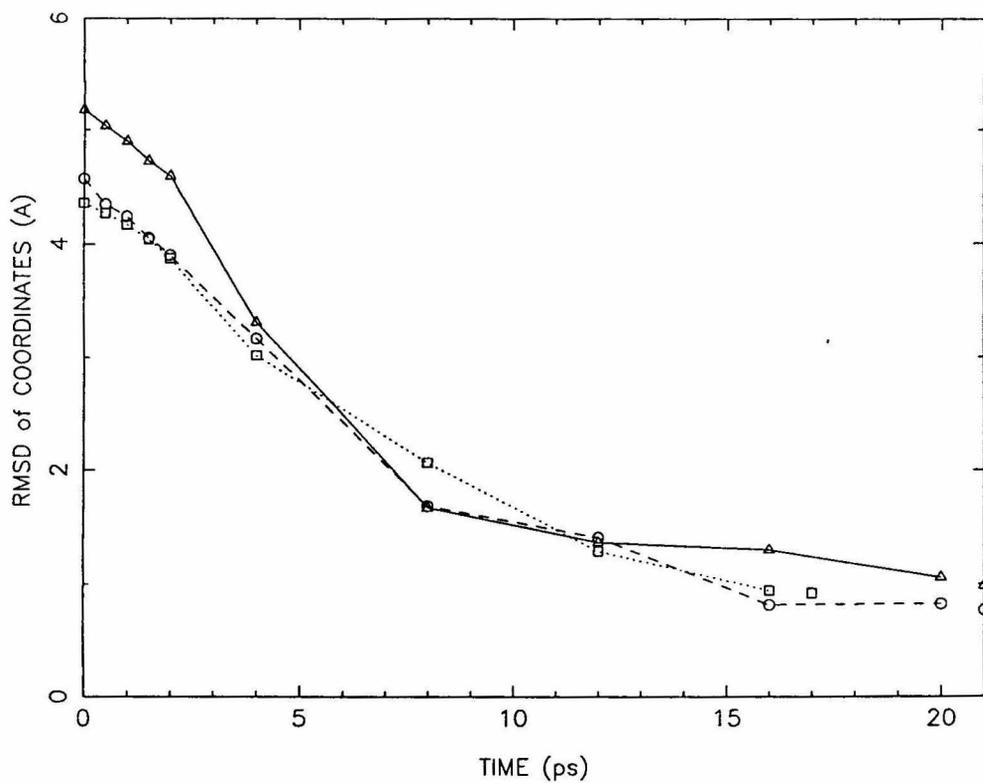


FIGURE 2-18: Atomic coordinate RMS differences in the course of the restrained molecular dynamics, relative to the average converged structure R2. Refinements were with Constraint #2. The *squares* and *dotted* line: refinement C2_I; the *circles* and *dashed* line: C2_II; the *triangles* and *solid* line: C2_III. The atomic RMS differences converges to a radius of 1.7 Å.

appears that rMD runs with Constraint #1 reach convergence faster than runs with Constraint #2. In addition to more distances in Constraint #1, the distances may have better qualities due to better resolution and signal-noise ratio in the first experiment ($\tau_m = 300$ ms).

Apparently, three rMD refinements with Constraint #1 have reached slightly different structures within a coordinate RMSD radius of approximately 1.7 Å; three refinements with Constraint #2 also converged to a similar radius among themselves. The convergence are shown in Figure 2-16 and Figure 2-18, where structures in rMD paths are plotted as their atomic RMS differences with the average converged structures. It appears, from the curves and Table 2-4, that the converged structures have the same travelling distances from Init_I and Init_II, but much farther from Init_III.

The refined structure R1 is made from averaging the coordinates of converged structures C1_I, C1_II and C1_III; likewise is R2 from C2_I, C2_II and C2_III. A comparison of these two structures gives an atomic RMS differences of 2.1 Å. It appears that the errors in the NMR constraints do not permit us to define structures more precise than this limit. This resolution is similar to that reported by Gochin & James (1990), Boelens et al. (1989) and Pardi et al. (1988) but is much lower than that by Baleja et al. (1990) and Nilges et al. (1987a,b).

Even the rMD refinements with the same constraint set can only be defined as moderately converged, from 3.6 Å to 1.7 Å with a travelling distances ranging from 3.4 to 5.4 Å. The general trend of less convergence with Init_III probably arises from the limited simulation

time of the rMD runs (Figure 2-15 to 2-18). In a refinement study with DNA I (Chapter 3), a similar rMD run was attempted to start from an A-DNA structure. Although the structure seemed to convert to B-DNA locally, the refinement failed to converge within the simulation time scale and resulted in a structure with significantly higher constraint and total energy.

The interproton distance violations, as shown in the first column of Table 2-5, appears to be large. Part of the distance violation should be accounted for from the errors in the NMR distance constraints, as this has been established as $\sigma = 0.26 \text{ \AA}$. The maximum force constant of $10 \text{ kcal/mol/\AA}^2$ appears to have reduced the distance violations to 2.6σ . A larger force constant will further reduce such distance violations and improve convergences, as shown in the case of C1_II. It is an open question as to how close we should bring down the the distance violation by increasing constraint force constants, if such practice is allowed by sufficient internal data consistency, which appears to have been reflected in the σ already. A more important consideration seems to be the mismatch of the energy minima defined by distance constraints and the empirical force field. This could arise from distance errors due to time and conformational averaging of intramolecular motions. It could also arise from imperfect forcefield, e.g., due to ignoring solvent and counterions. Therefore, excessive large force constants have been warned against (Gochin & James, 1990). This seems to be the case in this refinement and likely to contribute to the distance violation. The average structures R1 and R2 are found to have lowered distance violations and

constraint energies than the actual minimized and converged structures (Table 2-5). In terms of empirical potential energy, R1 and R2 are in a very bad conformation because they have coordinates averaging from three converged structures without optimizing geometries. This assessment is further supported by the observation that when R1/2, the average structure of R1 and R2, is energy-minimized to give Final_II, the distance violation and constraint energy goes back up to the level of converged (minimized) structures (Table 2-5). Therefore, it is difficult to justify any force constants much larger than 20 kcal/mol/Å² and our refinement accuracy for interproton distances seems to be limited to 2σ , or 0.52 Å.

The constraint energy, empirical potential energy and kinetic energy in two selected rMD runs are plotted as a function of the rMD time course in Figure 2-19 (C1_I) and Figure 2-20 (C2_III). Constraint energy reduces rapidly after being applied. All energies reach equilibrium at about 7 ps, but the conformation evolves further and converges at much later stages of the rMD run (as shown in Figure 2-15 to 2-18). This phenomenon appears to be universal, implying that energy convergence is necessary but not a sufficient condition for conformational convergence in restrained molecular dynamics. It should be pointed out that rMD is sampling conformational space not only in terms of energy, but also in terms of conformational accessibilities or the probability distributions of structures. In these energy curves, the bumps represent the heating and annealing processes

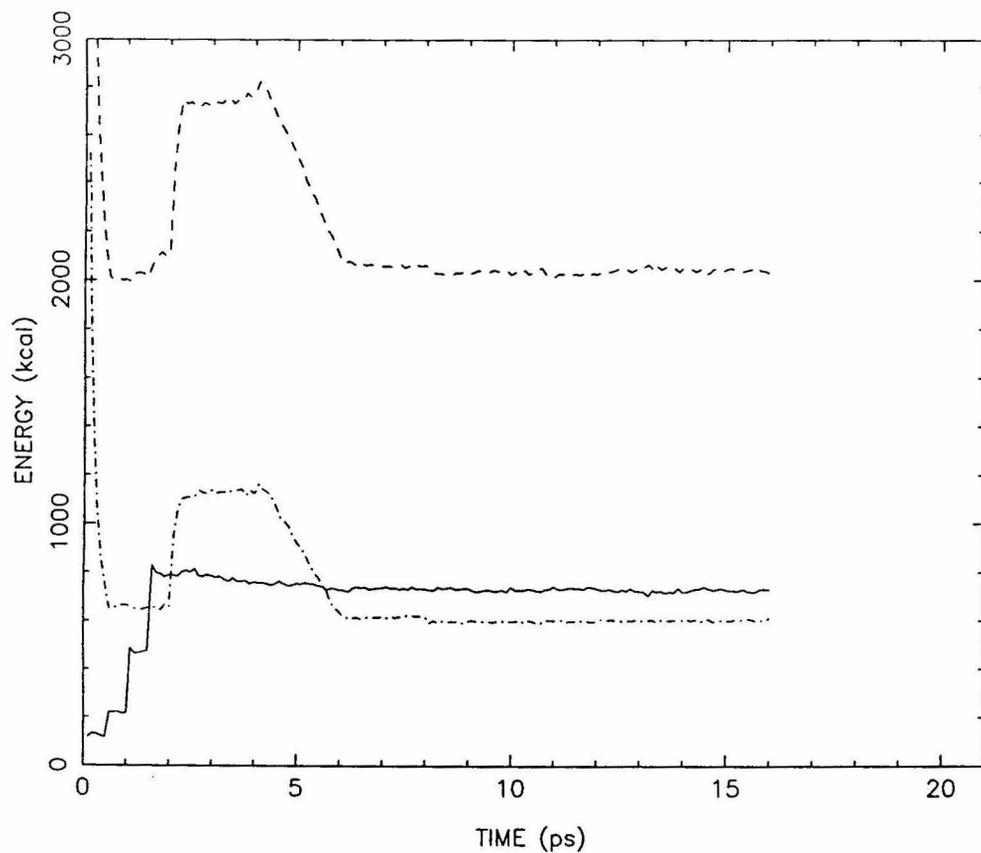


FIGURE 2-19: Energies in the course of refinement C1_I. The *solid* line: constraint energy; *dashed* line: empirical potential energy; *dotted* line: kinetic energy. The refinement was started from Init_I, a B-form DNA, with Constraint #1 applied. Constraint force constants were increased in 4 steps between 0-2 ps period. The molecule was heated up to 600K during simulation 2-4 ps period and cooled down to 300 K for the next 2 ps. MD at 300 K were continued until the simulation converges. The energy bumps reflect the heating and cooling process.

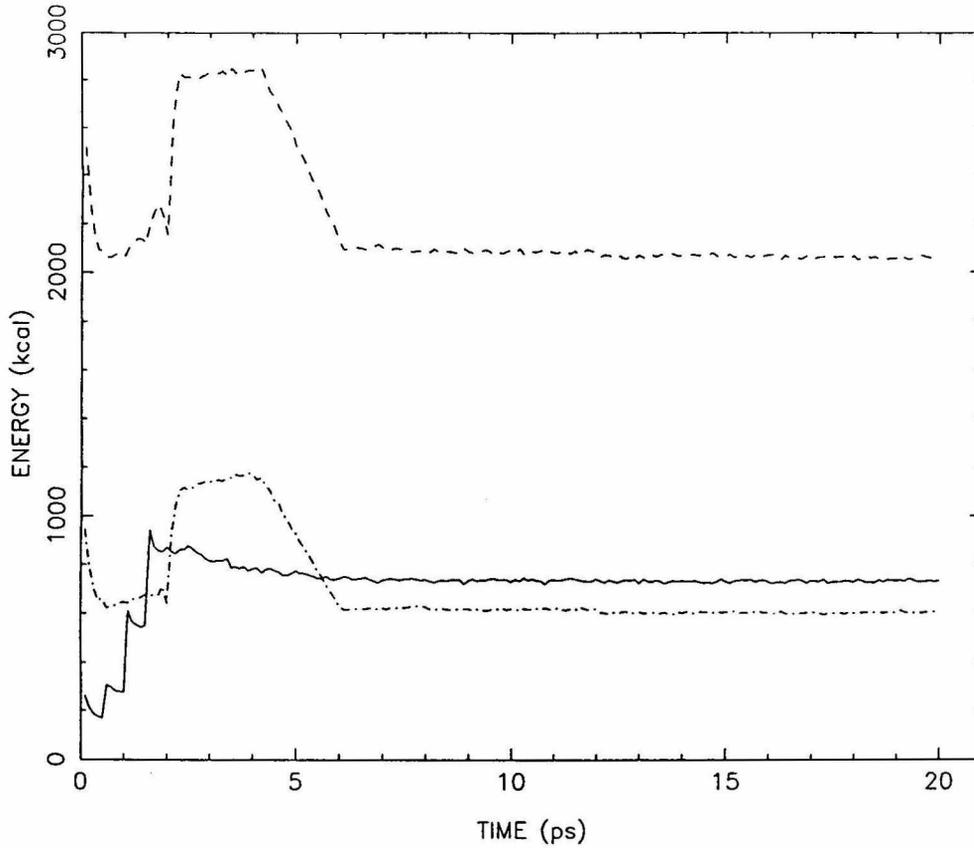


FIGURE 2-20: Energies in the course of refinement C2_III. The *solid* line: constraint energy; *dashed* line: empirical potential energy; *dotted* line: kinetic energy. The refinement was started from Init_III with Constraint #2 applied.

in the rMD. The interesting observation is that constraint energies are not affected by heating. This could be attributed to insufficient coupling between the constraint and empirical energy, or simply due to the rigid distance function's inability to absorb thermal energy.

We have chosen in this study a parabolic function without a flat bottom, such that we can in the future incorporate force constants of individual distances directly from individual standard deviations of constraint distances. This force function tends to emphasize the center value of a distance slightly than flat-bottom functions. Similar force functions have been used in other studies (e.g., Nilges et al., 1987). In such a force function, distance RMSD violation is proportional to the constraint energy divided by the force constant. Extrapolating from the energy convergence argument, similar distance violations do not guarantee similar converged structures. In order to determine solution structures at faithful accuracy, sufficient numbers of diversified rMD runs are necessary to sample a probability distribution of the conformers in the conformational space.

Control Experiments

The control experiments were designed to address two questions about the rMD-based structural refinement using NMR constraints. The first question is that how do we know the converged structures are, in fact, defined by the constraints, and not by other components like the empirical molecular forcefields? The forcefields, although playing a significant role in optimizing the chemical structures in the course of

rMD, are known to be not accurate enough to determine the three-dimensional structure of a macromolecule *per se*. The concern is that certain imperfections in the forcefield may result in structural artifacts, e.g., collapsed major grooves (Baleja et al., 1990). The second question is how do we know the refined structures, as defined by distance constraints, are the “true” solution structure which gives the distances?

The first question was answered with a *negative control* experiment. A simulated distance constraint set was produced using coordinates of a B-form DNA II. The simulated set has the same constraining atom-pairs and force constants as in the experimental constraint set. Random noise in Gaussian distribution ($\sigma = 0.25 \text{ \AA}$) was added with scaling to increase uncertainty at longer distances. The only difference between the simulated B-DNA constraint set and the experimental one is, therefore, the distance values. With the same refinement protocol, the simulated B-DNA constraints were applied to one of the initial structures, Init_III. The resulting structure after rMD convergence was N_III. As shown in Table 2-6, N_III converged to a B-DNA structure (atomic RMS differences 2.14 \AA) and was more distant from the experimentally-measured structure R1/2 (RMSD 3.02 \AA). The converged structure N_III has no bending feature as in the Final_II (*vide infra*). The negative control results vindicate the point that our refined structures were defined by the distance patterns in the experimental constraints.

The second question was addressed by a *positive control* experiment. A simulated constraint set was generated using the method

Table 2-6: Atomic RMS Differences of Structures in Control Experiments of rMD^a

	N_III	P_I	B-DNA ^b	R1/2
Init_III	3.05	3.95	3.56	4.52
N_III		2.40	2.14 ^c	3.02 ^d
P_I			3.09 ^d	1.09 ^c
B-DNA				3.66

^aNegative control was started from Init_III with constraints drawn from B-DNA coordinates. Positive control was started from B-DNA with constraints from R1/2 (Table 2-4). ^bB-DNA is an energy-optimized structure from Init_I. ^cConverged structures: Negative control goes to B-DNA, Positive to the experimentally-refined structure. ^dNot preferred structures.

mentioned above, with distance values calculated from the coordinates of the experimentally-determined structure R1/2. With the simulated constraints applied, the rMD starting with Init_I reached a converged structure P_I which is almost identical to R1/2 (RMSD 1.09Å, Table 2-6). The positive control did not answer directly the question we raised of whether the refined structure is the “true” solution structure, as this question may never be answered affirmatively. The result does suggest the consistency of our method, that the refined structure possesses distances that have the same constraining power and effect as those from the true solution structure. The highly accurate convergence of the positive control experiment came with surprise, since it was known the simulated positive constraints were very different from the original experimental constraints (0.64 Å, from distance violations of R1/2). It seems to indicate that our structure determination method is very robust to noise disturbance and errors contained in the distance constraint sets.

Refined Structures of DNA II

Due to limited convergence and resolution, specific helical parameters can not be fully determined to a confident accuracy. Nevertheless, the overall structure has been overwhelmingly defined as we have seen some of the structure features appear consistently in all the converged structures. Figure 2-21 and Figure 2-22 superimpose converged structures under Constraint #1 and #2, respectively. The two average structures, R1 and R2, are superimposed in Figure 2-23. The further average structure of R1 and R2, after energy minimization to

optimize coordinates, is shown in Figure 2-24 as the final refined structure of DNA II (Final_II). It can be noticed from Table 2-4, Final_II is substantially different from the initial structures (atomic RMSD range 3.7-4.5 Å), and also, converged structures are forming a cluster within a 2.1 Å radius of Final_II.

A striking feature in all converged structures is a pronounced bending of the DNA molecule in the middle of the helices (Figures 2-21 to 2-24). The bending occurs primarily at a Pyr-Pur step T7-G8 (C21-A22). The rest of the helices appears straight. The bending angle can be therefore estimated by the angle between axes of the top and the bottom halves of the helices. The angle is estimated as $32^{\circ} \pm 7^{\circ}$. The error margin was deducted by the atomic RMSD value. It appears so in angle variations among different converged structures.

It seems the cause of the bending is mainly the junction between the oligo(dA) tract and the other half of the helices. The oligo(dA) tract appears straight with significant propeller twists of base pairs. Such structural features are not novel, as they have been observed previously in crystal structures (Nelson et al., 1987) and been suggested in many studies of sequences containing oligo(dA), including NMR studies (Katahira et al., 1990; Kintanar et al., 1987). Such strong propeller twists enable possible bifurcated hydrogen bonds between neighboring bases, and maintain the tract straight and rigid. Energetically, such propeller twist helps base stacking and is therefore favorable (Calladine, 1982). What is peculiar is that the propeller twists are abruptly dampened at the 5' terminal of the oligo(dA) tract. The next base-pair G8-C21 appears

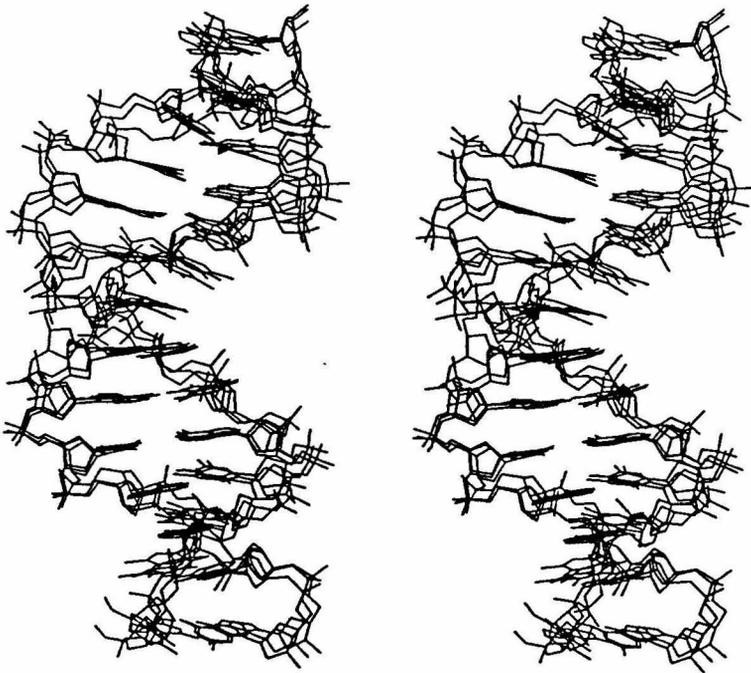


FIGURE 2-21: Superimpose of three converged structures C1_I, C1_II and C1_III in rMD refinements with Constraint #1 applied. Average atomic RMSD=1.7Å.

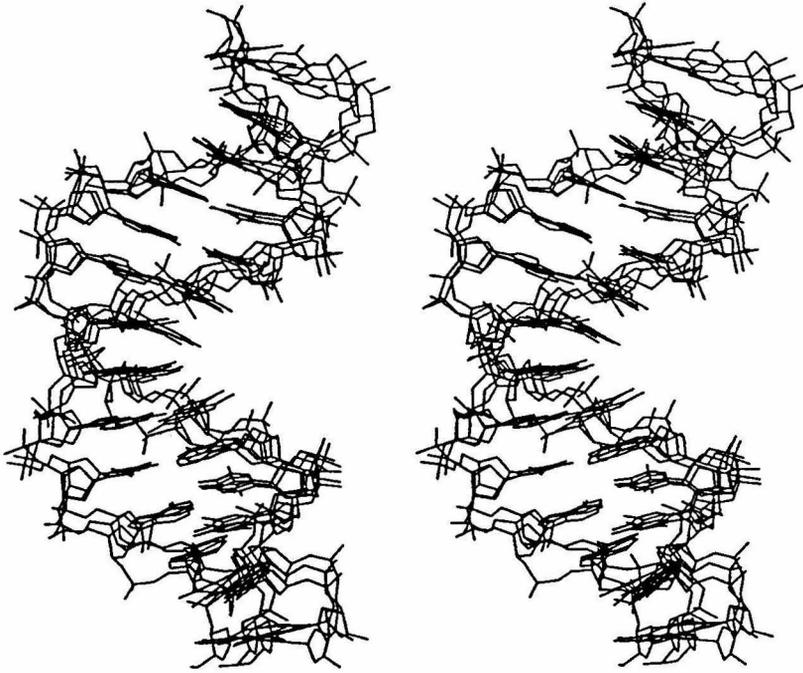


FIGURE 2-22: Superimpose of three converged structures C2_I, C2_II and C2_III in refinements with Constraint #2. Average atomic RMSD = 1.7 Å.

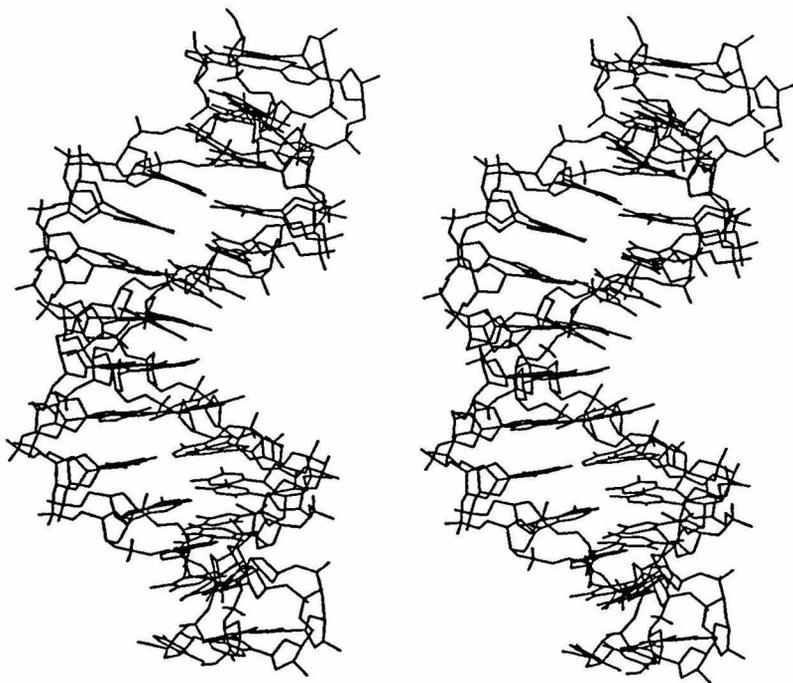
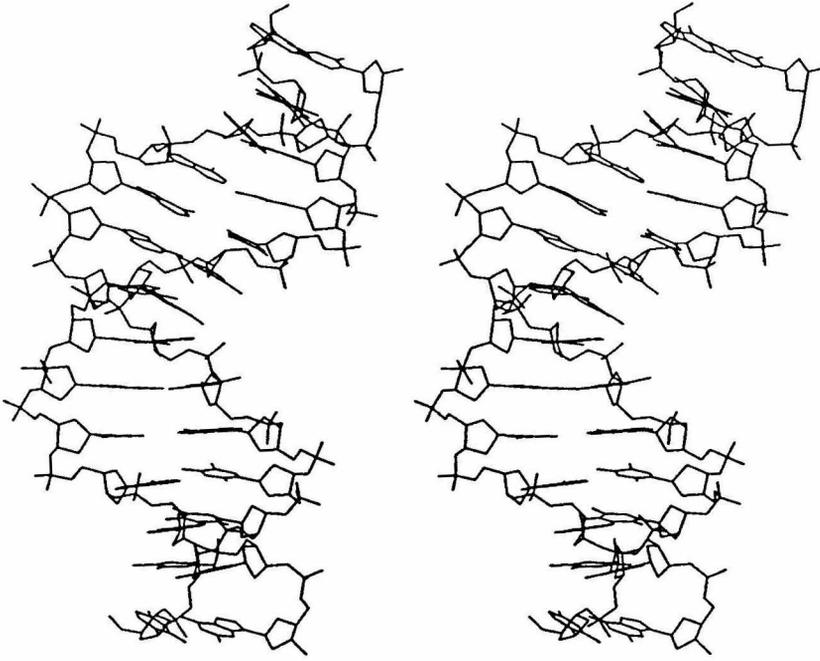
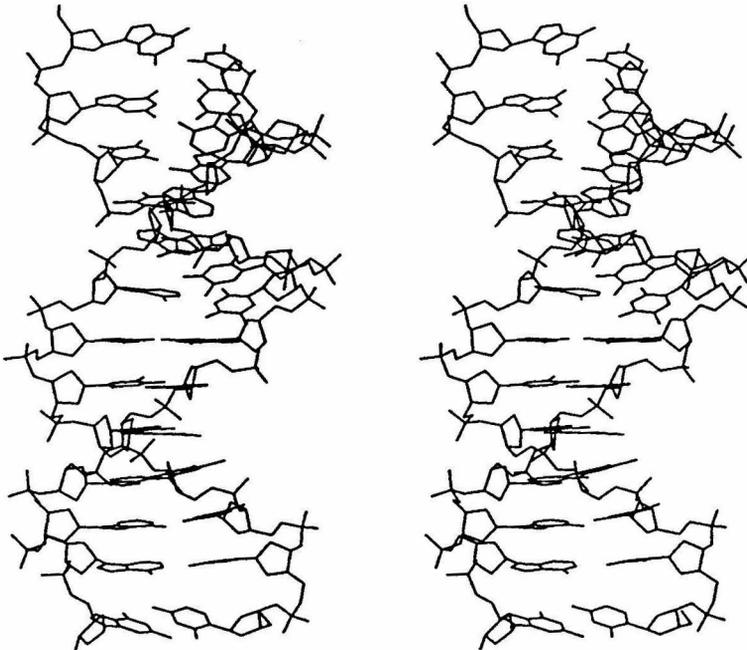


FIGURE 2-23: Superimpose of the refined structures R1 and R2. R1 is the average of three converged structures under Constraint #1. Likewise is R2 under Constraint #2. Average RMSD of the two structures is 2.1 Å.

A**B**

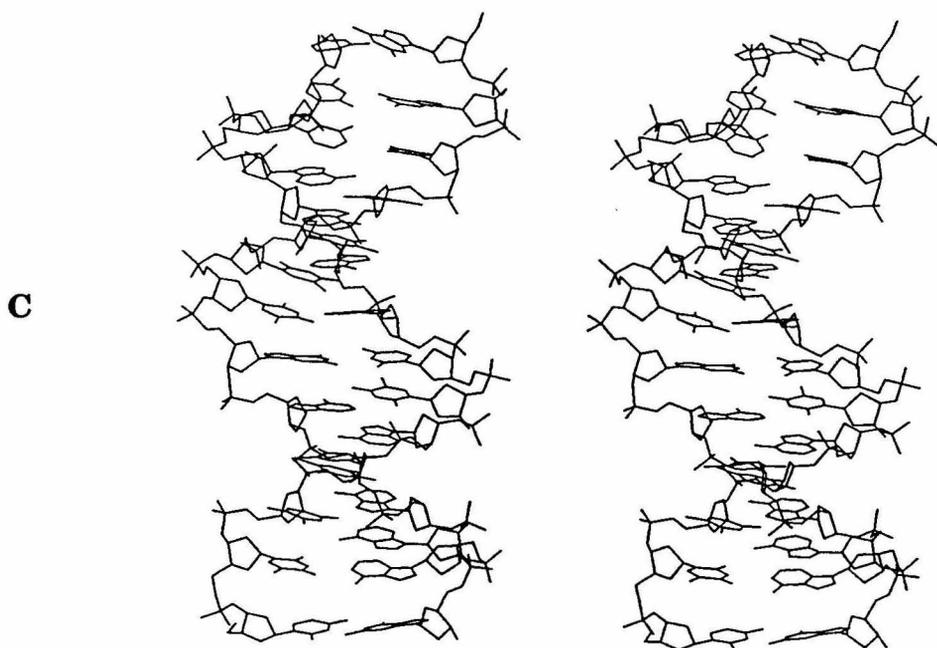


FIGURE 2-24: Final structure of refined DNA II in solution, Final_II. Structures were generated by averaging coordinates in six converged structures. The resulting coordinates have been optimized for force field potentials. **A.** Side view: bending of the DNA molecule and the positive roll angle are shown. **B.** Rear view: at the junction between base-pairs T7•A22 and G8•C21, a positive slide has taken place so the interstrand Pur-Pur overlap can be maximized. **C.** Front view: narrowed major groove is shown.

almost free of propeller twist, likely a combined effect of the base-stacking and the base-pair rigidity from a third hydrogen bond (Dickerson & Drew, 1981). To accommodate such a Pyr-Pur junction, a positive roll and slide have been introduced, resulting an inter-strand purine overlap. The T7-G8 (C21-A22) steps are thus transformed into an A-like conformation. Such a maneuver belongs to classic examples of Pyr-Pur junctions, and suggested by Calladine & Drew (1984) to be a result of an interstrand purine clash. Similar large roll angles and B-to-A transitions have been observed in Dickerson et al.'s crystal structure of $d(\text{CGCGAATTCGCG})_2$ at C3-G4 steps, and in Nilges et al.'s NMR solution structure of $d(\text{GCATGC})_2$ at C2-A3 steps. It is still not certain what are the sequences and structural elements that trigger such B-to-A transitions, as it appears not all Pyr-Pur steps adopt such changes. In the sequence of DNA II, we believe the triggering element is the long oligo(dA) tract which accumulates and rigidifies the strong propeller twisting nature of the top half of the DNA molecule. Pyr-Pur junctions have also been reported as unwinding helices (Patel et al., 1987) and wrinkled backbones (Gochin & James, 1990). All these observations are consistent with each other in a gross resolution. They seem to have a general consensus of deviated DNA structures at Pyr-Pur junctions.

The DNA II bears an identical segment of sequences $(d\text{CAAAA}) \bullet (d\text{TTTTG})$ as in Nelson et al.'s oligo(dA)•oligo(dT) tract studied by crystallographic methods. The structure determined by the

Table 2-7: Widths of Major and Minor Grooves (Å)^a

A. Major Groove

Strand 1:	5'-	G2	T3	T4	T5	T6	T7	G8	...
		(19.0)	18.0	16.6	15.9	16.8	17.9	(18.6)	
Strand 2:	...	A23	A22	C21	T20	A19	T18	T17	-5'
		(18.7)	17.8	16.3	15.9	16.9	18.0	(18.7)	
Average:		18.8	17.9	16.4	15.9	16.8	18.0	18.6	

B. Minor Groove

Strand 1:	...	T5	T6	T7	G8	A9	T10	A11	A12	A13	G14	-3'
		(9.0)	9.0	10.1	11.0	11.8	12.1	11.4	11.0	10.3	(10.1)	
Strand 2:	3'-	C28	C27	A26	A25	A24	A23	A22	C21	T20	A19	...
		(9.0)	9.0	10.4	10.9	11.8	11.8	11.5	10.8	9.9	(10.1)	
Average:		9.0	9.0	10.2	11.0	11.8	11.9	11.5	10.9	10.1	10.1	

^aWidths were measured by the distances between P-P on the backbones. Three nearest distances were averaged to give the individual widths.

two methods are also qualitatively similar. This includes structural features like straight and strong propeller twisting nature of oligo(dA) tract and the positive roll angle at steps C-A (T-G). The roll angle reported by Nelson et al. is much smaller (10.3°) though.

There has been great interest in the bending of DNA by phased runs of oligo(dA) tracts (Koo & Crothers, 1987; Ulanovsky et al., 1986). It is tempting to generalize the bending phenomenon in this DNA molecule to other systems. Nelson et al. (1987) have established the straight nature of oligo(dA) tract, as has been observed in this study. Thus the wedge model (Trifonov & Sussman, 1980) of DNA bending is not supported by these results. We have observed distances patterns (H2-H1') suggesting minor groove compressions as suggested by one of the junction models (Koo et al., 1986). Indeed, the minor groove of DNA II in the refined structure is compressed towards the 3' direction of the oligo(dA) tract in the refined structure (Table 2-7B). However, we did not see significant bending at the 3' junction of oligo(dA) tract in the structure. The 3' junctions are expected to have greater structural anomalies associated with DNA bending according to the junction model (Koo et al., 1986). In contrast, we found the DNA bends at the 5' junction of the oligo(dA). If the anomalies found in DNA II are not a general cause for DNA bending, then other DNA bending models may need to be explored (Nelson et al., 1987).

One of the consequences of the DNA bending is to create a short segment of narrowed major groove, since the bending arises from a positive roll angle opening base-pairs towards the minor groove and

compressing the major groove. As shown in Table 2-7A, The narrowest part of the major groove is located near the groove in between T5 and T20. Interestingly, this is precisely the location where the Hin recombinase recognize the DNA sequence, as according to Sluka (1988), Plaxco et al. (1989). The DNA binding domain of Hin recombinase, a putative helix-turn-helix motif, has been suggested to place its recognition helix in the major groove approximately in between T4-T6 and T20-A22. We believe what we have observed is not a mere coincidence, that such sequence-specific variations in DNA structure are correlated to the specific DNA-protein interactions. Such a correlation will be further discussed in the next chapter.

REFERENCES

- Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry* **29**, 4828.
- Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry* **28**, 6996.
- Bax, A. & Lerner, L. (1986) *Science* **232**, 960.
- Behling, R. W., & Kearns, D. R. (1986) *Biochemistry* **25**, 3335.
- Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct.* **173**, 299.
- Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson.* **82**, 290.
- Borgias, B. A., Gochin, M., Kerwood, D. J., & James, T. L. (1990) *Progress in NMR Spectroscopy* **22**, 83.
- Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson.* **87**, 475.
- Broido, M. S., Zon, G., & James, T. L. (1984) *Biochem. Biophys. Res. Commun.* **119**, 663.
- Calladine, C. R. (1982) *J. Mol. Biol.* **161**, 343.
- Calladine, C. R. & Drew, H. R. (1984) *J. Mol. Biol.* **178**, 773.
- Denk, W. (1986) *J. Magn. Reson.* **67**, 386.
- Dickerson, R. E., & Drew, H. P. (1981) *J. Mol. Biol.* **149**, 761.
- Gochin, M., & James, T. L. (1990) *Biochemistry* **29**, 11172.
- Gochin, M., & Zon, G., & James, T. L. (1990) *Biochemistry*, **29**, 11161.
- Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol.* **45**, 665.

- Havel, T. F., & Wüthrich, K., (1985) *J. Mol. Biol.* **182**, 287.
- Hughes, K. T., Youderian, P., & Simon, M. I. (1988) *Genes Dev.* **2**, 937.
- Jeener, J., Bachman, P., Meier, B. H., & Ernst, R. R. (1979) *J. Chem. Phys.* **71**, 4546.
- Katahira, M., Sugeta, H., & Kyogoku, Y. (1990) *Nucleic Acid Res.* **18**, 613.
- Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson.* **57**, 404.
- Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acid Res.* **15**, 5845.
- Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry* **30**, 3787.
- Koo, H. -S., & Crothers, D. M. (1987) *Biochemistry* **26**, 3745.
- Koo, H. -S., Wu, H. -W., & Crothers, D. M. (1986) *Nature* **320**, 501.
- Macura, S., & Ernst, R. R. (1980) *Mol. Phys.* **41**, 95.
- Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113**, 967.
- Mayo, S. L., Olafson, B. D., & Goddard III, W. A. (1990) *J. Phys. Chem.* **94**, 8897.
- Nelson, H. C. M, Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature* **330**, 221.
- Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., & Nilsson, L. (1987a) *Biochemistry* **26**, 3718.
- Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987b) *Biochemistry* **26**, 3734.

- Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1986) *J. Mol. Biol.* **188**, 455.
- Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson.* **66**, 187.
- Pardi, A., Hare, D. R., & Wang, C. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8785.
- Patel, D. J., Shapiro, L., & Hare, D. (1987) *Ann. Rev. Biophys. Biophys. Chem.* **16**, 423.
- Plaxco, K. W., Mathiowetz, A. M., & Goddard III, W. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 984.
- Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson.* **19**, 250.
- Reid, B. R., Banks, K., Flynn, P., & Nerdal, W. (1989) *Biochemistry*, **28**, 10001.
- Sluka, J. (1988) Ph.D. Thesis (California Institute of Technology, Pasadena, CA).
- Soloman, I. (1985) *Phys. Rev.* **99**, 559.
- Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry* **25**, 6854.
- Trifonov, E. N. & Sussman, J. L. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3816.
- Ulanovsky, L., Bodner, M., Trifonov, E. N., & Choder, M. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 862.
- Verlet, L. (1967) *Phys. Rev.* **159**, 98.
- Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Nucleic Acid Res.* **12**, 4035.

Wemmer, D. E., Chou, S. -H., & Reid, B. R. (1984) *J. Mol. Biol.* **180**, 41.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Chapter 3

**Solution Structure of the *L. HixL* DNA
d(GGTTCTTGAAAACC)•d(GGTTTTCAAGAACC):
Sequence-dependent Structural Variations Correlated to
DNA-Protein Interactions**

In this chapter, we present the solution structure of a DNA molecule (DNA I) containing the *left* half of the *hixL* recombination site of Hin recombinase, designated *L.hixL* (Hughes et al., 1988). *L.hixL* is one of the two natural sequences that have the strongest affinities for the Hin recombinase binding domain (Bruist et al., 1987). The other sequence, *R.hixL*, is contained in DNA II and has been presented in the previous chapter. These two DNA segments have equivalent binding affinities with the protein domain. Their sequences, although homologous to each other, have significant differences. Considering the highly specific interactions between the DNA and the protein binding domain, questions are raised whether there are common structural elements, other than the specific contacting residues, playing a role in the interactions. More explicitly, we are looking for some sequence-dependent structural variations in DNA conformation which are unique for these sequences. If such common and unique features exist, it is highly possible they contribute greatly to the specific DNA-protein interactions, as conformational complementarities are usually the most important determinants for specific protein binding interactions, typically seen in immunoglobulins and other macro-molecular

contained 5.0 mM DNA I in duplex, 10 mM phosphate, 100 mM NaCl, 0.1 mM EDTA, pH* 7.0.

NMR spectra were recorded on two 500 MHz spectrometers. NOESY experiment were performed using mixing times $\tau_m = 50, 75, 100$ to 500 ms in 50-ms increments, 600, 800 and 1000 ms. NMR spectra reported here were taken at 35°C in D₂O.

DNA I has 613 interproton distances that have possibilities of being shorter than 5Å and giving observable NOE cross-peaks, assuming a right-handed B or A type conformation. Statistics showed only 29% of these, or 178 cross-peaks were completely isolated ones and free of interference from overlapping with neighboring peaks. One of the problems with this sequence was its pseudo-palindromic symmetry. Because of this, the two nucleotides on the terminals are equivalent and completely overlapped with each other. Thus, all resonances of G1 and G15, G2 and G16, C14 and C28 and some resonances of the C13 and C27 are equivalent. Based on the argument that sequence determines the structure, we believe it can be justified to assume these nucleotides have equivalent structures also. Therefore, all intensities arising from intranucleotide distances and sequential distances within these nucleotides were split equally to two terminals. Such an option was incorporated in our automatic deconvolution program. After cross-peak processing, total numbers of valid intensities obtained from a 300-ms and 200-ms NOESY were 468 and 457, which were 76% and 74% of all possible cross-peaks. This is a remarkable improvement over the original intensity set. The two NMR experiments were done

independently with an NMR sample of DNA I. The 300-ms NOESY was collected with 1024 transients in the t_1 domain, whereas the 200-ms spectrum had 512 transients.

Distance constraints were calculated from intensities with the program MARDIGRAS (Borgias & James, 1990). A correlation time $\tau_c = 2$ ns or 1 ns was used for the calculations. As stated in the preceding chapter, the two calculations resulted in very similar distance sets (data not shown). The distance sets calculated with $\tau_c = 1$ were used for structural refinements since they contain several more distances.

The distance constraints involving thymine methyl groups were treated similarly as in Chapter 2. An additional 20 hydrogen-bond constraints were added to the total constraints. These included all hydrogen bonds with imino and amino protons in G•C base-pairs and the imino protons in A•T base-pairs, except at the terminal base-pairs.

The structural refinements were started with structures Init_I, Init_II and Init_III. Init_I is an idealized B-DNA built with crystallographic data. Init_II was generated by disturbing the B-DNA structure in an MD run of 3 ps at 300 K. Init_III is also a disturbed structure in 1.5 ps free MD at 600 K. The restrained molecular dynamics was conducted in the same protocol as in MATERIALS AND METHODS of the previous chapter. A maximum force constant of 10 kcal/mol/Å² was used for applying distance constraints.

RESULTS AND DISCUSSIONS

Resonance Assignments

The proton resonance assignments of the DNA I were relatively straightforward. The only special attention required is the equivalent nucleotides at two terminals due to a pseudo-palindromic symmetry. Therefore, in the sequential assignment pathways, the resonances of the two strands started at the same point, split at the third nucleotide and came back to a same point in the end (Figure 3-2). Assignments of nonlabile protons in DNA I are listed in Table 3-1.

Qualitative Analysis of the NOE Intensity Data

Similar to DNA II in the previous chapter, qualitative assessment of the relative NOE cross-peak intensities ($\tau_m = 50$ ms, 100 ms) establishes that the overall conformation of DNA I is in B-type family.

The most unusual aspect of the NMR spectra was in the intensities of the sequential NOE cross-peaks between base protons H6/8 and sugar protons H2' and H2". Significant variations in these NOE intensities were observed along the sequence. The most dramatic variation is the weakening of sequential H2"-H8 in steps T7-G8 and C21-A22 (Figure 3-3). In idealized B-DNA, sequential H2"-H8 distance is 2.1 Å, which should produce a cross-peak that is among the strongest ones. The actual intensities indicated the distances are at least 1 Å longer, implying significant sequence-dependent structural deviations from idealized B-type DNA conformation at the Pyr-Pur steps.

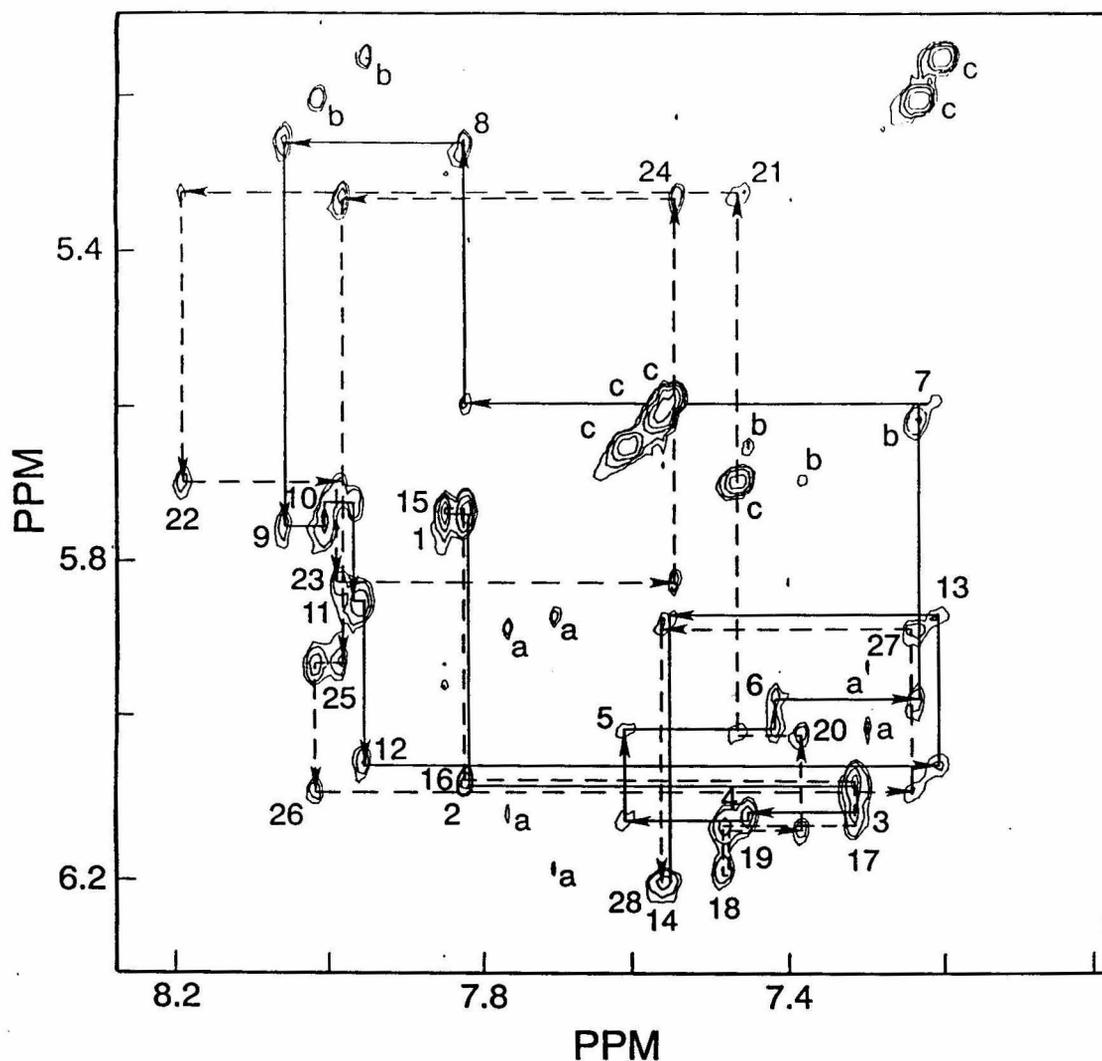


FIGURE 3-2: Assignment of base H6/H8 and sugar H1' of DNA I. The NOESY experiment was done at 35°C with a mixing time $\tau_m = 300$ ms. Spectral resolution is not enhanced. Sequential connectivities between H6/H8 and H1' are shown by the solid line for the strand 1 (residues 1-14) and by the dashed line for strand 2 (residues 15-28). Additional peaks labelled: (a) cross-peaks of adenine H2 to H1', interstrand-sequential or sequential; (b) sequential cross-peaks between the H5 base proton of cytosine and the H6/H8 proton of the preceding base; (c) intranucleotide H5-H6 NOE interactions in cytosine. Resonance assignments are given in Table 3-1.

Table 3-1: Proton Resonance Assignments for DNA I^a

Residue	H8/H6	H5/CH3/H2	H1'	H2'	H2"	H3'	H4'
G 1	7.851		5.735	2.567	2.696	4.816	4.194
G 2	7.824		6.079	2.649	2.846	4.978	4.436
T 3	7.316	1.376	6.117	2.163	2.628	4.881	4.294
T 4	7.456	1.618	6.124	2.273	2.601	4.920	4.238
C 5	7.616	5.651	6.010	2.138	2.544	4.812	4.226
T 6	7.419	1.644	5.967	2.093	2.502	4.861	4.173
T 7	7.235	1.673	5.615	1.886	2.230	4.820	4.031
G 8	7.827		5.262	2.627	2.631	4.945	4.263
A 9	8.062	7.165	5.753	2.584	2.763	5.007	4.353
A 10	8.006	7.052	5.729	2.545	2.772	5.000	4.358
A 11	7.968	7.115	5.850	2.518	2.825	4.994	4.395
A 12	7.957	7.706	6.055	2.500	2.799	4.945	4.406
C 13	7.207	5.149	5.864	1.984	2.378	4.721	4.134
C 14	7.559	5.590	6.206	2.234	2.235	4.520	4.018
G 15	7.851		5.734	2.570	2.705	4.816	4.194
G 16	7.824		6.077	2.647	2.849	4.977	4.436
T 17	7.316	1.375	6.128	2.185	2.637	4.891	4.305
T 18	7.488	1.619	6.190	2.256	2.669	4.922	4.238
T 19	7.483	1.659	6.135	2.177	2.655	4.914	4.237
T 20	7.385	1.655	6.018	2.115	2.491	4.888	4.171
C 21	7.466	5.693	5.324	1.932	2.242	4.809	4.061
A 22	8.192	7.176	5.701	2.700	2.788	5.009	4.322
A 23	7.988	7.405	5.823	2.522	2.729	4.997	4.339
G 24	7.552		5.332	2.438	2.593	4.925	4.279
A 25	7.986	7.299	5.927	2.561	2.840	5.004	4.385
A 26	8.018	7.769	6.089	2.534	2.795	4.977	4.411
C 27	7.238	5.206	5.882	2.002	2.391	4.725	4.138
C 28	7.571	5.612	6.205	2.238	2.239	4.520	4.018

^aChemical shifts (ppm) are at 35°C, relative to 3-(trimethylsilyl)propionic acid (external standard).

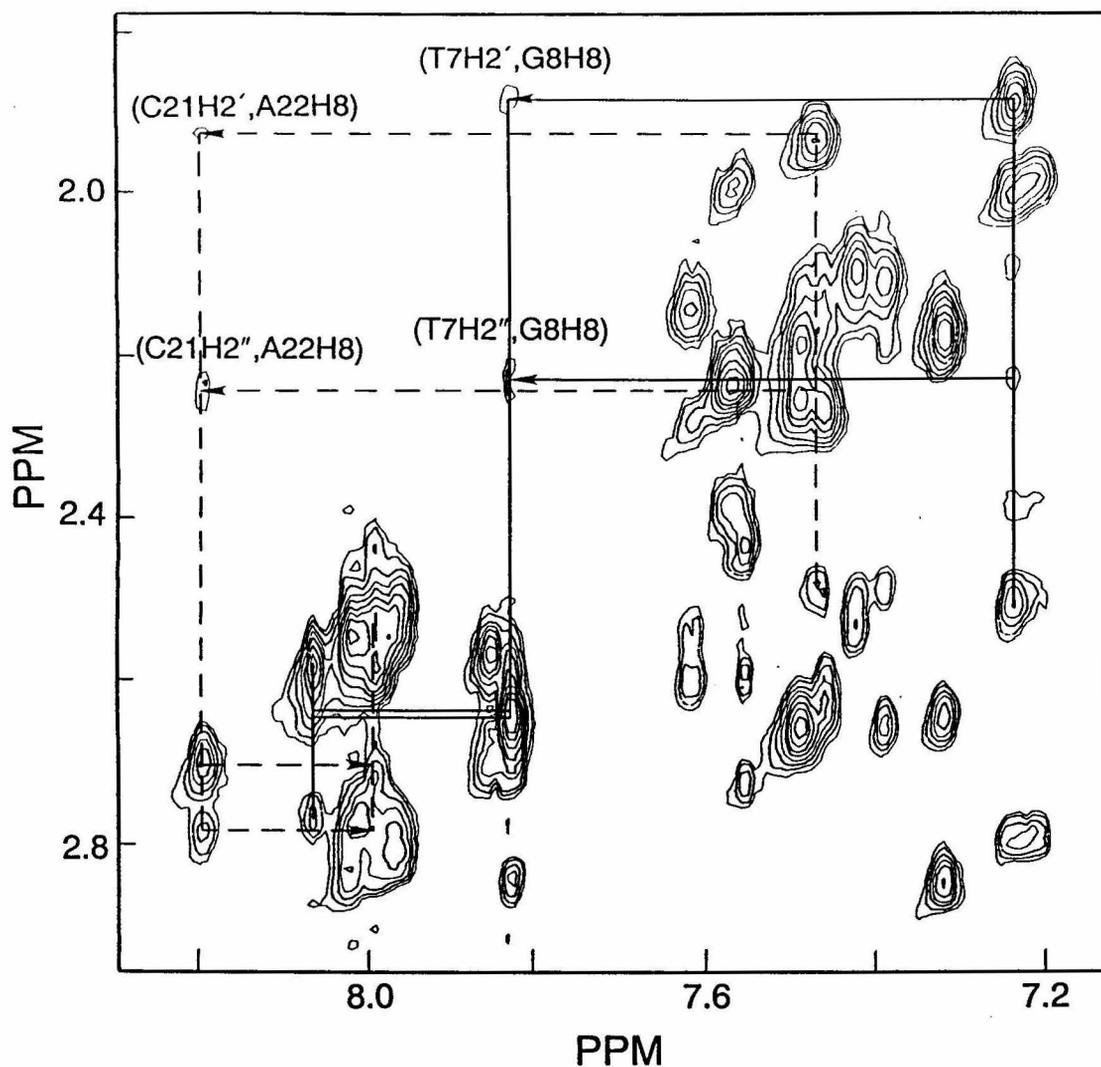


FIGURE 3-3: The 2-D NMR region showing NOE interactions between H2'/H2'' and H6/8 of DNA I. NOESY experiment was done with a mixing time $\tau_m = 100$ ms. Resolution of the 2-D spectrum was not enhanced. Sequential distances H2'-H8 and H2''-H8 are expected to be 3.8 and 2.1 Å, respectively. The labelled H2''-H8 cross-peaks are, therefore, significantly weaker than expected. Compare similar variations in DNA I (Figure 2-6).

The NOE cross-peaks arising from interstrand-sequential and sequential H2-H1' were observed for the tetra(dA) tract and for some of the adenines in the tract d(AAGAA) (Figure 3-2). The distances (Table 3-2C and 3-3C) showed similar "minor groove compression" phenomenon as mentioned in studies of DNA II and other 2-D NMR studies of oligo(dA) tract (Kintanar et al., 1987; Katahira et al., 1990). The lack of interstrand-sequential H2-H1' cross-peaks between A22 and G8 was explained later by a widened minor groove due to a positive roll at T7-G8 and bending of the DNA molecule towards the major groove (*vide infra*).

Distance Constraints

Two complete sets of distances from NOESY experiments $\tau_m = 300$ ms (Constraint #1) and 200 ms (Constraint #2) are listed in Table 3-5 and Table 3-6 at the end of this chapter. Constraint #1 and Constraint #2 comprise 398 and 362 distance constraints, respectively. This represents about 65% and 59% of all interproton distances possibly shorter than 5 Å in the molecule. Figure 3-4 and Figure 3-5 summarize the statistics of distance constraints.

The DNA molecule is well-defined by a great number of constraints except for residue T19 and the step leading to it (T18-T19) in Constraint #2. There are only 2 intranucleotide and 1 sequential distance constraints available in this region. The lack of constraints is obviously a result of overlapped resonances in T18 and T19.

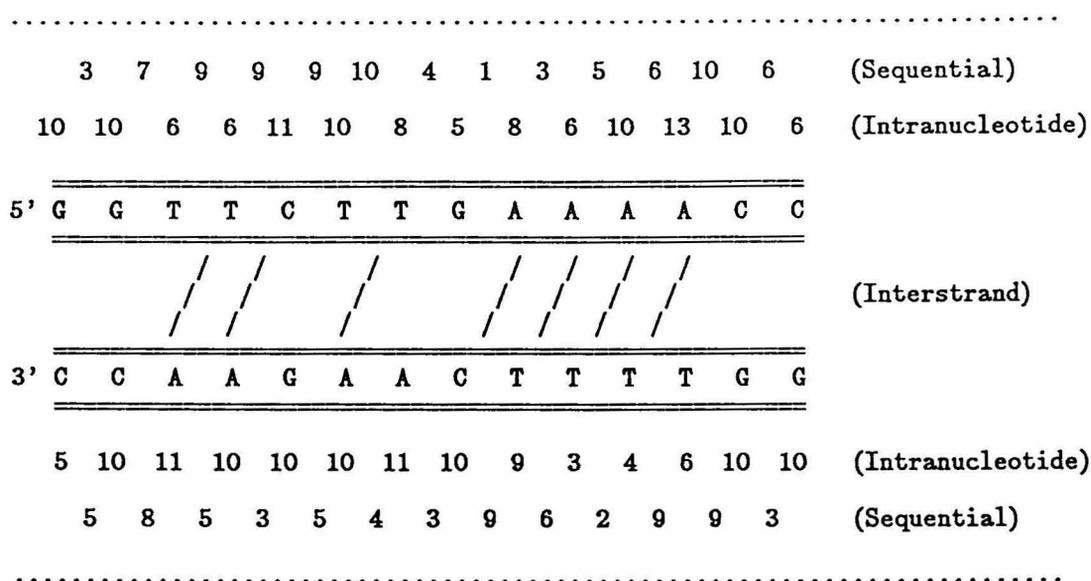


FIGURE 3-4: Statistics of distance Constraint #1. NOESY experiment was performed with $\tau_m = 300$ ms. Distances were calculated using a correlation time $\tau_c = 2$ ns. Number of intranucleotide distances for each residue is indicated near the residue, sequential distances placed in between two residues. Interstrand distances are indicated by connection lines between the strands. Total number of intranucleotide distances is 238, sequential 153, interstrand 7. Total distance constraints, 398.

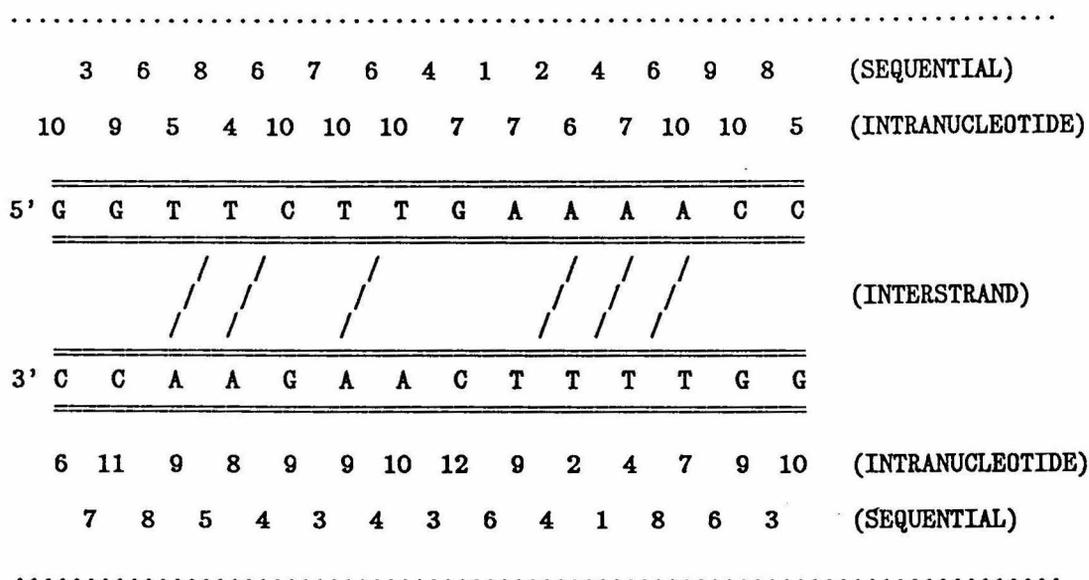


FIGURE 3-5: Statistics of distance Constraint #2. NOESY experiment was done with $\tau_m = 200$ ms. Total number of intranucleotide distances is 225, sequential 132, interstrand 5. Total number of all distance constraints, 362.

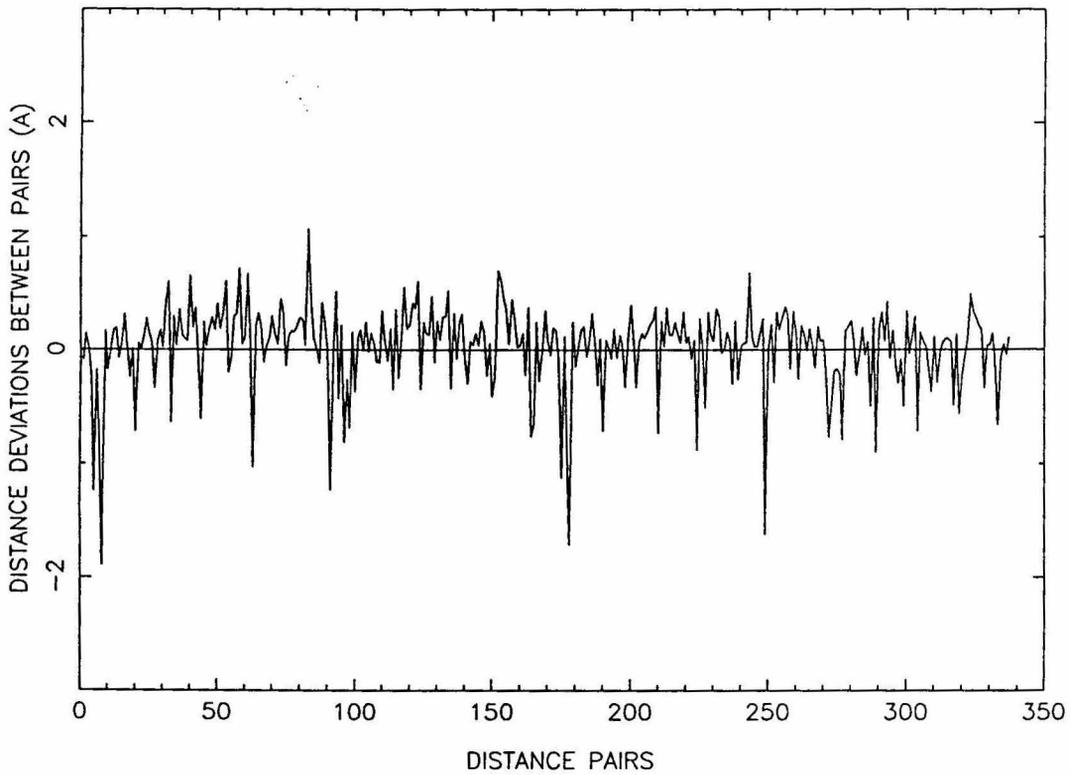


FIGURE 3-6: Distance Consistency: Deviation Plot. The difference between distances in a pair (distance in Constraint #1 minus distance in Constraint #2) is plotted against the pair-number. The distance RMSD is 0.37 Å, implying a standard deviation $\sigma = 0.26$ Å.

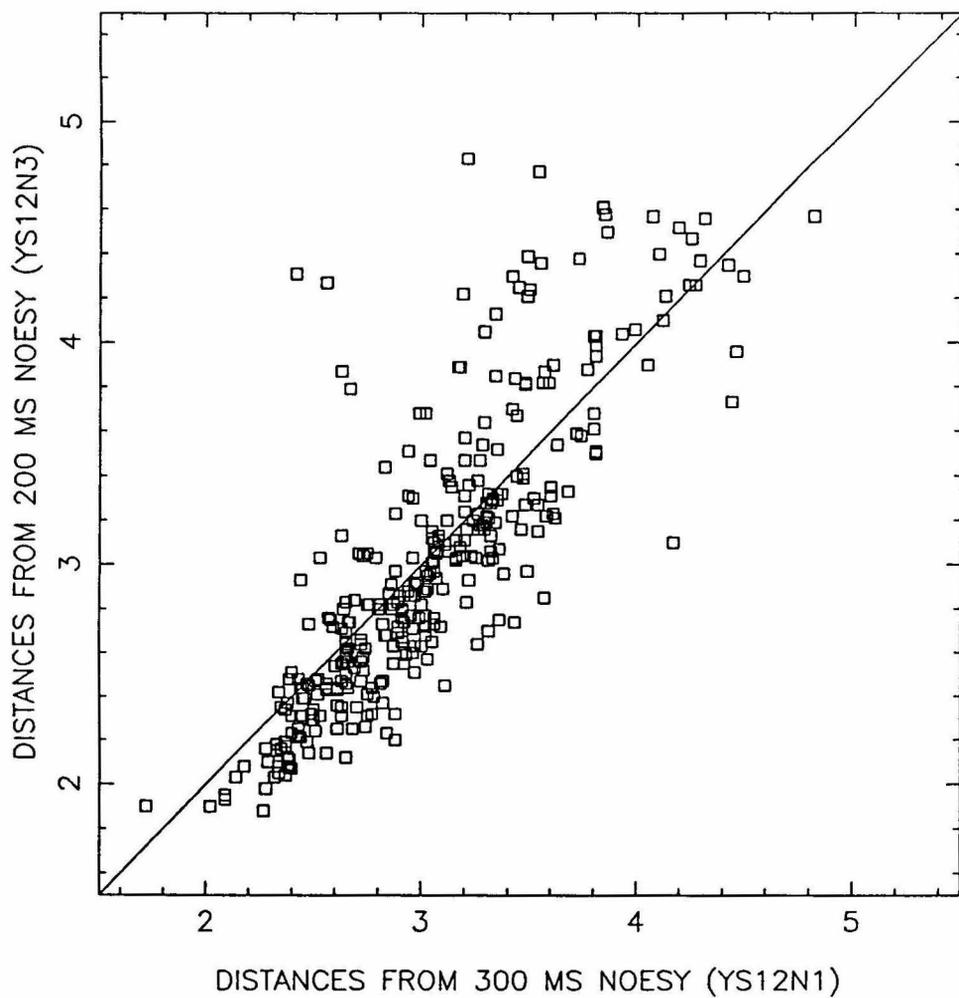


FIGURE 3-7: Distance Consistency: Correlation Plot. The distances in Constraint #2 (y-axis) are plotted against distances in Constraint #1 (x-axis). Greater deviations at longer distances were expected.

The distance constraints from the parallel experiments are compared as their deviations and correlations in Figure 3-6 and Figure 3-7, respectively. The data shows a distance RMS difference of 0.37 Å, which is the same as the value in DNA II. The standard deviation of the average distances is $\sigma = 0.26$ Å.

Structural Refinements

The structural refinements using restrained molecular dynamics were started with structures Init_I (B-DNA), Init_II and Init_III (disturbed B-DNA) (Initial structures not shown), having an average atomic RMS difference of 2.2 Å. The rMD runs with Constraint #1 were C1_I, C1_II and C1_III; with Constraint #2, C2_I, C2_II and C2_III. The atomic RMS differences between all pairs of structures are given in Table 3-2. The distance violations, total empirical potential energy and energy decompositions of all structures are listed in Table 3-3. The overall convergence radius for this molecule is 2.4 Å. The RMS distance violation of the converged structures is 2.4σ or 0.62 Å.

The refinements were conducted satisfactorily for all rMD simulations with the exception of the base step T18-T19. When refinements were started with disturbed initial structures and Constraint #2, the region was not converged, apparently due to limited numbers of the distance constraints in this region. Consequently, the segment G15-T18 on the second strand are not well-determined relative to the other part of the molecule. The situation reaffirms the importance of

Table 3-2: Atomic RMS Difference (Å)^a

	Init_II	Init_III	C1_I	C1_II	C1_III	C2_I	C2_II	C2_III	R1	R2	R1/2	Final_I
Init_I	2.02	2.56	2.93	3.45	4.79	2.92	2.80	2.92	3.59	2.66	2.95	2.92
Init_II	-	1.90	2.75	3.15	4.50	3.23	2.82	3.01	3.31	2.86	2.88	2.87
Init_III		-	3.58	3.96	5.14	3.67	4.44	3.57	4.09	3.42	3.60	3.56
C1_I			-	1.77	2.74	2.89	2.44	2.02	1.41	2.26	1.51	1.43
C1_II				-	1.83	2.76	2.20	2.20	0.78	2.17	1.16	1.39
C1_III					-	3.85	3.32	3.08	1.44	3.26	2.23	2.43
C2_I						-	1.28	2.21	2.93	1.08	1.88	1.88
C2_II							-	1.61	2.37	0.63	1.31	1.36
C2_III								-	2.13	1.22	1.32	1.35
R1									-	2.28	1.14	1.31
R2										-	1.14	1.17
R1/2											-	0.51

^aRefinements were started with Init_I, Init_II or Init_III, using Constraint #1 or #2. C1_I, e.g., was started with Init_I and using Constraint #1. Constraints used were calculated using $\tau_c = 1$ ns. R1 is a refined structure from averaging C1_I, C1_I and C1_III; R2 is from C2_I, C2_II and C2_III. R1/2 is from averaging R1 and R2. Final_I is from energy minimization of R1/2 under Constraint #1.

Table 3-3: Constraint Violations and Energies^a

	RMS Violation	Total Energy	Bonds	Angles	Torsions	VDW	Electrost.	Hbond	Constraint
Init_I	0.772	7908	871	2286	316	3785	-384	-261	1290
Init_II	0.906	3688	445	1417	281	530	-521	-223	1726
Init_III	1.154	5473	748	1708	357	630	-476	-144	2596
C1_I	0.617	2088	126	1183	231	543	-596	-251	849
C1_II	0.622	2081	126	1204	241	548	-650	-251	860
C1_III	0.622	2055	126	1200	248	561	-684	-255	856
C2_I ^b	0.739	1884	122	1158	229	511	-606	-257	722
C2_II ^b	0.732	1886	125	1148	229	511	-591	-260	720
C2_III ^b	0.734	1835	126	1176	245	510	-686	-254	716
R1	0.593	23743	5529	2879	262	15018	-472	-257	780
R2 ^b	0.723	10446	4754	2033	264	3461	-491	-263	685
R1/2	0.600	40839	7064	3462	266	29749	-441	-274	810
Final_I	0.620	2082	121	1174	233	543	-610	-255	873

^aConstraint violations are in Å, all other terms (energies) are in kcal. Inversion energy is not listed (<5kcal).

^bViolations based on Constraint #2. Others are based on Constraint #1.

a complete distance set (Gochin & James, 1990). In contrast to this, another single-constraint step G8-A9 (Figure 3-4, Figure 3-5) was well-converged in all rMD refinements. The difference is, for step G8-A9, both nucleotides have considerable more intranucleotide constraints defining the local structures. These two situations seem to draw a line separating complete and incomplete constraint sets in defining the local structures of DNA.

Refined Structures of DNA I

The converged structures under Constraint #1 and #2 are superimposed in Figure 3-8 and Figure 3-9, respectively. The two average structures, R1 and R2, are superimposed in Figure 3-10. The further average structure of R1 and R2, after energy minimization to optimize coordinates, is shown in Figure 3-11 as the final refined structure of DNA I (Final_I). The converged structures form a cluster within a 2.1 Å radius of Final_I (Table 3-2).

The general features of DNA I can be summarized as following: 1) A pronounced bending of the DNA molecule primarily at a Pyr-Pur step T7-G8 (C21-A22) with the angle amounts to $25^{\circ} \pm 8^{\circ}$. 2) Narrowed major groove in between G2-T6 on the first strand and T19-A23 on the second strand (Table 3-4A). 3) Compressed minor groove towards 3' of oligo(dA) tract (Table 3-4B).

Microscopically, the persistent propeller twists are likely the cause for the compressed and deepened minor groove. The relatively widened

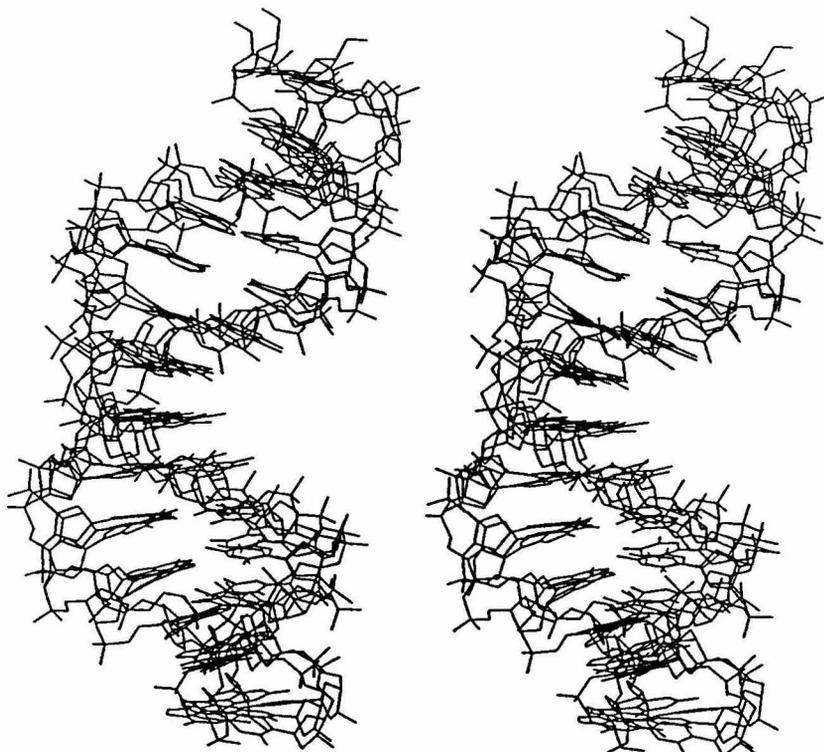


FIGURE 3-8: Superimpose of three converged structures C1_I, C1_II and C1_III in rMD refinement with Constraint #1 applied. Average atomic RMSD = 2.1 Å.

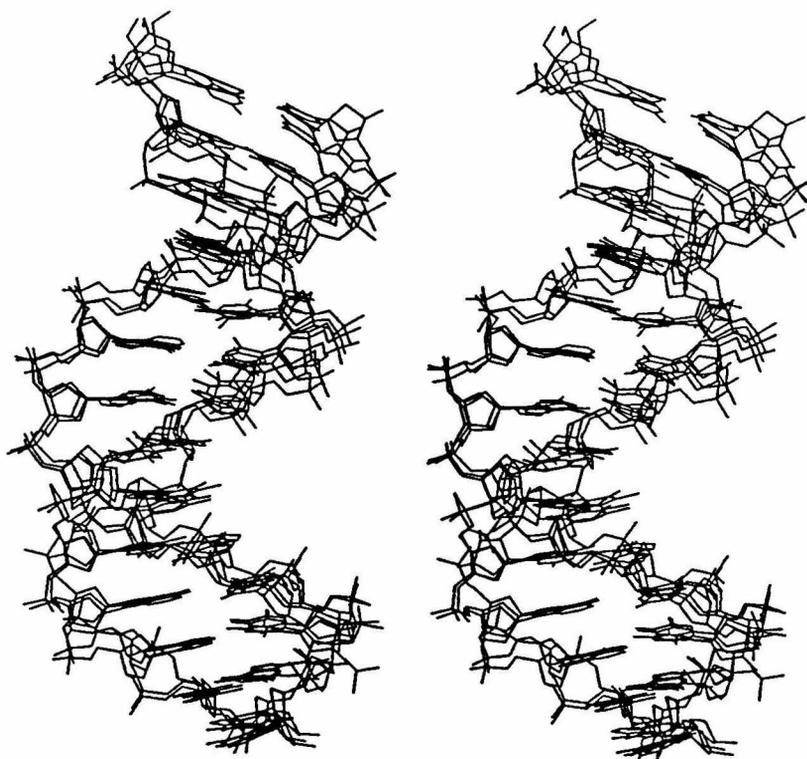


FIGURE 3-9: Superimpose of three converged structures C2_I, C2_II and C2_III in refinements with Constraint #2. Average atomic RMSD = 1.7 Å.

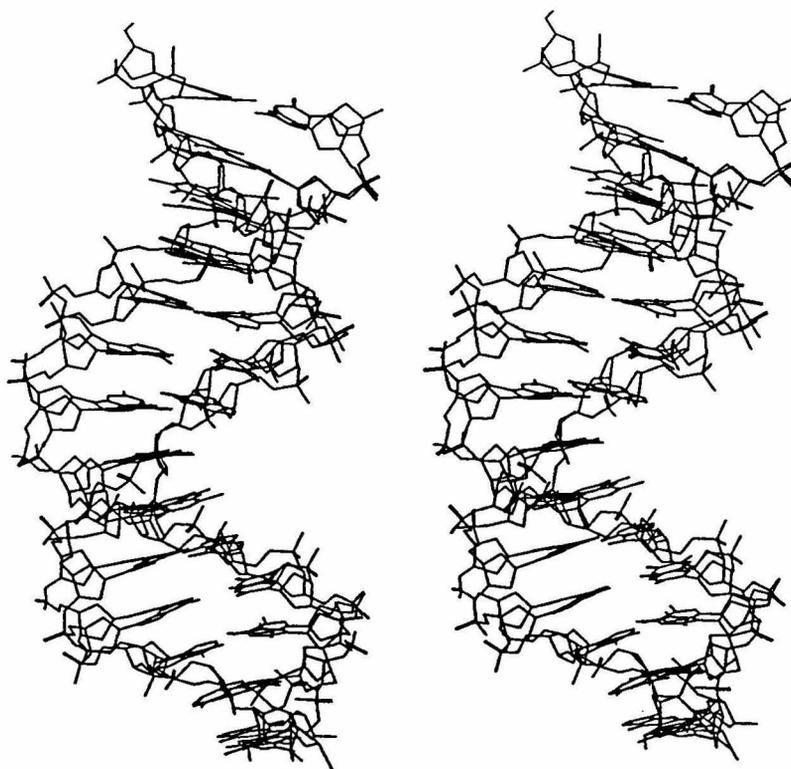
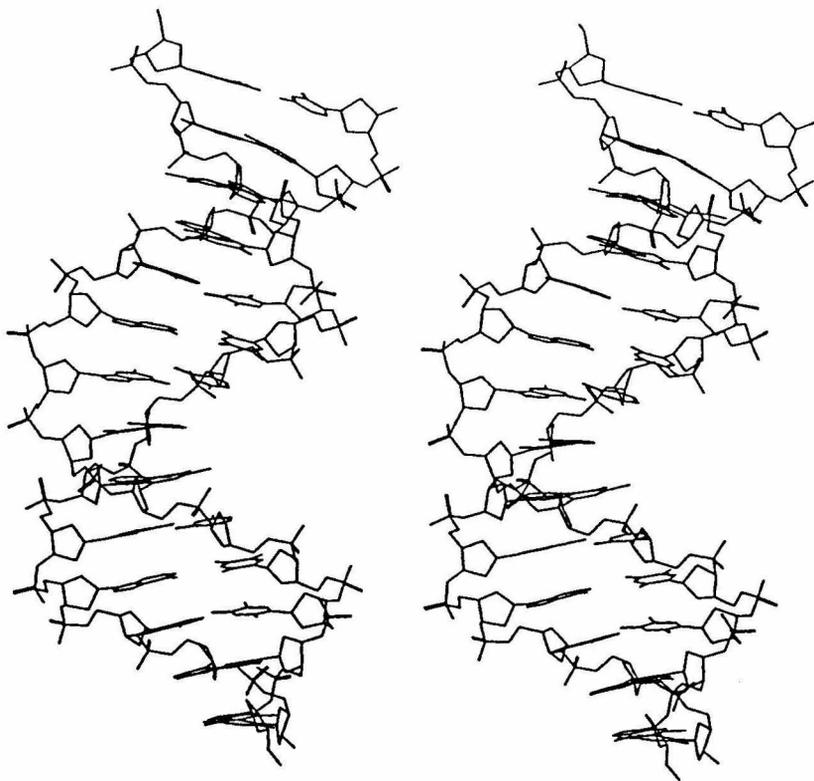
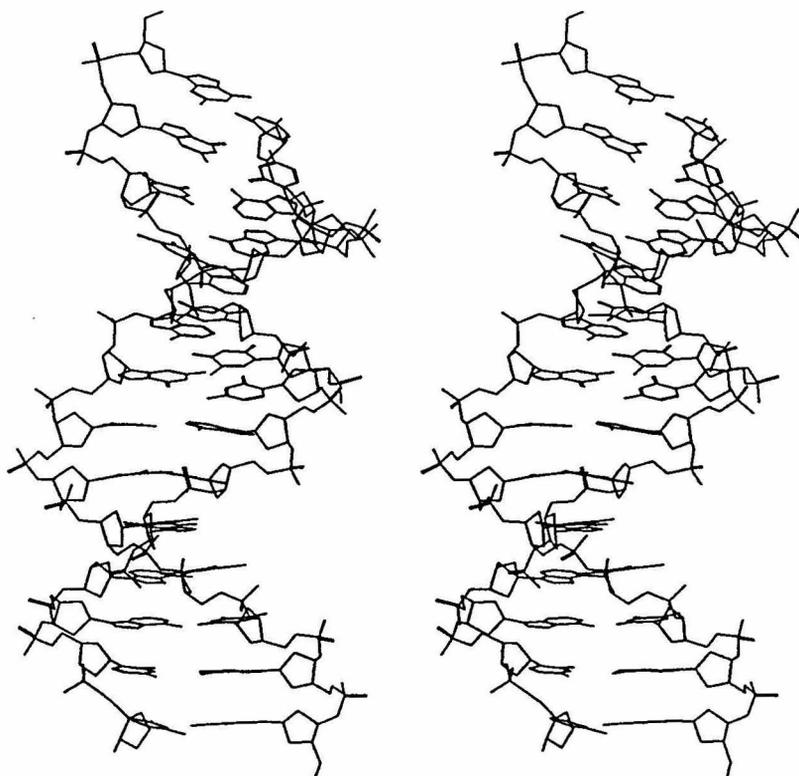


FIGURE 3-10: Superimpose of the refined structures R1 and R2. R1 is the average of three converged structures under Constraint #1. R2 is under Constraint #2. Atomic RMSD of the two structures is 2.3 Å.

A



B



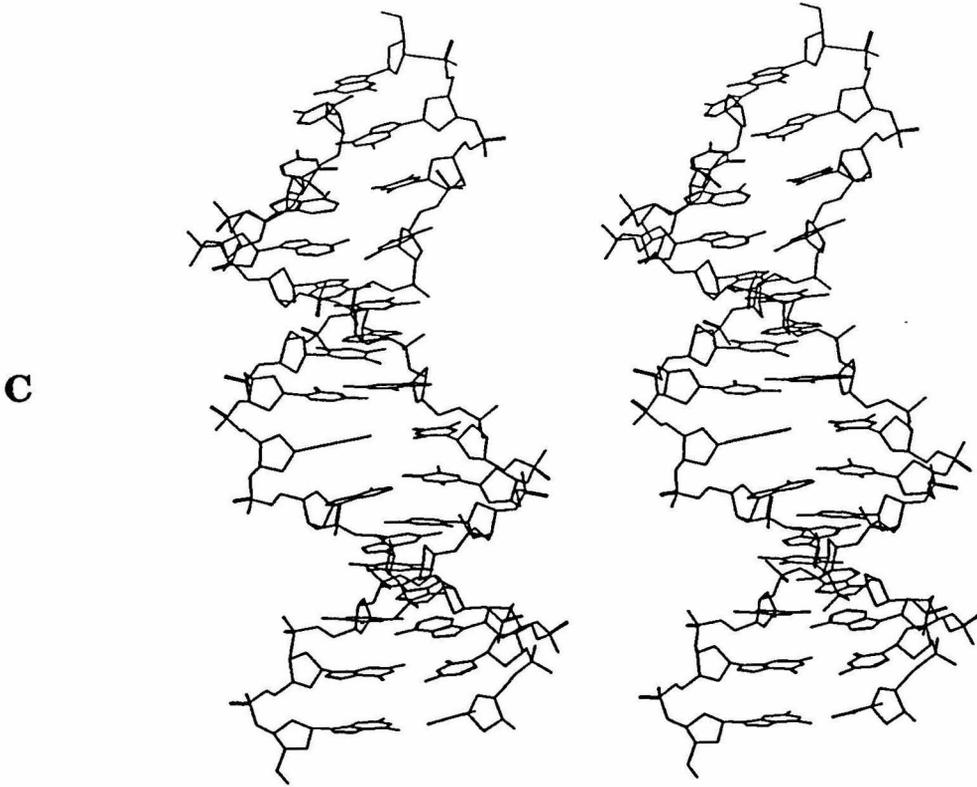


FIGURE 3-11: Final structure of refined DNA I in solution, Final_I. The structure was generated by averaging coordinates in six converged structures. The resulting coordinates have been optimized for force field potentials. **A.** Side view: bending of the DNA molecule and the positive roll angle are shown. **B.** Rear view: at the junction between base-pairs T7•A22 and G8•C21, a positive slide has taken place so the interstrand Pur-Pur overlap can be maximized. **C.** Front view: narrowed major groove is shown.

Table 3-4: Widths of Major and Minor Grooves (Å)^a

A. Major Groove

Strand 1:	5'-	G2	T3	T4	C5	T6	T7	G8	A9	...
		(16.8)	16.8	16.6	16.5	17.3	18.8	19.5	(19.2)	
Strand 2:	...	A23	A22	C21	T20	T19	T18	T17	G16	-5'
		(18.0)	17.1	16.9	16.5	17.8	18.5	19.4	(18.9)	
Average:		17.4	16.9	16.8	16.5	17.6	18.6	19.4	19.0	

B. Minor Groove

Strand 1:	...	C5	T6	T7	G8	A9	A10	A11	A12	C13	C14	-3'
		(10.5)	10.2	10.8	11.7	12.2	12.0	11.2	10.3	10.0	(9.9)	
Strand 2:	3'-	C28	C27	A26	A25	G24	A23	A22	C21	T20	T19	...
		(9.9)	10.2	11.3	11.8	13.0	12.4	11.9	10.3	10.0	(10.0)	
Average:		10.2	10.2	11.0	11.8	12.6	12.2	11.6	10.3	10.0	10.0	

^aWidths were measured by the nearest distances between phosphates on the backbones. Three such distances were averaged to give the individual widths. At terminals the average of two is used.

minor groove and narrowed major groove in the middle of the molecule is likely an effect of the DNA bending, which with a positive roll angle, opens up the minor groove and compresses the major groove. The propeller twists were significantly dampened at the base-pair G8-C21, where positive slide creating interstrand purine overlap is observed in combination with the positive roll. The obvious structural variation is, therefore, a B-to-A transition at the Pyr-Pur junction G8-C21 (T7-A22). These features demonstrate an almost classic sceneries of a Pyr-Pur junctions (Calladine and Drew, 1984).

Comparison of DNA I and DNA II

The recombination mechanism (Hughes et al., 1988) advises us that DNA I and DNA II should be homologous to each other in inverted repeats (palindrome). Just from the sequences, it seems they could match each other in direct repeats by only two base changes, whereas a three-base-change is needed for an inverted-repeat matching (Sequence in Figure 3-13, central 10 bp only). The reason they are not related in direct repeats is that the oligo(dA) tracts in these DNA molecules are only matched palindromically. It should be noted in DNA I the tract d(AAGAA), although similar in sequence to d(AAAAA) in DNA II, should be considered distinctly different tracts. There are evidences that the G in the sequence d(AAGAA) is an oligo(dA) breaker. It removes the otherwise possible structural role of the oligo(dA) tract (Koo & Crothers, 1987). This palindromic sequence similarity between DNA I and DNA II seems to emphasize the importance of the oligo(dA) tract in the sequence.

In fact, the oligo(dA) tract is part of the consensus sequence of the *Hin* recombinase binding sequences (Glasgow et al., 1989).

In Figure 3-12 these two DNA molecules are superimposed in inverted orientations. Their structural features are strikingly similar to each other. The only significant discrepancy is between the structure of the strand segment G15-T18 in DNA I and G1-T4 in DNA II. This is apparently caused by the poor convergence at this region in DNA I as discussed before. The major groove widths and minor groove widths of the two DNA molecules are plotted in Figure 3-13. The major groove width reaches minimum width near residues T5 and T20 in DNA II. The narrowest point for DNA I is one residue apart, but this is likely an artifact, since the poor convergence of segment G15-T18 is responsible for the rapid rise of major groove width of DNA I (Figure 3-13). Due to the limited length of the DNA molecule, major groove widths can not be measured for the 3'(strand 1) half of the molecule. The widths of the minor grooves of the two DNA tracts demonstrated a very consistent correlation. The maxima of the minor groove widths are the result of Pyr-Pur positive roll angles and the bending of the DNA molecules. Overall, the structural features of the two DNA molecules are very well-correlated, and all those common structural variations, including the bending, narrowing of the major groove and widening of the minor groove, are centered at the Pyr-Pur junctions, which are steps T7-G8 (C21-A22) in DNA I and C21-A22 (T7-G8) in DNA II.

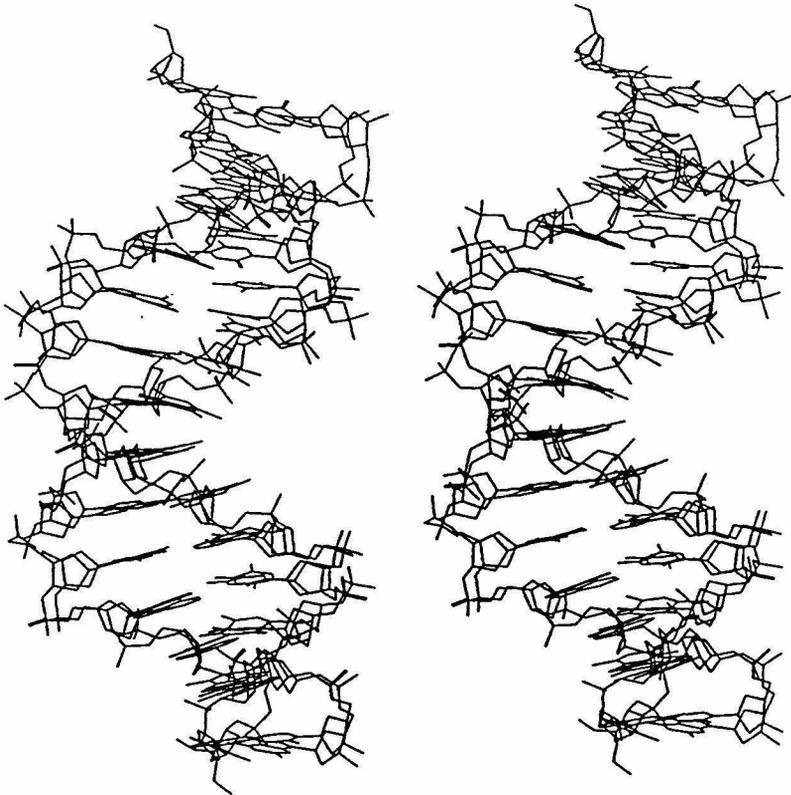


FIGURE 3-12: Superimpose of the Final_I and Final_II. The two molecules are orientated in opposite directions, with the 5'-strand 1 of the Final_I pointing downwards, Final_II upwards. Despite differences in sequences, two structures are strikingly similar. Discrepancies are observed between G1-T4 of Final_II and G15-T18 of Final_I at the top of the molecules, an artifact possibly arising from the poor convergence of DNA I refinements in that region.

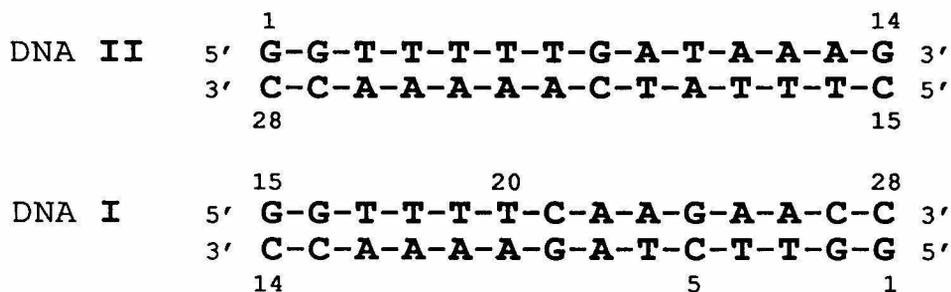
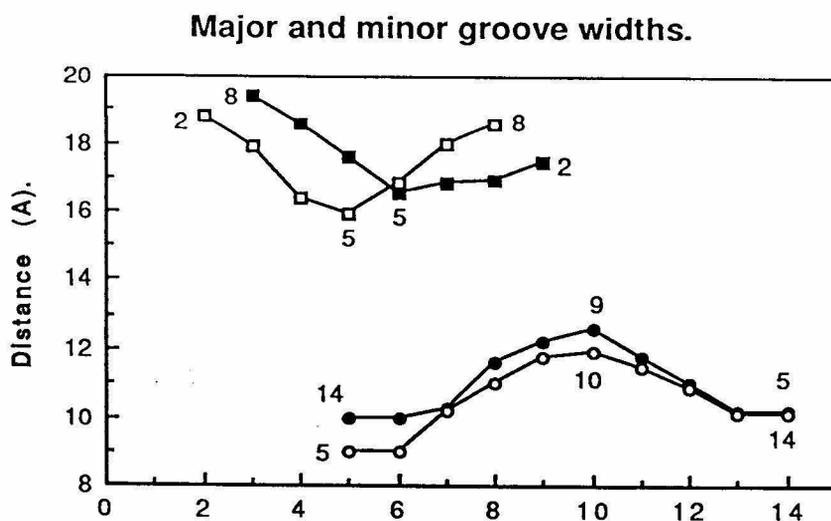


FIGURE 3-13: The major and minor groove widths of the refined structure Final_I and Final_II. *Close squares*: major groove widths in DNA I; *open squares*: major, DNA II; *close circles*: minor, DNA I; *open circles*: minor, DNA II. The curves are plotted to match the sequences below the figure. Annotated numbers indicate one of the two residues across the grooves. The two residues are always shifted to 5' for major groove, and 3' for minor groove, by 4 residues. Thus, (5) for major groove indicates widths between 5 and 20; for minor groove, 5 and 28 (as labelled in the sequence of DNA I). The widths are calculated by averaging three nearest P-P distances on the backbones.

The Common Structural Features Correlated with Protein Binding

The *Hin* recombinase binding domain, a 52 amino acid peptide, has been suggested to adopt a helix-turn-helix motif (Pabo & Sauer, 1984) recognizing the specific DNA sequences (Plaxco et al., 1989; Sluka, 1988). Figure 3-14 shows the refined DNA I been recognizing by the binding domain. The coordinates of peptide binding domain were from the model of Plaxco et al. (1989). The recognition helix, according to the model (Figure 1-4), was placed in parallel (orientated N- to C-terminals) to the DNA major groove, covering the DNA tract T6-G8 (C21-A23). The equivalent tract is T20-A22 (T7-A9) in DNA II. As discussed before, the DNA bending and major groove narrowing are originated and centered at the Pyr-Pur junctions T7-G8 (C21-A22) in DNA I or DNA II. It is very precisely this position where the recognition α -helix contacts the DNA major groove (to take advantage of the helix dipole interaction, the α -helix may tilt slightly so the C-terminal is the closest to the major groove). It can be left for speculation whether a narrowed major groove binds the α -helix better, and if yes, whether it is a universal phenomenon for helix-turn-helix DNA binding interactions. We would like to conclude by saying, the DNA bending, narrowed major grooves and other sequence-dependent structural variations have been found as common elements in different binding sites of the protein. These common and unique structural variations are well-correlated with the protein

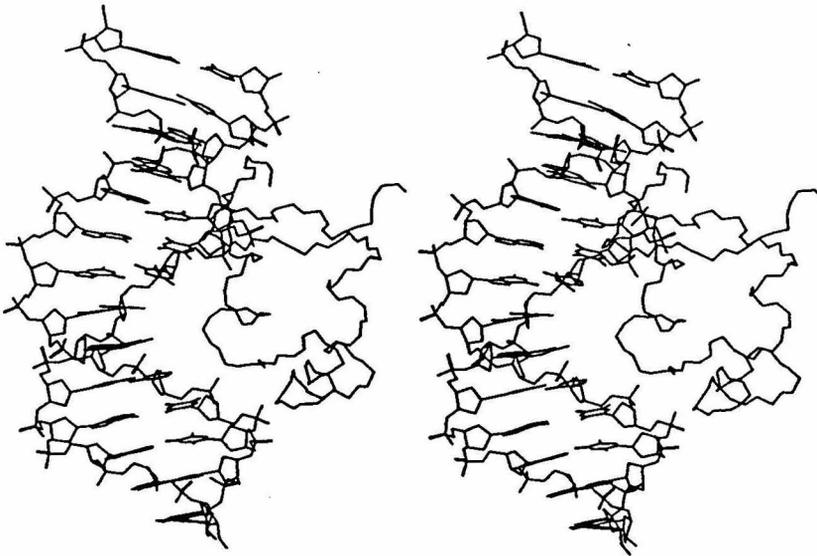


FIGURE 3-14: The main chain of the Hin binding domain (Plaxco et al., 1989) was docked to the refined structure of DNA I by matching the DNA in the model Hin binding complex with the correspondent base sequence of DNA I. The contacting backbone positions of the DNA molecules are superimposed.

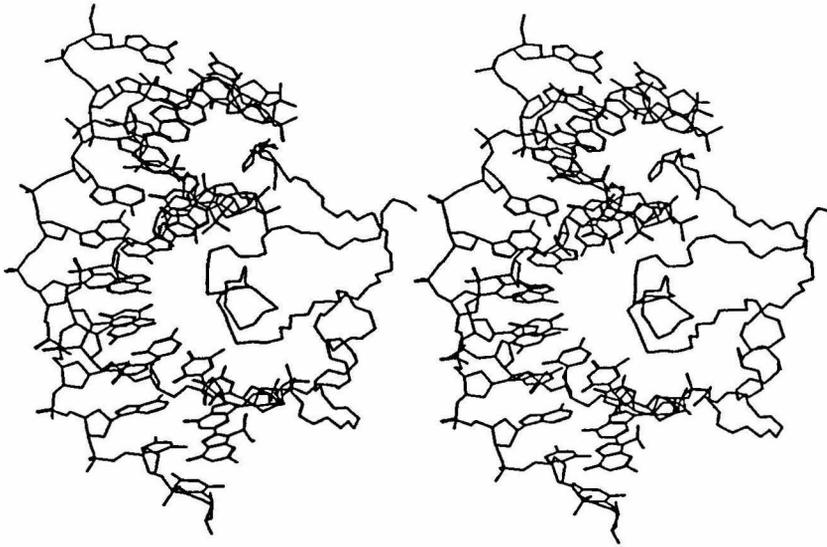


FIGURE 3-15: The same structure as in Figure 3-14. View from an angle parallel to the grooves. Narrowed major groove and compressed and deepened minor groove are shown to be recognized by the peptide chain.

interactions. It is likely, such unique DNA structural features contribute to the specificity of the interaction by enhancing the complementarities of the binding protein and the DNA recognition sites.

It is also worthwhile to mention the correlations of the minor groove alterations and the specific minor groove binding of the Hin peptide as suggested by Sluka et al. (1990). The suggested minor groove interactions occur in the region of the oligo(dA) tract, where a compressed and deepened minor groove has been observed (Figure 3-15). Such minor groove binding interactions are commonplace in DNA-drug interactions, where such minor groove structures provide increased van der Waals interactions with drugs, and the propeller-twisted bases provide possibilities of bifurcated hydrogen bondings with drugs (Sarma et al., 1990; Pelton & Wemmer, 1988). Similar interactions of the altered minor groove with arginines of the N-terminal of the protein binding domain are, therefore, feasible and may contribute to the protein binding affinities. It is interesting to note, that by bending the DNA towards the protein as discussed before, the major groove binding motif and minor groove binding peptide segment may practice their bindings and stay as a single globular domain more readily. Such an arrangement can further enhance the specific protein-DNA interactions of the binding domain.

In conclusion, we have presented certain DNA structural variations that appear to be common in the recombination system and contribute to the specific DNA-protein interactions. Ever since the the publication of the first high resolution structure of DNA (Dickerson & Drew, 1981), it has been well-recognized that DNA structures contain

significant sequence-dependent variations that are important for their biological functions such as DNA-protein interactions. Such a correlation has never been observed so vividly and in detail as in this system. It is of future interest to discuss how much of the cooperativity is involved in the DNA-protein interactions, that both DNA and protein binding domain could experience significant structure rearrangements upon the specific interactions. Our work here laid a foundation for future work along this line, that examinations of DNA and protein structure in the binding complexes vs. the free DNA and protein structure could be expected to provide clear insights to the question.

Table 3-5: Distance Constraint Set #1^a

A. Intranucleotide distances

Strand #1: (1-14)

i	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
H1'	H8/H6	2.6	3.0			3.2	3.2		2.9	3.1	3.2	2.8	2.9	3.3	3.3
H2'	H8/H6	2.4		2.7	2.7	2.8	2.8	2.5		2.6		3.4	2.8	2.7	
H2 ^W	H8/H6	2.8	2.9		2.8	3.5	3.0	2.9		2.8		3.3	2.6	3.2	
H3'	H8/H6	3.5	3.1	3.1	3.4	3.4	3.0	3.2	3.5	3.3			3.1	3.1	2.8
H4'	H8/H6		3.9	4.2	4.0			4.2			4.4	4.1	3.8		
H2'	H5/CH3					3.3								3.0	3.5
H2	H1'											4.8	4.3		
H1'	H2'	2.6	2.6		2.8		2.6	2.7		2.9	2.4	2.8	2.4	2.3	
H1'	H2 ^W	1.7	2.3			2.3	2.3			2.1	2.7	2.2	2.2	2.0	
H1'	H3'			3.6		3.3	3.1		3.2	3.2	3.2	3.5	3.6		3.8
H1'	H4'	2.4	2.7	2.4	2.2	3.2	3.8	2.6	3.5	2.8	2.9	2.9	2.9	2.6	2.8
H2'	H3'	2.3	2.3			2.7	2.8	2.4					2.4		
H2 ^W	H3'	2.3	2.4			2.4						2.6	2.7	2.6	
H3'	H4'	2.3	2.4	2.2		2.8	2.5	2.5	2.3				2.6	2.6	2.5

Strand #2: (15-28)

i	i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
H1'	H8/H6	2.6	3.0			2.9	3.0	3.2	3.2	3.1	3.0	3.0	3.2	3.0	3.3
H2'	H8/H6	2.5		2.7		2.7	2.9	2.7	2.8	2.4	2.6	2.6		2.6	
H2 ^W	H8/H6	2.9	2.9				3.0	2.6	2.9	3.0	2.8	2.6	3.0	2.9	
H3'	H8/H6	3.4	3.1	3.1			3.4	3.3	3.3		3.5		4.1	3.2	2.8
H4'	H8/H6		3.9	4.1						3.8		4.4	4.4		
H2'	H5/CH3							2.9						3.6	
H2	H1'								4.5	4.0		4.1	4.2		
H1'	H2'	2.6	2.6		2.5			2.7	2.4	2.8	2.6	2.6	2.5	2.4	
H1'	H2 ^W	2.0	2.4		2.2		2.3	2.3	2.3	2.3	2.4	2.3	2.3	2.0	
H1'	H3'			2.6			3.4		3.7		3.8	3.6			3.8
H1'	H4'	2.5	2.7	2.6	2.9	2.9	2.8	2.6	3.1	3.0	3.3	3.0	3.0	2.7	2.8
H2'	H3'	2.3	2.3				2.7	2.5	2.4		2.5		2.6		
H2 ^W	H3'	2.4	2.5		2.1					2.1		2.6	2.5	2.6	
H3'	H4'	2.4	2.4	2.3			2.4	2.3	2.6	2.5	2.5		2.6	2.6	2.5

 B. Sequential distances

Strand #1: (1-14)

i	i+1	1-	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-
		2	3	4	5	6	7	8	9	10	11	12	13	14
H1'	H8/H6	2.4	2.9		3.2	3.1	3.1	3.5	3.0	3.0	3.0	2.7	2.9	
H2'	H8/H6	2.8		3.0	3.0	2.9	4.1	3.4		2.6	2.4	2.6	3.2	2.8
H2 ⁿ	H8/H6	2.7		3.1	3.3	3.0	2.9	3.0			3.3	2.7	3.0	
H3'	H8/H6		4.3		4.4		4.2					3.8	4.4	3.4
H1'	H5/CH3		3.4	3.5	3.6	3.9	4.4						3.4	
H2'	H5/CH3		3.0	3.3	2.9		3.3						3.4	3.2
H2 ⁿ	H5/CH3		3.3	3.3	3.3	3.4	3.5						2.9	3.2
H3'	H5/CH3		3.9	3.8		3.9	4.0							
H2	H1'										4.3	3.9	3.3	
H8/H6	H8/H6			3.7	3.9	4.5	3.9	3.5					3.9	
H8/H6	H5/CH3		3.2	3.2	3.5	3.4	3.2						3.1	3.6
H5/CH3	H5/CH3			3.7		3.7								3.7
H2	H2									3.7	4.1	3.5		

Strand #2: (15-28)

i	i+1	15-	16-	17-	18-	19-	20-	21-	22-	23-	24-	25-	26-	27-
		16	17	18	19	20	21	22	23	24	25	26	27	28
H1'	H8/H6	2.5	2.9	2.8		3.1	3.2	3.8	3.1	3.2	2.9	2.7	3.3	
H2'	H8/H6	2.8		3.0		3.3	3.4	3.6	3.3	3.2	3.4			2.8
H2 ⁿ	H8/H6	2.7	3.0			3.0	3.0	3.5	2.6	2.9	2.4	2.6	2.9	
H3'	H8/H6		3.7							4.5		3.4	4.4	3.8
H1'	H5/CH3		3.4	3.6	3.6	3.2	3.3						3.8	
H2'	H5/CH3		3.0	3.3	2.8	3.0	3.2							3.0
H2 ⁿ	H5/CH3		3.3	3.3			3.0						2.9	4.2
H3'	H5/CH3		3.9	3.8										
H2	H1'									4.2		3.6	3.3	
H8/H6	H8/H6		4.8	3.8		3.6	4.1						4.2	
H8/H6	H5/CH3		3.2	3.2			3.4						3.0	
H5/CH3	H5/CH3			3.7			4.1							3.7
H2	H2								3.8			3.6		

C. Interstrand distances

Strand #1: (1-14)

i	j	dps(-)	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	14-
			28	27	26	25	24	23	22	21	20	19	18	17	16
H2	H1'								4.4	4.1	4.0	3.7			

Strand #2: (15-28)

i	j	dps(-)	16-	17-	18-	19-	20-	21-	22-	23-	24-	25-	26-	27-	28-
			14	13	12	11	10	9	8	7	6	5	4	3	2
H2	H1'								4.3		3.8	3.8			

^aNOESY experiment was performed with $\tau_m = 300$ ms. Distances (\AA) were calculated using the full spin matrix analysis. In table C, $d_{ps}(-)$ denotes interstrand-sequential distances to the 3'-neighboring nucleotide of the complementary base (shorthand by Wüthrich, 1986). A statistics of the distances is available in Figure 3-4.

Table 3-6: Distance Constraint Set #2^a

A. Intranucleotide distances

Strand #1: (1-14)

i	i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
H1'	H8/H6	2.8	3.0			3.1	3.3		3.6	3.0	3.6	3.2	3.0	3.2	3.2
H2'	H8/H6	4.3		2.3	2.4	2.5	2.3	2.1		2.3			2.6	2.3	
H2 ⁿ	H8/H6	2.7	2.8		2.9	2.8	2.6	2.6		2.8		2.7	2.3	2.9	
H3'	H8/H6	4.2	3.8	3.0	3.2	3.1	2.9	3.3	3.8	3.3			3.2	3.1	2.4
H4'	H8/H6							4.2			4.3				
H2'	H5/CH3					3.1									
H2	H1'											4.5			
H1'	H2'	3.8	2.5		2.2		2.4	2.5	2.2	3.3	2.5	2.3	2.3	2.1	
H1'	H2 ⁿ	1.9	2.0			2.0	2.0	1.9	2.5	2.0	2.3	2.1	1.9	1.9	
H1'	H3'						4.2		3.2		3.3				4.0
H1'	H4'	2.7	2.8	2.1		3.0	3.6	2.4	4.7	2.6	2.9	2.8	2.6	2.6	2.6
H2'	H3'	2.1	2.1	2.4		2.4	2.2	2.2					2.3	2.1	
H2 ⁿ	H3'	2.3	2.2			2.4		2.5				2.2	2.2	2.6	
H3'	H4'	2.4	2.4	2.1		3.4	2.4	2.4	2.3				2.5	2.5	2.4

Strand #2: (15-28)

i	i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
H1'	H8/H6	2.8	3.0				3.1	4.8	3.5	3.3	3.1	2.8	3.5	3.0	3.1
H2'	H8/H6	4.2		2.5		2.4	2.5	2.5	2.4	2.3	2.6	2.2		2.4	
H2 ⁿ	H8/H6	2.7	2.8				2.7	2.3	2.7	2.9	2.6	2.7	3.2	3.3	
H3'	H8/H6	4.2	3.8	3.0			3.3	3.3	3.5		3.8		4.1	3.0	2.9
H4'	H8/H6			4.5				4.6	4.8			4.3			4.7
H2'	H5/CH3							2.8						3.5	
H2	H1'									4.5		4.4	4.2		
H1'	H2'	3.7	2.5		2.2			3.0	2.1	2.8	2.4	3.1	2.4	2.9	
H1'	H2 ⁿ	1.9	2.0		1.8		2.1	2.0	2.1	2.1	2.2	2.0	2.0	1.9	
H1'	H3'			2.7			4.3			4.0					3.9
H1'	H4'	2.7	2.8	2.4	2.7		2.7	2.5	3.4	2.8	4.1	2.9	3.1	2.6	2.6
H2'	H3'	2.1	2.1	2.4		2.5	2.6	2.3	2.2		2.7			2.4	
H2 ⁿ	H3'	2.3	2.2		2.0			3.2					3.0	2.5	
H3'	H4'	2.4	2.4	2.4			2.3	2.3	2.6	2.3	2.7		2.7	2.5	2.4

 B. Sequential distances

Strand #1: (1-14)

i	i+1	1- 2	2- 3	3- 4	4- 5	5- 6	6- 7	7- 8	8- 9	9- 10	10- 11	11- 12	12- 13	13- 14
H1'	H8/H6	2.4	2.7		3.4	3.0	3.0	3.3	2.8	3.1	2.7	2.4	2.6	4.7
H2'	H8/H6	2.8		2.9	2.7	2.8	3.1	2.9		2.5	2.2	2.1	2.6	2.8
H2 ⁿ	H8/H6	3.0	2.6	2.4	2.7	2.7	2.5	3.4			2.9	3.0	2.5	2.4
H3'	H8/H6					4.7						3.5	3.7	3.7
H1'	H5/CH3		3.8	3.2	3.2									3.8
H2'	H5/CH3		2.8	3.0	2.7		3.0						3.6	2.8
H2 ⁿ	H5/CH3		3.2	3.2		3.2	3.2						2.9	4.0
H3'	H5/CH3			4.5										
H2	H1'									4.5	4.0	3.1		
H8/H6	H8/H6				4.0			4.3						
H8/H6	H5/CH3		3.1	3.1		3.4	3.2						3.3	3.3
H5/CH3	H5/CH3			3.5		3.8								4.3
H2	H2										3.8			

Strand #2: (15-28)

i	i+1	15- 16	16- 17	17- 18	18- 19	19- 20	20- 21	21- 22	22- 23	23- 24	24- 25	25- 26	26- 27	27- 28
H1'	H8/H6	2.4	2.7			3.0	3.2	3.5	2.8	3.1	2.6	2.6	3.0	4.5
H2'	H8/H6	2.8		2.7		3.8	2.7	3.3	3.3		4.3		3.2	2.8
H2 ⁿ	H8/H6	3.0	2.6			2.6	2.9	3.1	2.8	2.6	2.2	2.5	2.5	2.5
H3'	H8/H6			4.5							4.4	4.2	3.9	
H1'	H5/CH3		3.8	3.2									4.0	4.5
H2'	H5/CH3		2.8	3.0	2.8	3.0	3.4							3.6
H2 ⁿ	H5/CH3		3.2	3.3			3.0						3.5	4.3
H3'	H5/CH3			4.5										
H2	H1'									4.4		3.3	3.2	
H8/H6	H8/H6													4.9
H8/H6	H5/CH3		3.1	3.1			3.4						3.1	
H5/CH3	H5/CH3			3.5										
H2	H2								4.6			3.9		

C. Interstrand distances

Strand #1: (1-14)

i	j	dps(-)	2-	3-	4-	5-	6-	7-	8-	9-	10-	11-	12-	13-	14-
			28	27	26	25	24	23	22	21	20	19	18	17	16
H2	H1'										4.2	3.9	3.5		

Strand #2: (15-28)

i	j	dps(-)	16-	17-	18-	19-	20-	21-	22-	23-	24-	25-	26-	27-	28-
			14	13	12	11	10	9	8	7	6	5	4	3	2
H2	H1'											3.9	3.6		

^aNOESY experiment was performed with $\tau_m = 200$ ms. A statistics of the distances is available in Figure 3-5.

REFERENCES

- Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson.* **87**, 475.
- Bruist, M. F., Horvath, S. J., Hood, L. E., Steitz, T. A., & Simon, M. I. (1987) *Science* **235**, 777.
- Calladine, C. R. & Drew, H. R. (1984) *J. Mol. Biol.* **178**, 773.
- Dickerson, R. E., & Drew, H. P. (1981) *J. Mol. Biol.* **149**, 761.
- Glasgow, A. C., Bruist, M. F., & Simon, M. I. (1989) *J. Biol. Chem.* **264**, 10072.
- Gochin, M., & James, T. L. (1990) *Biochemistry* **29**, 11172.
- Hughes, K. T., Youderian, P., & Simon, M. I. (1988) *Genes Dev.* **2**, 937.
- Katahira, M., Sugeta, H., & Kyogoku, Y. (1990) *Nucleic Acid Res.* **18**, 613.
- Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acid Res.* **15**, 5845.
- Koo, H. -S., & Crothers, D. M. (1987) *Biochemistry* **26**, 3745.
- Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293.
- Pelton, J. G., & Wemmer, D. E. (1988) *Biochemistry* **27**, 8088.
- Plaxco, K. W., Mathiowetz, A. M., & Goddard III, W. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 984.
- Sarma, M. H., Gupta, G., Garcia, A. E., Umemoto, K., & Sarma, R. H. (1990) *Biochemistry* **29**, 4723.
- Sluka, J. (1988) Ph.D. Thesis (California Institute of Technology, Pasadena, CA).

Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., & Dervan, P. B. (1990) *Biochemistry* **29**, 6551.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Chapter 4

Studies of the Protein-DNA Interactions between the Hin 52mer Peptide and DNA Oligomers

The Hin 52mer peptide is the C-terminal 52 amino acid segment of Hin recombinase, a 190 amino acid protein (Hughes et al., 1988). The Hin peptide was obtained by chemical synthesis and has been established as the DNA binding domain of the recombinase. The binding specificity of the peptide is similar to the intact Hin protein (Glasgow et al., 1989; Bruist et al., 1987). The large quantity of homogeneous peptide available through peptide synthesis enables us to investigate the properties of the peptide and its interactions with the DNA oligomers bearing specific recognition sites of the Hin recombinase.

The following studies of the Hin peptide are reported in this chapter: 1) the structural properties of the peptide in physiological solution and in conformation-enhancing solutions, i.e., solutions containing trifluoroethanol (TFE); 2) the binding assay of the peptide to DNA oligomers with specific or non-specific binding sequences; 3) preliminary NMR studies of the DNA-protein domain binding complex.

MATERIALS AND METHODS

Peptide Synthesis

The Hin 52mer peptide was synthesized on an ABI peptide synthesizer using the solid-phase method with phenylaceticamidomethyl (PAM)-resin and butoxycarbonyl (BOC) amino acids. A purification

process using affinity chromatography with a DNA binding site linked to a cellulose resin was planned but aborted later since the peptide appear to adhere to the resin. Nevertheless, the peptide seems to possess satisfactory purity for further studies after the crude product from hydrogen fluoride (HF) cleavage was purified using reverse phase HPLC on a C₄ column. The fractions corresponding to the main peak were collected and confirmed as the Hin peptide by mixing with a previous sample supplied by Sluka et al. The amino acid compositions and chemical purity of the peptide were established by Bruist et al. (1987). The peptide was desalted, as needed to recover from other buffers, by eluting through a gel-filtration column packed with Sephadex G-10 (size 40-120 μ) in a Pharmacia HR 10/30 column. Examinations of the peptide by NMR under denaturing conditions indicate the correct compositions of the aromatic amino acid (His/Phe/Tyr, 2:2:2). No apparent impurities were observed in NMR spectra. The biochemical purities, i.e., percentages of folding and functioning, were established later by binding assays and NMR of the folded peptide under conformation-enhanced conditions.

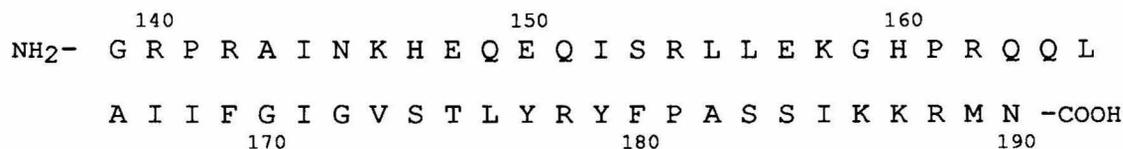


FIGURE 4-1: The amino acid sequence of the Hin 52mer peptide. The 52mer sequence is that of amino acids 139 to 190 of the Hin protein, as labelled. One-letter codes are used for amino acid residues in the sequence.

The Binding of the Peptide to the DNA

The binding experiment was done on a Bio-Rad TSK-125 Bio-Sil column with an aqueous eluate of 0.5 M NaCl, 10 mM phosphate, pH 7.0. HPLC runs were conducted with measured amounts of the peptide, one of the DNA oligomers to be tested for binding activities, and the corresponding peptide-DNA mixture (the binding complex), respectively. The resulting chromatograms were compared at 214 nm where both peptide and DNA absorb UV signals. If no interaction happens between the DNA and the peptide, the chromatogram of the mixture would be the sum of two other chromatograms, as was seen in control experiments with nonspecific DNA oligomers. However, the additive relationship would not be valid if a binding between the DNA and the peptide occurred. The peptide peak would have its retention time shifted, or be reduced with a corresponding increase in the intensity of the peak representing both the DNA and the DNA-peptide complex (DNA and DNA-peptide complex are not resolved). A spectral check of that chromatographic peak would reveal if it is from pure DNA or DNA with protein. The result was clear when the experiment was done with a molar ratio of 1:1 (peptide/DNA) in the injecting mixture, where the resulting peptide peak was either undisturbed or disappeared in the mixture chromatogram, depending on the oligonucleotide used. This is explained by the absence or presence of the binding of the peptide to the oligonucleotide (Figures 4-10c and Figure 4-11c).

The DNA-peptide binding complex can be redissolved from the binding precipitates or maintained soluble while mixing DNA oligomers and the Hin peptide, by a solution of 0.2 M KCl, 0.1 M potassium phosphate, and pH* 6.5 in D₂O. The solubility was about 0.5 mM for 1:1 DNA-peptide binding complex. At increased ionic strength and temperature, the solubility can be further improved, with the undesirable possibilities of disrupting the binding complex. The DNA-protein binding can be disrupted by high ionic strength (1M KCl), as seen by NMR in this study (data not shown) and by Boelens et al. (1987).

NMR Spectroscopy and CD Spectroscopy

All NMR spectra were recorded on Bruker AM500 spectrometers at room temperature (25°C) unless otherwise specified. NOESY experiments in water or in TFE/water were performed with a 1-3-3-1 pulse (Hope, 1983) incorporated as the detection 90° pulse in the NOESY pulse sequence (Figure 2-2) to suppress the water resonance. In all other experiments, residual HOD signal was suppressed by presaturation. The COSY of the Hin peptide in TFE/D₂O, as reported in this chapter, is obtained by using the double-quantum filtered (DQF) COSY techniques. The NOESY of the DNA-peptide binding complex is performed in D₂O with mixing times τ_m of 150 ms or 250 ms.

The circular dichroism (CD) spectra was measured on a JASCO J-600 Spectropolarimeter using cells with pathlengths of 0.01, 0.1 or 1 cm. The α -helical contents of the peptide were estimated using the SSEAX program from the same company.

The Folding of the Hin Peptide in Solutions with Trifluoroethanol

The Trifluoroethanol used in the CD measurement of the peptide helical contents was from Aldrich. The deuterated TFE with 98%D and 99.96%D used in NMR experiments were from ICN Biomedicals and Cambridge Isotope Laboratory, respectively. The buffer conditions for TFE titrations and peptide folding are always 50 mM sodium phosphate and pH* 3.0 (in D₂O) or pH 3.4 (in H₂O). The peptide concentrations were between 0.1 μ M to 0.16 mM for CD measurement, and 0.05 mM to 1.6 mM for NMR experiments. The CD spectra and the secondary structure contents estimated were independent of the peptide concentration, in contrast to the result with the peptide in aqueous solutions in the absence of TFE (Figure 4-3). The peptide effectively folds into a stable three-dimensional structure in the above-mentioned aqueous buffer with 12%(v) TFE. It starts a reversible aggregation, as observed by NMR under this condition, at a concentration of 2.0 mM.

RESULTS AND DISCUSSIONS*The Structural Properties of the Hin Peptide*

The NMR spectrum of the Hin peptide in physiological conditions is shown in Figure 4-2. Similar spectra were obtained over a wide range of trial conditions, including varying ionic strength, phosphate buffer concentration, temperature and the concentration of the peptide. The broad resonances are usually an indication of oligomerizations or

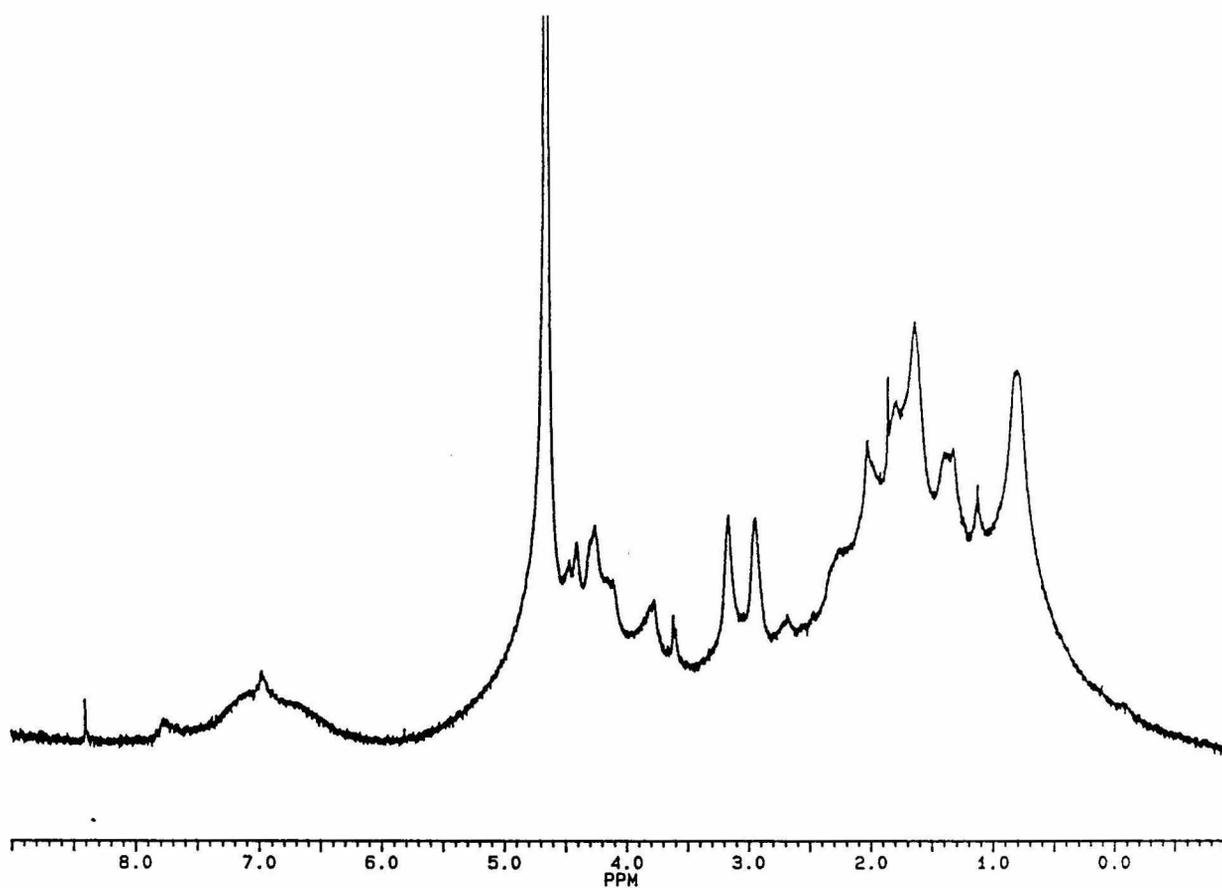


FIGURE 4-2: The NMR spectrum of the Hin 52mer peptide under 20 mM NaCl, 20 mM Pi, pH* 7.0 in D₂O. The broad linewidths are probably due to aggregation of the peptide, implying intermolecular interactions and unstable peptide conformation. The spectrum is relatively featureless.

aggregations of the peptide, causing shortened T_2 relaxation times, which usually happens at high concentrations of certain proteins. For the Hin peptide, however, spectra of this appearance started from a very low concentration at which a conventional NMR experiment can be carried out (0.01 mM). The molecular mechanism to account for this broadening of signals can be two related aspects of the structure, namely, unstable conformation intramolecularly and strong intermolecular interactions. For Hin peptide, both are likely to be the cause.

The concerns of an unstable conformation for the Hin peptide started early on in the research. Although the synthetic peptide was identified as a single component and sequence by various methods, the chromatograms of the purified product on heptyl-agarose (Bruist et al., 1987) or on Sepharose size-exclusion columns were showing a very broad peak for the peptide. It also appears to adhere strongly to certain resins, e.g., cellulose and the media in ion-exchange columns. The peptide, with a molecular weight of 6000, is also found to escape dialysis membranes of 1000 molecular weight cutoff (Arnold, F. H., unpublished results). In a circular dichroism (CD) measurement, the helical content estimated for the peptide was 23% at 0.1 mM, and decreases with lowering concentrations (Figure 4-3). Such a varying helical content has been interpreted before as a result of unstable peptide conformation and peptide intermolecular interactions (Brems et al., 1987).

The aggregation of the peptide may be broken by lowering the pH to about 3.4. Under such conditions the spectra of the peptide indicated typical denatured or random-coil conformation. As shown in Figure. 4-

7A, the two histidines, phenyl alanines and tyrosines are all at identical chemical shifts and not resolved from each other.

It is, therefore, fair to conclude that the peptide does not have a stable three-dimensional conformation alone in aqueous solutions. This does not exclude the possibility of a stable conformation in the oligomerized peptide chains (Wade-Jardetzky et al., 1978).

There is a disadvantage with the approach of synthetic peptides vs. the traditional way of obtaining protein domains by protease cleavage. That is, we may never know, before solving the complete protein structure, whether the synthetic polypeptide actually exists as a stable globular domain, which would be a prerequisite for the traditional approach to success. Although such a disadvantage can be turned around to synthesize arbitrarily modified versions of the original sequence so the conformation can be stabilized and intermolecular oligomerization eliminated, such practise is only possible with well-known protein structures (Wright, P. E., communications). Nevertheless, this could be a future direction of this research.

The Peptide Folding under Conformation-enhancing Conditions

Trifluoroethanol (TFE) is commonly used as an α -helix stabilizing reagent. Figure 4-4 shows the monotonically increased α -helical content with the additions of trifluoroethanol to a peptide solution at acidic pH (3.4), measured by CD spectroscopy between wavelengths 180-260 nm (Figure 4-3). Such an increase can be observed in almost all peptides (Lehrman et al., 1990; Nelson et al., 1989; Mammi et al., 1988). More

interestingly for the Hin peptide, the CD signals near 275 nm reaches maximum at 12% TFE, but is reduced at higher TFE concentration (Figure 4-5). The UV absorbances of the Hin peptide at 275 nm are attributed to the two tyrosines, and corresponding CD signals should arise from the conformational fixations of the tyrosines, implying that a stable *three-dimensional structure* appears at a TFE content of near 12%.

NMR experiments provided conclusive evidences of a folded peptide conformation under such conditions. Figure 4-6 shows the resonances of the peptide disperses and collapses with the titration of TFE, with an optimum concentration being 12%(v) or 3.4mol% (estimated by combining other experimental results). Figure 4-7 compares the NMR spectra of the denatured and the folded Hin peptide. The upfield-shifted methyl resonances are particularly strong evidence for a stable protein conformation. The large-scale upfield shifts of methyl resonances were caused by the proximity of an aromatic ring and the ring current it carries (Wüthrich, 1986). A single set of resolved resonances for each of the aromatic residues (His, Phe and Tyr, *vide infra*) indicated a unique conformer. The peptide folding is also strongly supported by the observations of exchangeable amide protons in D₂O lasting up to three hours (data not shown).

The aromatic residues are identified through two methods. The protons from two His were obviously seen from the spectrum in Figure 4-7B. The aromatic protons of two Phe and two Tyr were identified in a

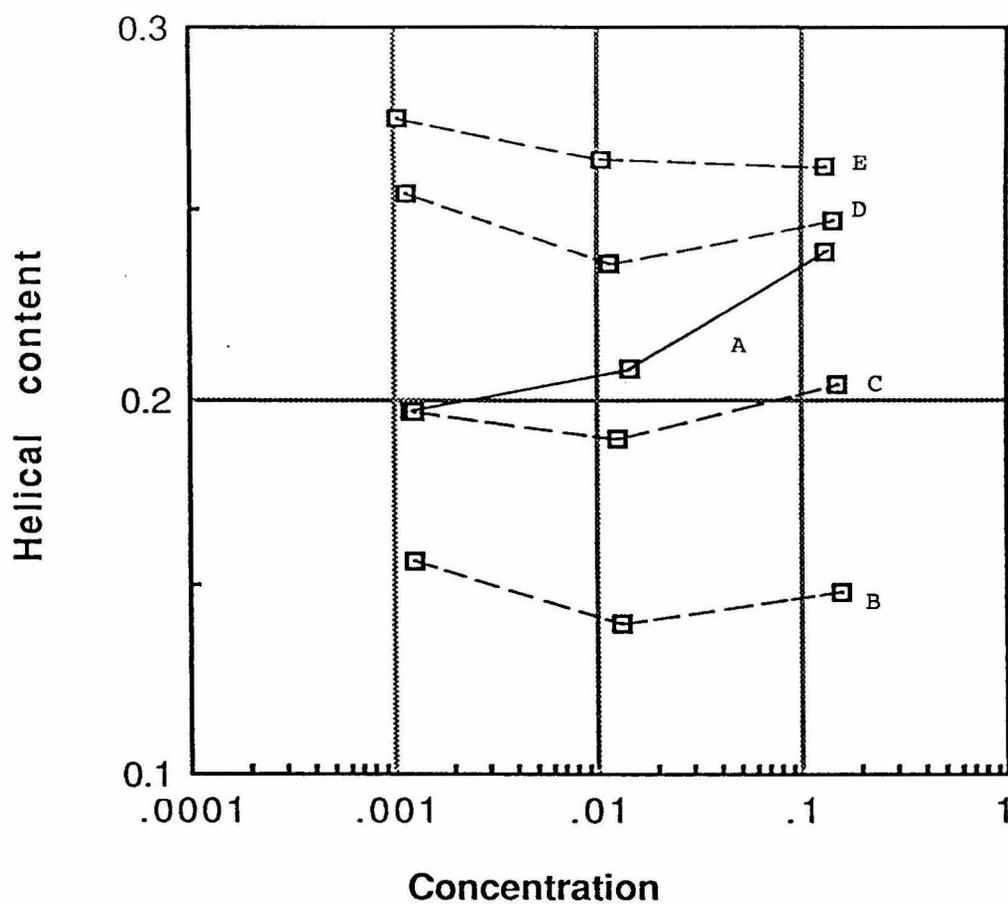


FIGURE 4-3: Concentration dependence of the helical contents of the peptide, estimated from CD measurements. *Solid line:* A. pH 7.6, 20 mM phosphate and NaCl. *Dashed lines:* pH 3.4, 50 mM phosphate; B. 0%(v)TFE, C. 5%TFE, D. 12%TFE, E. 20%TFE.

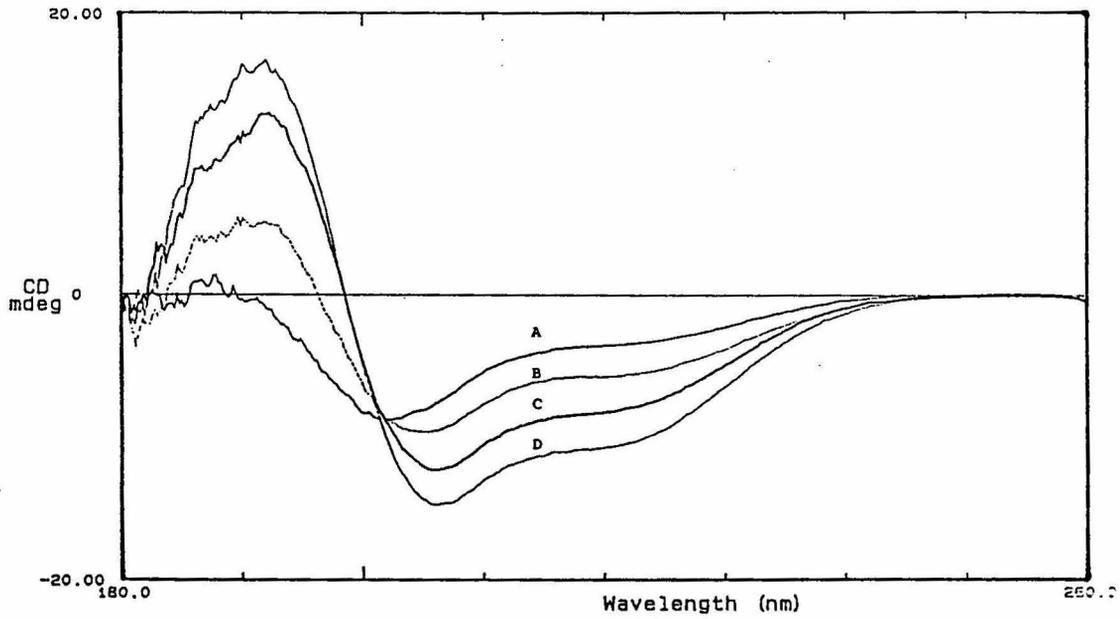


FIGURE 4-4: Circular dichroism spectra of the Hin peptide in buffer of pH 3.4 and 50 mM phosphate. Spectrum range is between 180 nm and 260 nm. A. 0%TFE, B. 5%TFE, C. 12%TFE, D. 20%TFE. The helical content of the peptide increases as percentage of TFE increases.

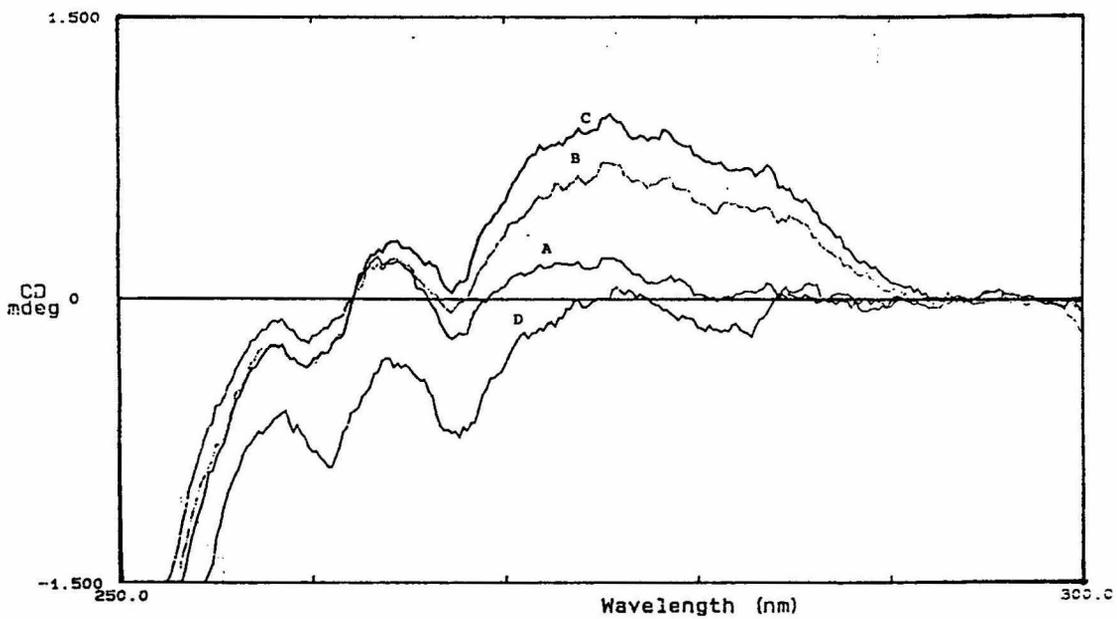


FIGURE 4-5: CD of the peptide in the range of 250 nm to 300 nm. **A.** 0%TFE, **B.** 5%TFE, **C.** 12%TFE, **D.** 20%TFE. The signal at 275 nm peaks at 12%TFE (C).

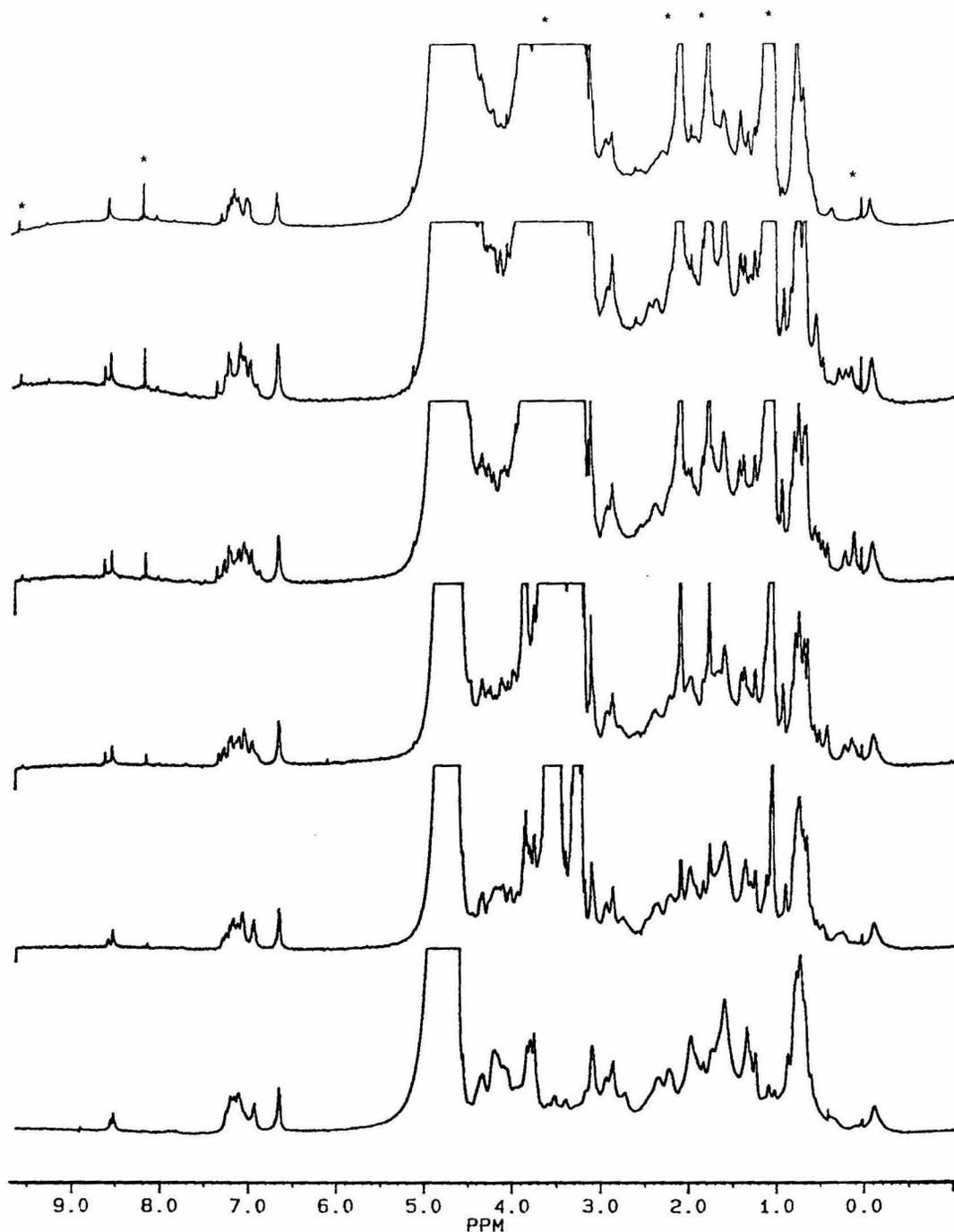


FIGURE 4-6: NMR spectra of the Hin peptide showing conformation changes corresponding to TFE titrations in pH 3, 50 mM buffer. From the bottom: 0%, 5%, 9%, 13%, 17%, 23%(v) TFE. The optimum is near 13%TFE where the resonances have the largest dispersions. *Impurities from TFE (98%D).

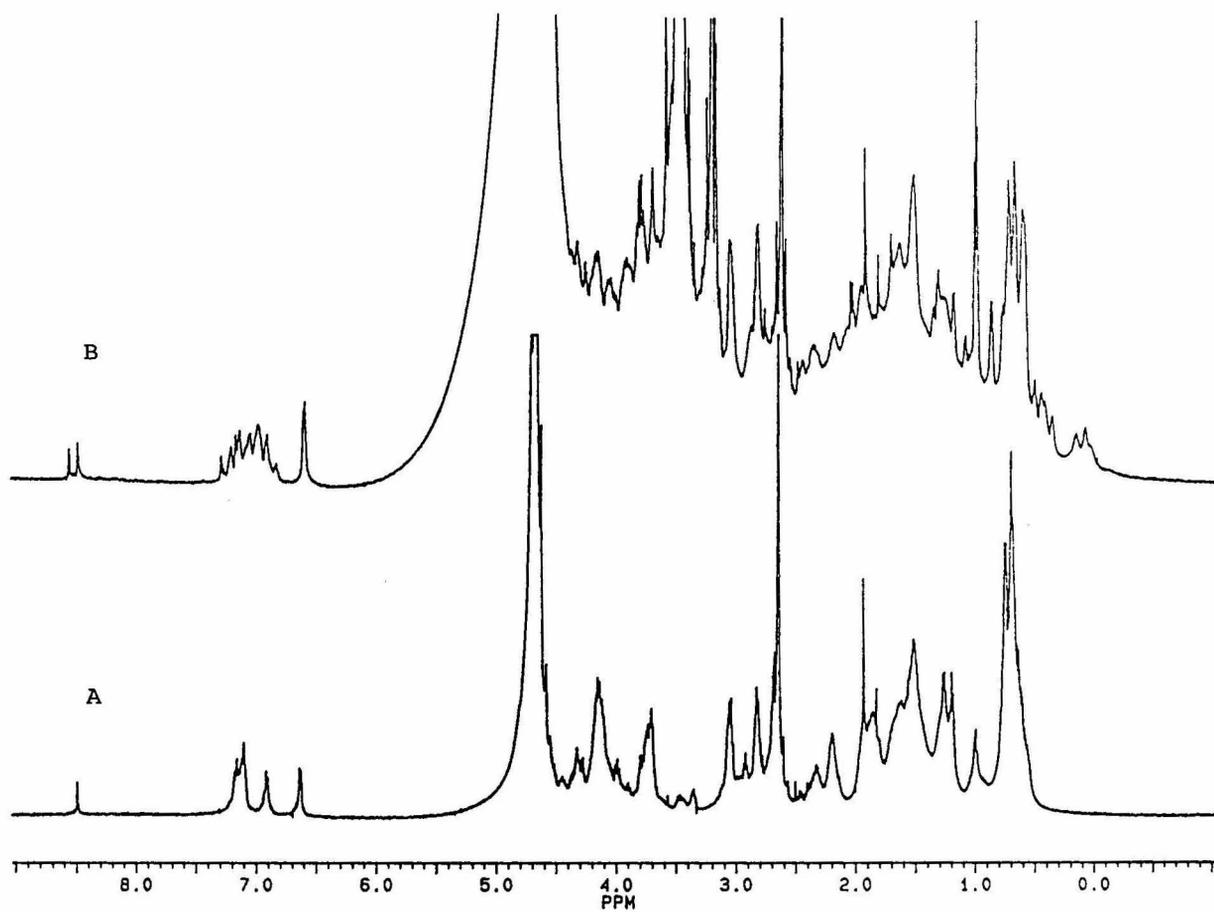


FIGURE 4-7: The Hin peptide before and after folding in pH 3 solutions. **A.** salt free; **B.** 50 mM phosphate, 12%(v) TFE.

double-quantum filtered COSY experiment (Figure 4-8). In the same experiment, 35 α -carbon protons can be directly identified to form J-coupling cross-peak patterns with β -protons on the same residue. In the case of Gly, geminal α -proton couplings were seen (Figure 4-9). Full sequential resonance assignments can not be completed at this stage as individual spin systems need to be identified. The main difficulties were that some of the antiphase cross-peak patterns were cancelled out or obscured due to broadened linewidths, a phenomenon possibly arises from the larger spin-spin relaxation properties of TFE molecules compared to D₂O. A possible future solution to this is the use propanol instead of TFE (Acharya et al., 1990).

In further NMR studies, we observed continuous stretches of peptide amide proton-amide proton (NH-NH) connectivities (d_{NN}) in the NOESY experiment of the Hin peptide in water (Figure 4-10). These stretches of NOE connectivities are one of the characteristics of α -helical secondary structures, which are part of the helix-turn-helix model proposed for the binding domain.

To conclude these experiments and observations, Hin 52mer peptide can be folded in a relatively artificial environment (pH 3.4, 12% TFE). The folded peptide has a well-defined three-dimensional structure as seen by the dispersed aromatic residues, shifted methyl groups, slow exchanging amide protons, resolution of the majority of α -carbon protons and existence of stretches of d_{NN} NOE in water.

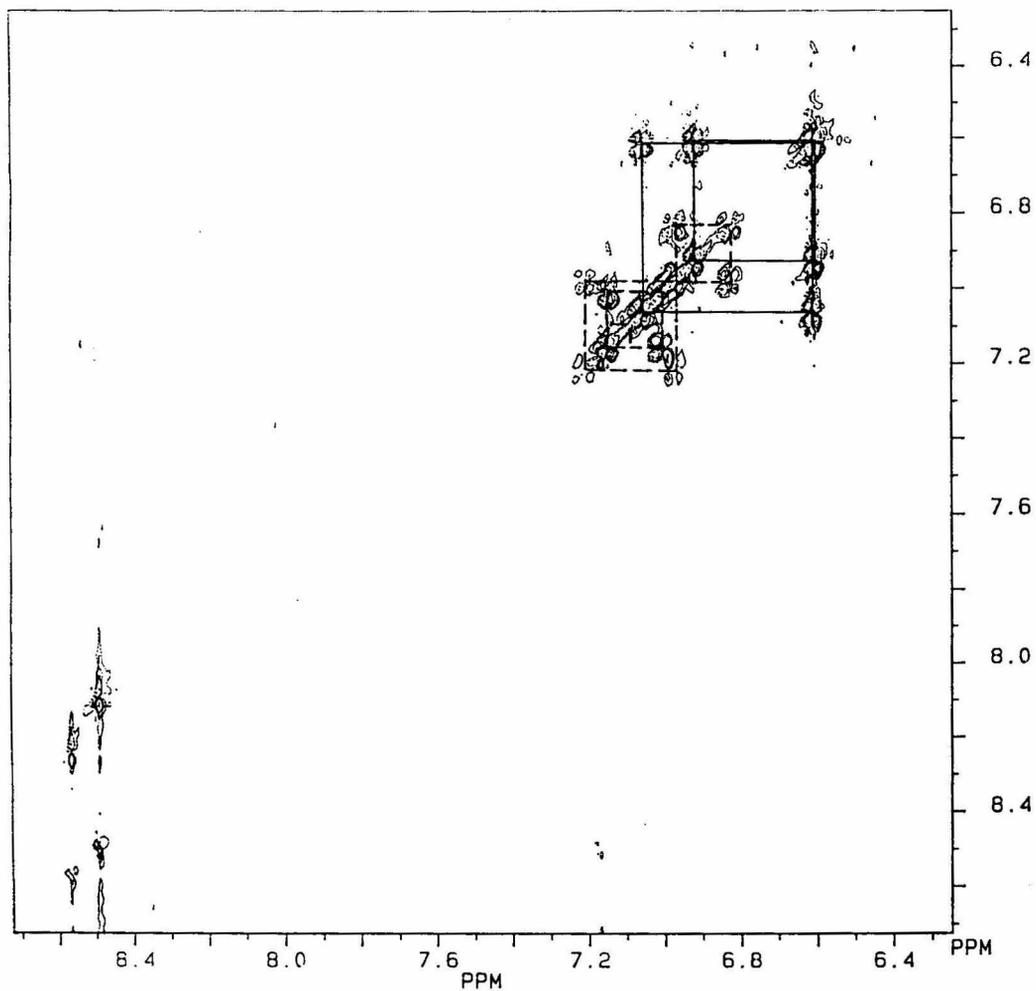


FIGURE 4-8: The aromatic region of a DQF-COSY of the folded Hin peptide in buffer 50 mM phosphate, 12%(v) TFE, pH 3. Cross-peaks from protons of Tyr and Phe are shown.

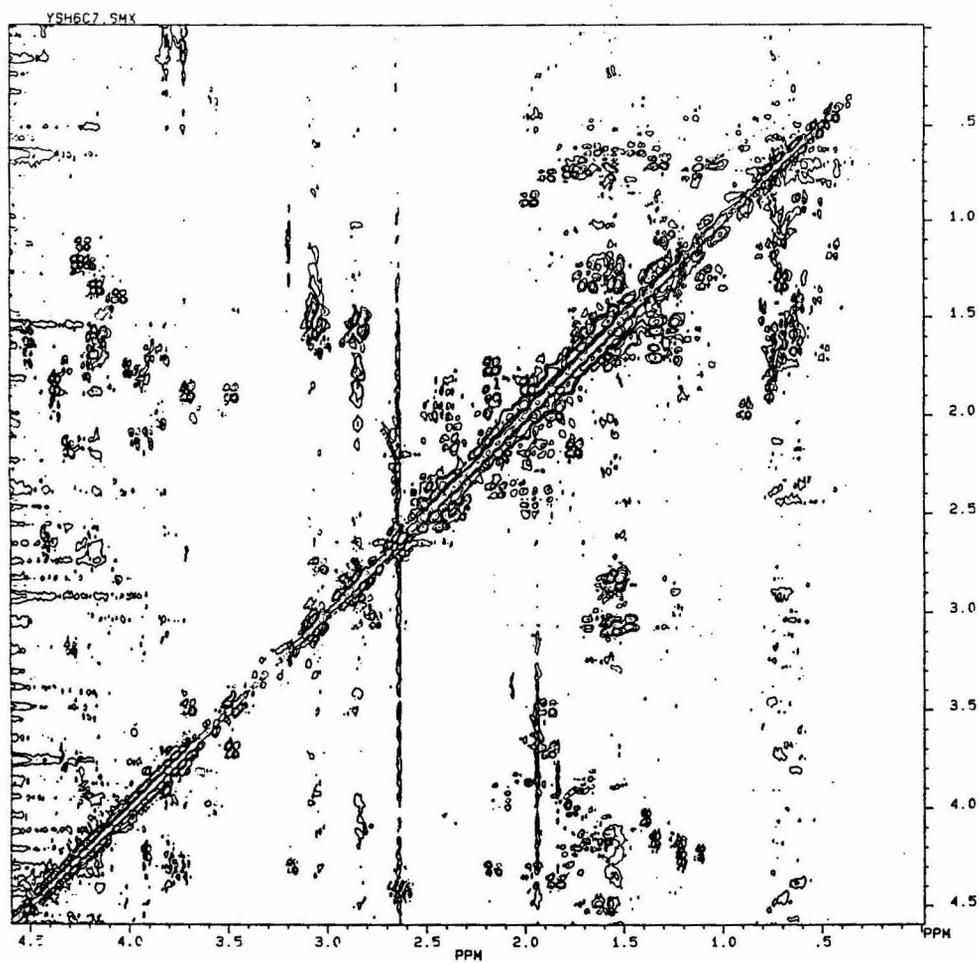


FIGURE 4-9: The aliphatic region of a DQF-COSY of the folded Hin peptide in buffer 50 mM phosphate, 12%(v) TFE, pH 3. The α -carbon protons are in the region 4.5-3.5 ppm.

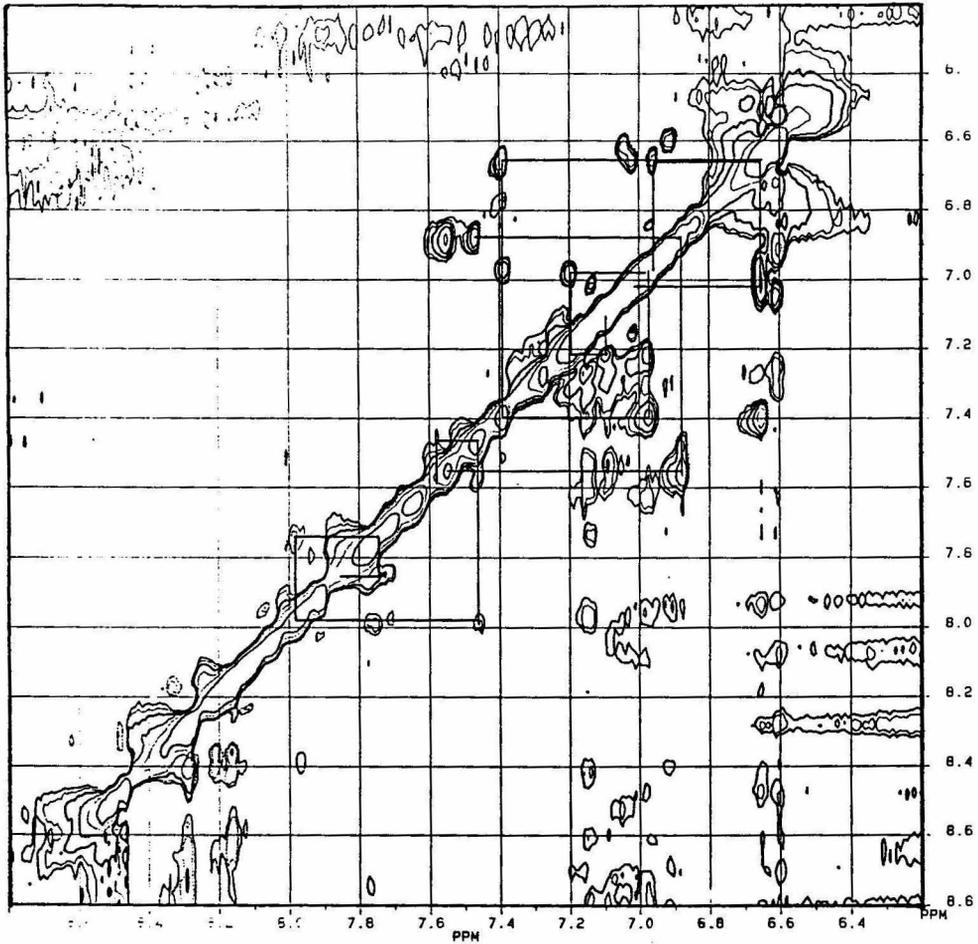


FIGURE 4-10: The amide-amide proton NOE (d_{NN}) in a NOESY spectra of the Hin peptide in 50 mM Pi, 12%TFE, pH 3. The experiment was performed in H₂O with 1-3-3-1 pulses incorporated to suppress the water signal.

It is, however, an open question as to whether the structure folded under such conditions indeed resembles the three-dimensional structure of the binding domain, either within the free unbound protein species or within the binding complex. We believe it would be reasonable for the two structures to have a resemblance. Most of the concerns that propel the question is that TFE will promote almost any peptide into α -helical secondary structure. What we have observed here, however, is a different phenomenon, i.e., TFE stabilizes the tertiary structures of a peptide, which is rarely observed (Shin et al., 1990). In our experiment, only a small concentration of TFE is used. We have demonstrated through the CD and NMR titration experiments, that the effect of excessive amounts of TFE is to, increase the helical content of the peptide still, but destroy the three-dimensional structure of the peptide at the same time. It must also be pointed out, that it is very rare for any peptide to adopt two distinctively different tertiary structures and both of them are well-defined. Therefore, the structure of the folded peptide under such TFE condition does appear to give us useful insight into the real three-dimensional structure of the binding domain.

The Protein-DNA Binding Assay

Naturally, the most important phase of this research is to investigate the physical interactions between the protein binding domain and the corresponding DNA sequences. Although the specific interactions of the peptide and the DNA sequence encoded in larger DNA molecules were confirmed through various methods (Bruist et al., 1987;

Sluka et al., 1987), there is no evidence that the DNA oligomer containing the binding sequences still acts in the same way. We have developed a binding assay to address the concern. This method can also be used to isolate and purify the binding complexes of the DNA and the peptide.

The assay uses gel filtration chromatography to separate molecules or molecule-complexes according to their effective sizes. The gel filtration chromatography employs mild eluate conditions free from the organic components of reverse phase chromatography and the high ionic strength of ion exchange chromatography. Under such chromatographic conditions, the interactions between the DNA and the peptide remain undisrupted. The assay, therefore, resembles the gel-retardation experiment in principle. Due to limited resolutions of the size-exclusion (gel filtration) chromatography, however, we can not resolve the peaks that correspond to the DNA and the DNA-peptide binding complexes. Yet we can still take advantage of the well-separated peptide peak. Chromatographic methods have an advantage of detecting the spectra of the fractions at each retention time, which not only enable us to monitor both protein and DNA simultaneously at different wavelengths, but also allow us to distinguish a peak that is either DNA or DNA-protein binding complex.

Similar to the gel-retardation experiment, when binding interactions happens between the DNA and the peptide, the peptide will simply migrate with the DNA. If the molar ratio of peptide/DNA is less than one, the corresponding peptide peak will disappear from the chromatograms and the peak corresponding to the DNA will now consist

of DNA plus the peptide. The detail of this binding assay is explained in MATERIALS AND METHODS and in Figure 4-11 and 4-12. We believe the technique of using chromatography to perform the binding assay can be further utilized, e.g., as preparative chromatography to isolate and purify the DNA-peptide binding complexes from nonhomogeneous peptide, an approach that resembles the affinity chromatographic methods.

Using such an assay, we observed binding between the Hin 52mer peptide and the DNA I and DNA II molecules, both of which contain a natural binding site for the Hin recombinase. The chromatograms with DNA I is shown in Figure 4-11. No binding interactions were observed for a control DNA molecule with no specific binding sequences (Figure 4-12). A mutated sequence from DNA I (Figure 3-1) with sequence d(GGTTTTTCGAAAACC)₂, has been assayed and concluded no specific binding activities. The method, therefore, provides a convenient way to check the binding activities of a DNA with a protein binding domain. The dissociation constants for the specific binding are estimated to be smaller than 1 μ M, assuming a 1000 folds of dilution in the process of chromatography (Frankel et al., 1985). Given the high salt eluate we have used (ionic strength 0.5 M), the binding of the peptide and the DNA oligomers under milder conditions should be stronger.

It appears that, although the peptide does not have a stable conformation in solution, the binding activities are nonetheless unhindered. Possible explanations are that the peptide conformation is

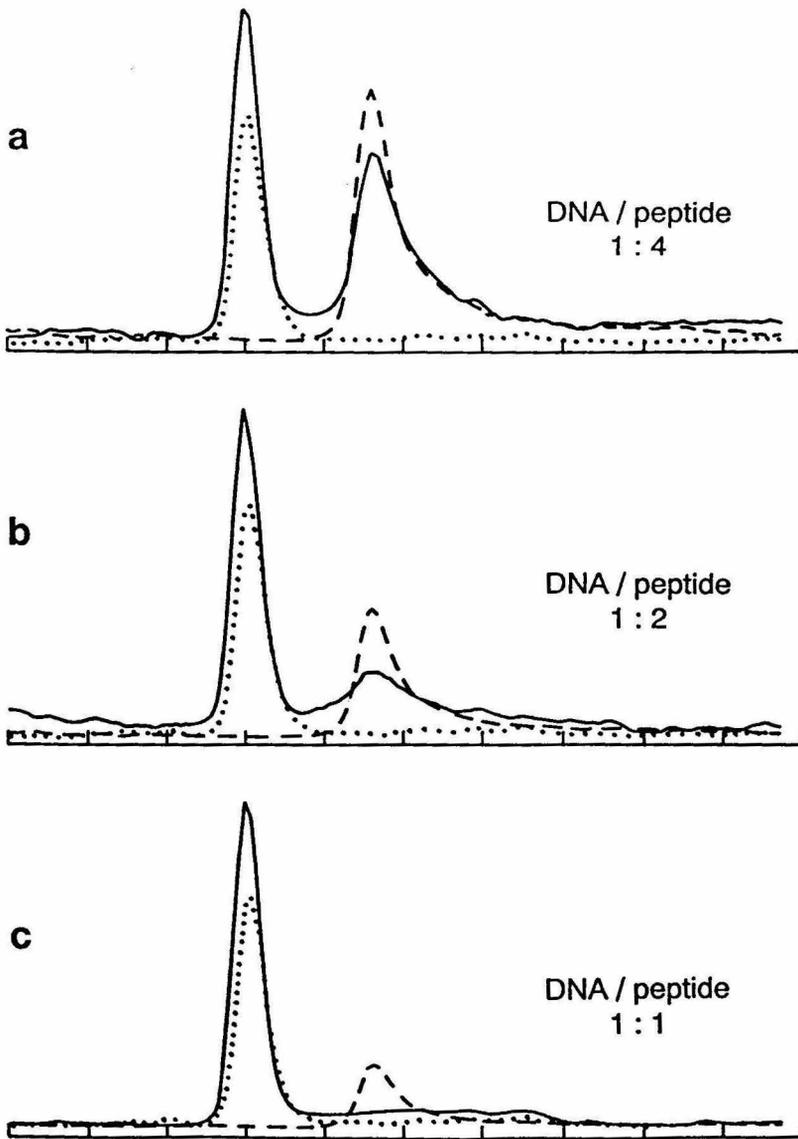


FIGURE 4-11: The chromatograms showing the binding of the Hin peptide to DNA I. *Dotted* trace: DNA; *dashed* trace: peptide; *solid* trace: DNA + peptide mixture. Notice the rise of the peak corresponding to DNA (DNA-peptide complex) and the reduction of peptide peak in the mixture chromatograms. In C, the corresponding peptide peak is missing from the mixture chromatogram.

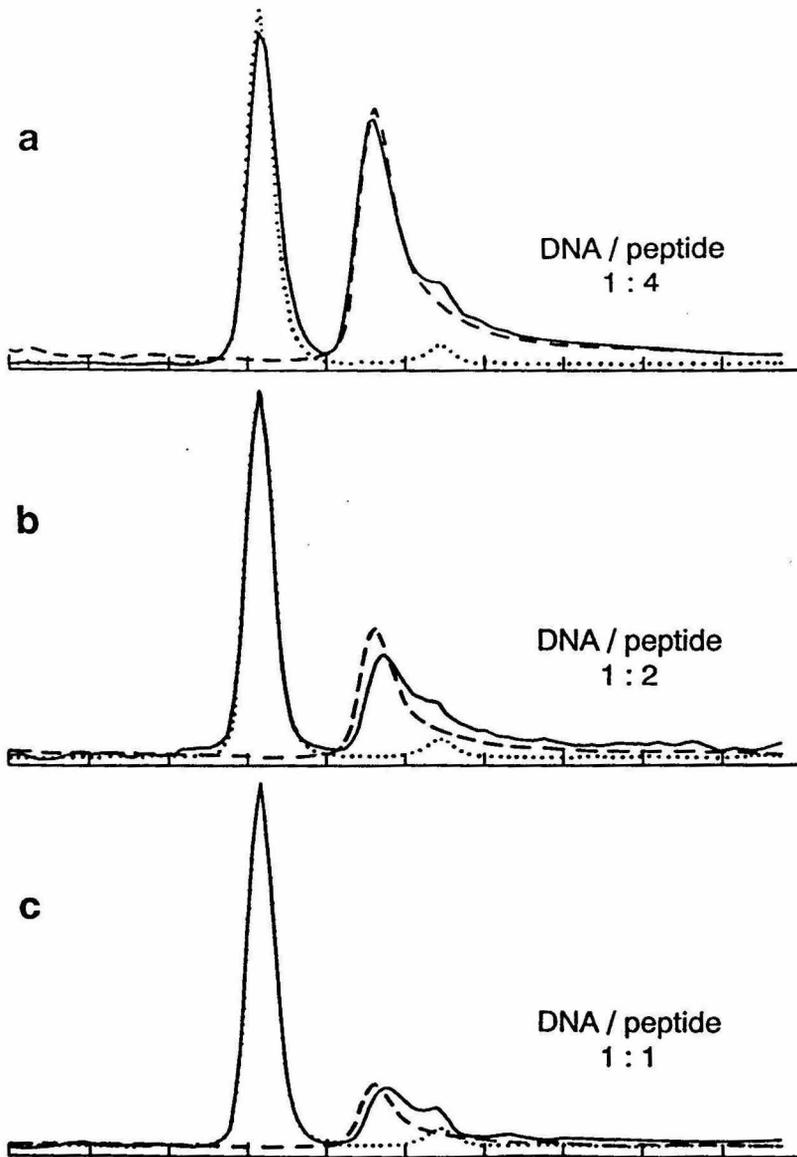


FIGURE 4-12: A control experiment for the DNA binding assay. *Dotted* trace: DNA with non-specific sequence; *dashed* trace: peptide; *solid* trace: DNA + peptide mixture. The mixture chromatograms are almost additions of the other two chromatograms.

induced or greatly stabilized in the presence of DNA binding sites. Such an explanation was confirmed later in the NMR studies (*vide infra*). Interestingly, circular dichroism (CD) examination of the peptide did not suggest a major change in the composition of secondary structures of the peptide upon binding to the DNA (Figure 4-13). It is possible that the secondary structures of the peptide exists in the solution, either transiently or in multiconformers, even if no stable tertiary conformation is formed, yet folding into the latter is highly situation-dependent. The existence of a DNA molecule is certainly a strong promoting factor for such a process.

The DNA-peptide Binding Complexes

As we come to the end of the presentation of this research, greatest interest is naturally focused on the structure of the protein-DNA binding complex. However, the structure determination of protein-DNA binding complexes is not a trivial job. Great efforts have been put into their studies (Lamerichs et al., 1989; Billeter et al., 1990). Up to this point, there is not yet a DNA-protein binding complex solved to atomic details by NMR methods. Our work reported here on the subject is also preliminary in nature.

The binding assay has indicated that the Hin 52mer peptide binds very strongly to, and forms complexes with, DNA sites containing specific Hin recombinase binding sequences. The titration of the peptide with DNA I, one of the DNA binding site (or the reverse) was

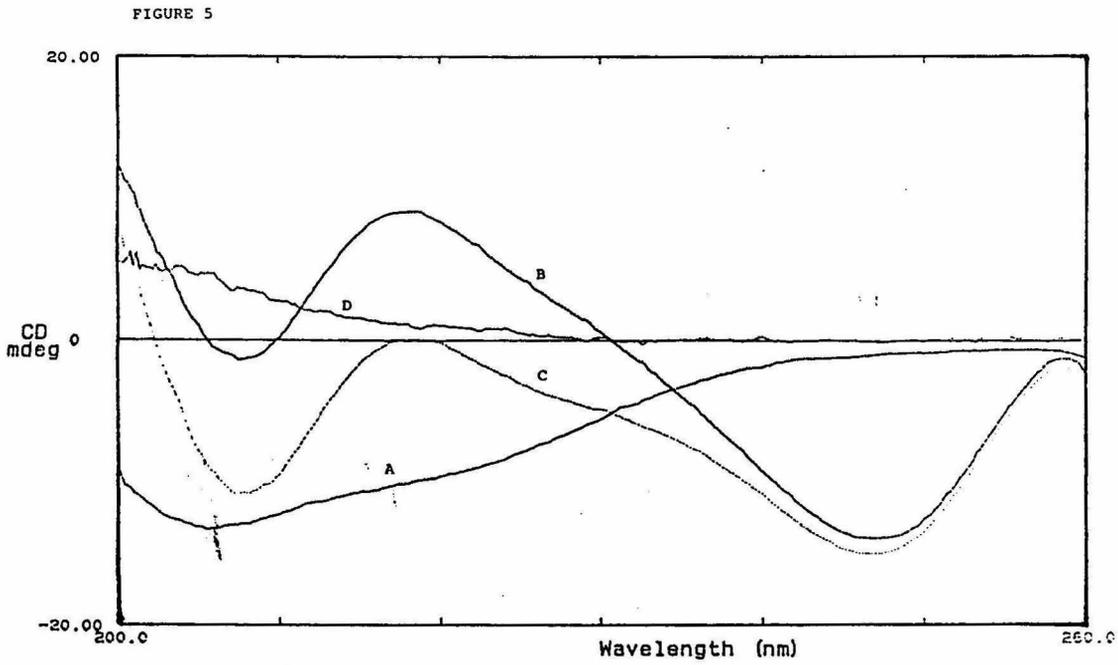


FIGURE 4-13: CD of the Hin peptide (A), the DNA II (B), the peptide and the DNA mixture (C), and the difference spectrum of the three (D). The change in secondary structure of the peptide upon binding is small. Spectra were measured in 20 mM Pi, pH 7.6 aqueous buffer.

accompanied by an instant precipitation of a white solid, which was later confirmed to be 1:1 DNA-protein complex by HPLC and NMR. The binding complex may be redissolved in a buffer with higher ionic strength (0.2 M KCl) and phosphate concentrations (0.1 M potassium phosphate). The 1-D and 2-D NMR studies on the binding complex have been carried out at a concentration of 0.5 mM with the above condition (Figure 4-14, 4-15).

Figure 4-14 shows a 1-D NMR spectrum of the binding complex in comparison to the spectra of the DNA and the peptide alone. The spectrum of the binding complex is significantly different from what would be expected from a simple addition of the two spectra of the free DNA and peptide. The aromatic protons of the Tyr and His have dispersed resonances in the binding complex (8.5-6.0 ppm). The Most important change for the peptide is the appearances of several upfield-shifted methyl groups (0.5 to -0.5 ppm) which are characteristic of a folded and stable protein conformation. For the resonances from DNA, base protons and sugar H1' protons seemed to have changed patterns (8.0-6.5 ppm and 6.0-5.0 ppm) and more significant are the dispersed resonances of thymine methyl groups (1.5-1.0 ppm). These chemical shift changes indicate binding interactions between the DNA and the protein. Since chemical shifts are very sensitive to the conditions and environment of the molecule, the changes in DNA resonances are not necessarily evidences of a distinctive DNA structural change. The peptide, however, experienced a dramatic change from undefined conformation to a unique and stable three-dimensional structure. Thus,

the presence of the DNA molecule seemed to stabilize the peptide conformation greatly. This is markedly different from a similar helix-turn-helix system studied by NMR. In the *lac* repressor headpiece, the DNA binding domain are reported to have similar structures in the complex and in free forms (Boelens et al., 1987; Lamerichs et al., 1989).

The NOESY of the binding complex with a mixing time $\tau_m = 150$ ms is shown in Figure 4-15. Region *g* contains DNA sequential base proton NOE interactions, and the strong peptide intraresidue NOE between aromatic protons of Tyr and Phe (Figure 4-16). Region *d* and *f* are mainly NOE connectivities in DNA. Region *a* and *b* are mainly long range NOE interactions in the folded peptide. Some of the NOE can be tentatively assigned, especially in region *a*, where the NOE interaction, i.e., space proximities between aromatic protons and aliphatic protons, are directly correlated to the upfield shiftings of these aliphatic methyl groups due to aromatic ring currents (Figure 4-17). The tentative assignments take advantage of the model structure of the protein-DNA binding complex (Plaxco et al., 1989). Thus, in the model structure, Ile 30 and Phe 31, Leu 38 and Phe 42 have been found to have the aliphatic methyl groups sitting on the aromatic rings. The observations of these long-range NOE interactions are the conclusive evidence of a folded and stable three-dimensional structure. The specific NOE interactions between the DNA and the peptide can not be identified at this stage, but many NOE intensities in region *c* and *e* appeared to be good candidates.

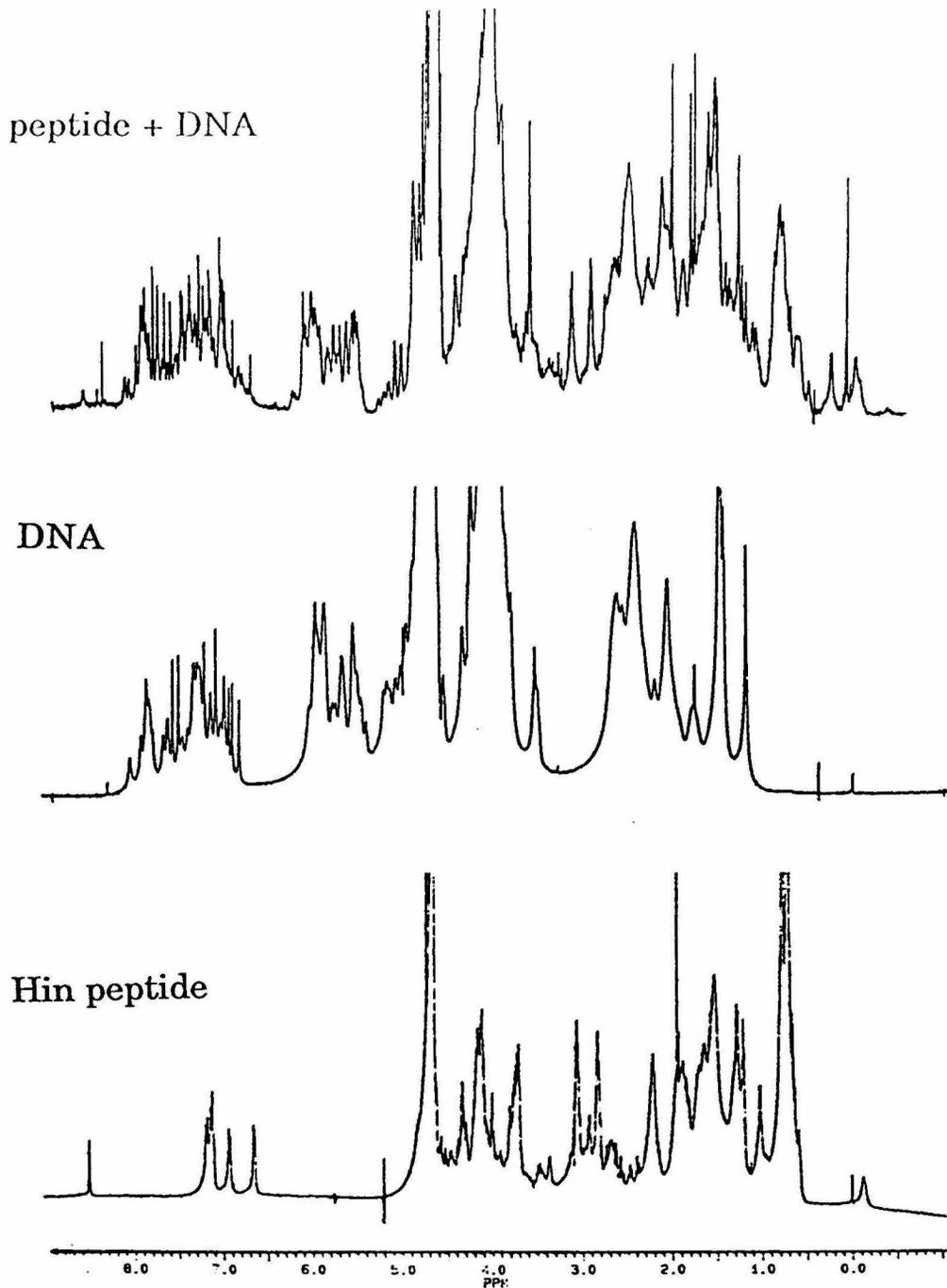


FIGURE 4-14: NMR spectra of the Hin peptide in unfolded form, the DNA I, and the peptide-DNA binding complex in 0.2 M KCl, 0.1 M potassium phosphate and pH 7.0, D₂O. See text for more explanations.

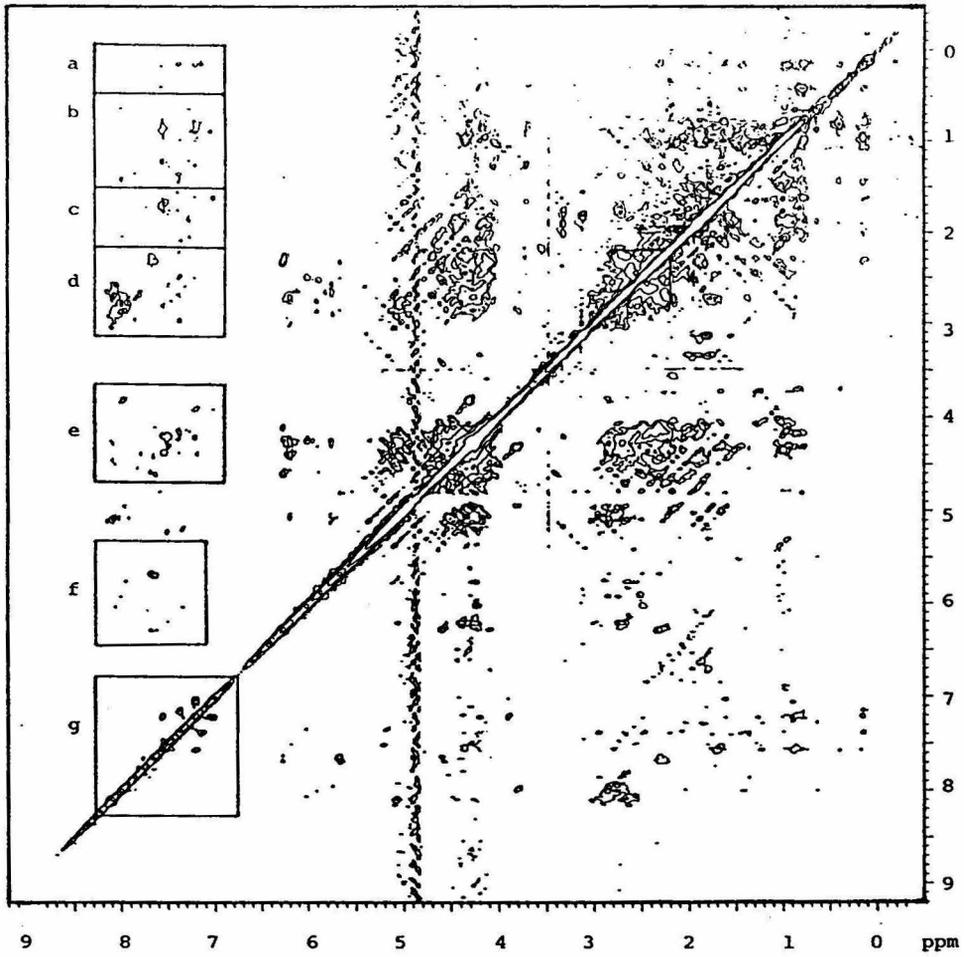


FIGURE 4-15: A NOESY ($\tau_m = 250$ ms) of the Hin peptide-DNA binding complex in 0.2 M KCl, 0.1 M potassium phosphate, pH 7.0 and D₂O.

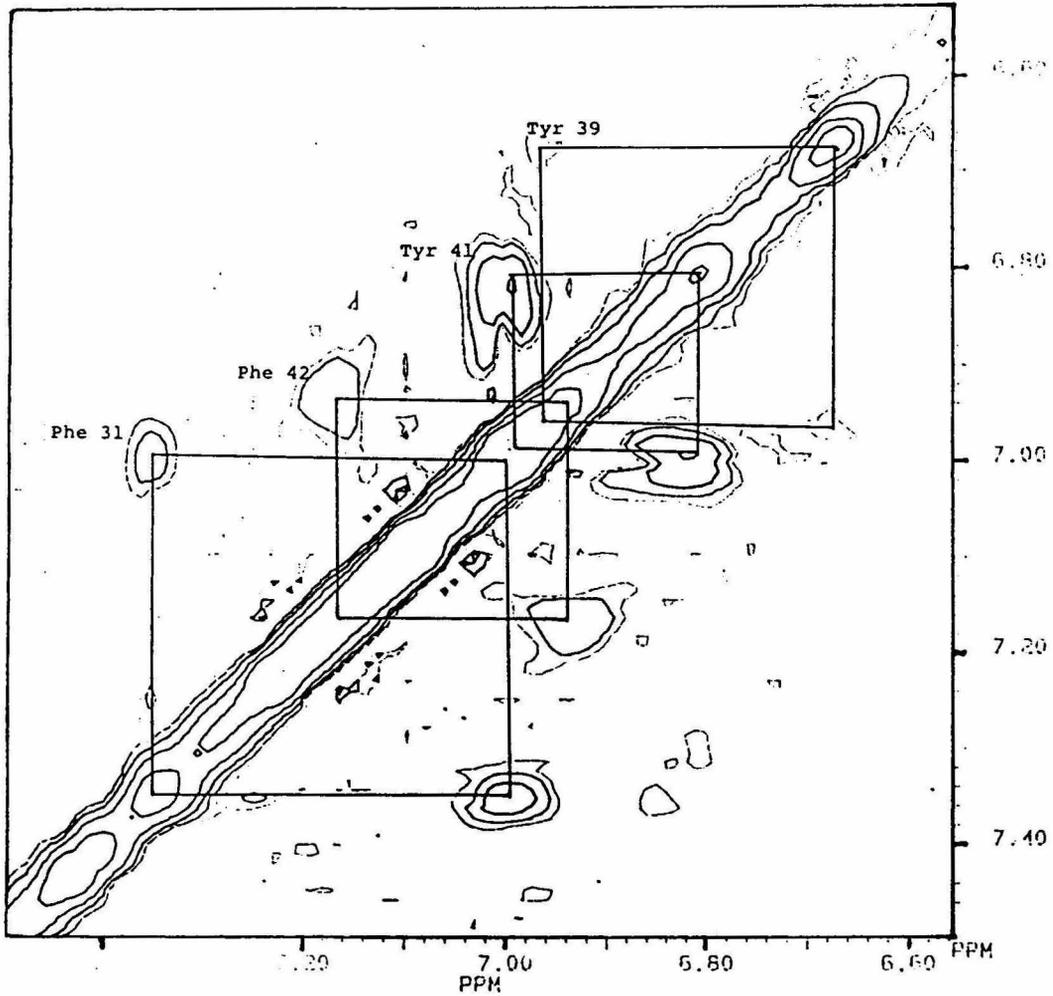


FIGURE 4-16: The aromatic region of Tyr and Phe in the NOESY of the Hin peptide-DNA binding complex as in Figure 4-14, region *g*.

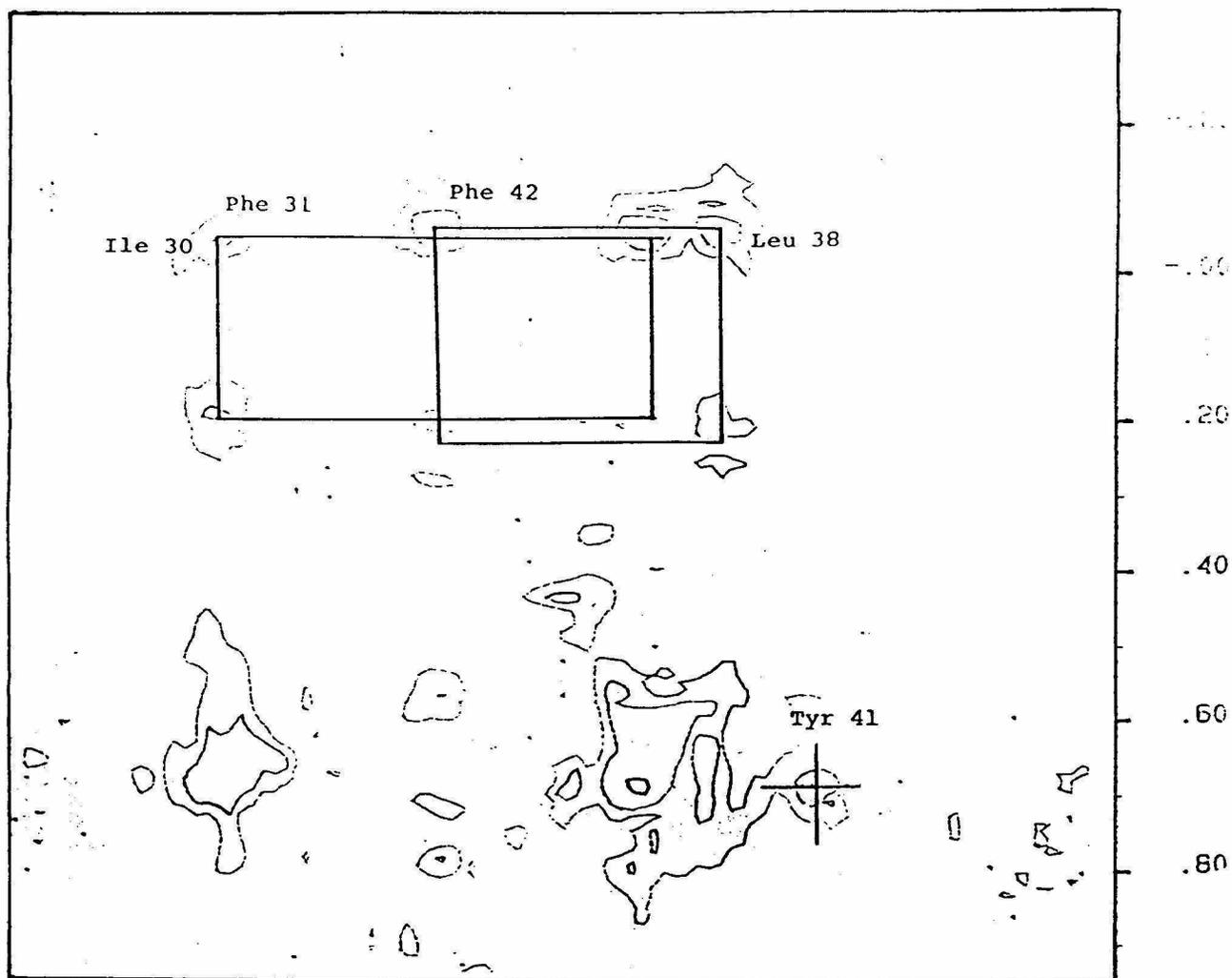


FIGURE 4-17: The cross-peak region of aromatic protons (Tyr, Phe) - aliphatic methyl groups (Leu, Ile) in the NOESY of the Hin peptide-DNA binding complex as in Figure 4-14, region *a*.

In conclusion, the Hin 52mer peptide can fold into a stable and unique conformation and form a binding complex in the presence of DNA oligomers with specific binding sequences. The binding complex appears to be structurally well-defined and solvable to a good resolution. Any further research attempting to solve the three-dimension structure of the binding complex, however, would need to address the following two difficulties: 1) overcrowded spectra, due to the large size of the protein-DNA binding complex; 2) poor spectrum quality, mainly due to low signal/noise from relatively low solubility of the binding complex. Possible solutions to the first problem are to expand the spectra into one or more new dimensions by combined usage of elaborated pulse sequences and isotope (e.g. ^{15}N) labelling (Kay et al., 1990; Fairbrother et al., 1991); or by selective deuterations of all or part of riboses or amino acids (Tsang et al., 1990). A future suggestion for dealing with the second difficulty is to engineer mutants of the synthetic peptide that have reduced aggregation possibilities and enhanced solubilities without disrupting the three-dimensional structures, e.g., to substitute certain surface residues with more hydrophilic ones. This approach is also applicable to the structural determination of the Hin 52mer peptide alone, i.e., modified sequences may reduce intermolecular interactions and, therefore, may help to stabilize the peptide conformation in solution.

ACKNOWLEDGMENTS

We thank James P. Sluka, David P. Mack, Peter B. Dervan and Suzanna J. Horvath for providing some of the Hin peptide and assisting the chemical synthesis of the peptide used in this research.

REFERENCES

- Acharya, A. S., Khandke, K. M., Manjula, B. N. & Roy, R. P. (1990) *Abstract of the Protein Society 4th Symp.* M6.
- Billeter, M., Qian, Y. -Q., Otting, G., Müller, M., Gehring, W. J., Wüthrich, K. (1990) *J. Mol. Biol.* **214**, 183.
- Boelens, R., Scheek, R. M., Lamerichs, R. M. J. N., de Vlieg, J., van Boom, J. H., & Kaptein, R. (1987) in *DNA-ligand Interactions* pp.191-215.
- Brems, D. N. et al. (1987) *Biochemistry* **26**, 7774.
- Bruist, M. F., Horvath, S. J., Hood, L. E., Steitz, T. A., & Simon, M. I. (1987) *Science* **235**, 777.
- Fairbrother, N. J., Cavanagh, J., Dyson, H. J., Palmar III, A. G., Sutrina, S. C., Reizer, J., Saier, M. H., & Wright, P. E. (1991) *Biochemistry* **30**, 6896.
- Frankel, A. D., Ackers, G. K., & Smith, H. O., (1985) *Biochemistry* **24** 3049.
- Glasgow, A. C., Bruist, M. F., & Simon, M. I. (1989) *J. Biol. Chem.* **264**, 10072.
- Hope, P. J. (1983) *J. Magn. Reson.* **55**, 283.
- Hughes, K. T., Youderian, P., & Simon, M. I. (1988) *Genes Dev.* **2**, 937.
- Kay, L. E., Clore, G. M., Bax, A., & Gronenborn, A. M (1990) *Science* **249**, 411.

- Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B., & Rüterjans, H. (1989) *Biochemistry* **28**, 2985.
- Lehrman, S. R. et al., (1990) *Biochemistry* **29**, 5590
- Mammi, S. et al., (1988) *Biochemistry* **27**, 1374.
- Nelson, J. W. et al. (1989) *Biochemistry* **28**, 5256.
- Plaxco, K. W., Mathiowetz, A. M., & Goddard III, W. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 984.
- Shin, H. -C., Waltho, J. P., Dyson, H. J., & Wright, P. E. (1990) *Abstract of the Protein Society 4th Symp.* M5.
- Sluka, J. P., Horvath, S. J., Bruist, M. F., Simon, M. I., & Dervan, P. B. (1987) *Science* **238**, 1129.
- Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I., & Dervan, P. B. (1990) *Biochemistry* **29**, 6551.
- Tsang, P., Wright, P. E., & Range, M. (1990) *J. Am. Chem. Soc.* **112**, 8183.
- Wade-Jardetzky, N. et al. (1978) *J. Mol. Biol.* **128**, 1259.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, pp.30-31.

APPENDIX: PROGRAMS AND PARAMETER FILES
FOR DATA PROCESSING

	file size (blks)
ASG_PRIORITY.DAT;6	1
ATHYD.DATA;1	7#
ATOM_NAME_DIFF.DAT;10	2
BRUKER.EXE;2	16*
CHEM.CON;1	1#
CNSTRNT_GEN.COM;10	1
CNSTRNT_GEN.EXE;12	21
COMPAREXPK.EXE;17	15
COMPARE_CNSTRNT.EXE;9	14
CONSTRAIN.DAT;2	11#
CONVERT_BKV_PDB.EXE;2	11
CONVERT_CNSTRNT_CONSTR.EXE;7	22
CORMA_IN.EXE;2	12#
CROSSPEAK.COM;205	141
DCNVL_FILEGEN.EXE;3	8
DECONVLXPK.EXE;79	28
EXAM_CNSTRNT.EXE;13	22
EXTRACT_E.EXE;7	9
FTGP22.EXE;1	415*
GAUSS_CURV.DAT;3	13
GENPLOT_CNSTR_COMP.COM;14	8
GENPLOT_MDE.COM;10	4
GENPLOT_RMSD0.COM;2	5
GENPLOT_RMSD1.COM;6	7
MARDI.EXE;11	126#
MERGE_FILEGEN.EXE;5	13

MERGE_XPKNM\$XPK.EXE;23	9
MERGE_XPKNM\$XPKDC.EXE;6	9
NEWHYD.EXE;4	45#
OFFSETXPK.EXE;5	7
PSEUDO_APPROX.DAT;2	3
SORT_ASG.EXE;11	11
SORT_CNSTRNT.EXE;27	32
SUBTRT1_1024.MAC;2	1
SUBTRT1_512.MAC;1	1
XFBQSB45.MAC;2	2
XFBQSS45.MAC;2	2
XPKASG.EXE;40	44
XPKDC_PRINT.COM;3	3
XPKNM_PRINT.COM;11	3
XPKPICK_FILEGEN.EXE;3	9
XPK_GEN.EXE;36	26
XPK_GEN_FILEGEN.EXE;9	11

* FTNMR associated softwares (Hare Research)

MARDIGRAS associated softwares (T.L.James, UCSF)

All programs can be accessed through the master command procedure CROSSPEAK.COM. Program usages are either self-explanatory or instructions are given in the command procedure.