U

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION (U)	1b. RESTRICTIVE MARKINGS NA
2a. SECURITY CLASSIFICATION AUTHORITY NA	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE NA	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NA	5. MONITORING ORGANIZATION REPORT NUMBER(S) NA
6a. NAME OF PERFORMING ORGANIZATION California Institute of Technology; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Division of Chemistry 147-75 California Institute of Technology Pasadena, California 91125	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION DARPA/DSO; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0775
8c. ADDRESS (City, State, and ZIP Code) 1400 Wilson Blvd. Arlington, Virginia 22209	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Biopolymers: Proteins and Nucleic Acids

12. PERSONAL AUTHOR(S)
John H. Richards; John N. Abelson; Peter B. Dervan; Leroy H. Hood; Melvin I. Simon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 9/15/89 TO 9/14/90	September 15, 1990	15

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Biopolymers/Macromolecular Interactions/Protein-DNA Binding/Peptide Synthesis/DNA Cleavage/NMR of Proteins and Nucleic Acids/tRNA Specificity

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Our work focuses on molecular recognition by biopolymers as a model for interactions between macromolecules in general. Specifically we are learning the rules that govern protein-nucleic acid (both DNA and RNA) binding. Chemical synthesis, molecular genetics, biophysical observations, including nmr and CD, and computer simulations serve as general experimental approaches used in this work.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	(U)
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Ira Skurnick & Dr. Robert W. Newburgh	22b TELEPHONE (Include Area Code) 202/694-3147 & 696-1212; 22c. OFFICE SYMBOL DARPA/DSO & ONR

ANNUAL REPORT

September 15, 1990

Biopolymers: Proteins and Nucleic Acids

California Institute of Technology
Pasadena, California 91125

Co-Investigators: John N. Abelson, Leroy E. Hood and Melvin I. Simon, Division of Biology; Peter B. Dervan and John H. Richards, Division of Chemistry

Contract: N00014-86-K-0755

Summary

We have continued to gain insights into the central question of interactions between biopolymers that are essential to information processing and control in biological systems. The work largely focuses on interactions involving nucleic acids and proteins that have profound influences on the expression of genetic information and on the behavior of all living organisms. Insights from studies of biological systems have been translated into synthetic chemical approaches to the creation of powerful and specific reagents for the *in vitro* manipulation of genetic material.

In addition, progress continues to be made in elucidation of the detailed molecular structure of the important DNA protein complexes, largely by nmr, though x-ray crystallography is beginning to hold considerable promise.

Another crucial aspect of information processing concerns the charging of a particular tRNA by its cognate amino acid, the fundamental control step in translating a nucleic acid sequence into an amino acid sequence. The basic aspects of this essential decoding process have been another focus of our work which has been successfully pursued in the past year.

JOHN N. ABELSON

In previous years we have been studying the problem of tRNA identity by changing the identity of one tRNA to another. By changing an amber suppressor $tRNA^{Leu}$ to $tRNA^{Ser}$ we were able to determine much about those features of a $tRNA^{Ser}$ molecule which allow its unique recognition by the serine tRNA synthetase. We are continuing this same approach with $tRNAs^{Leu}$ and $tRNA^{Cys}$. In preliminary results it appears that we have been able to change $tRNA^{Ser}$ to $tRNA^{Leu}$ and $tRNA^{Ala}$ to $tRNA^{Cys}$.

We have also taken a completely new approach. Drs. Jeffrey Sampson and Margaret Saks have made a review of existing data and ideas that have led to the deduction of putative identity sites for all 20 *E. coli* synthetases. The basic premises of the approach developed for deducing recognition sites are that (1) synthetase topology dictates regions of interaction with tRNAs and (2) when taken individually, identity elements need not be unique to only one isoaccepting group. Because bases and base pairs do not have to be unique to be recognition elements, synthetases may interact with groups of adjacent nucleotides and the chemical environments that they form. To test this idea, we analyzed acceptor stem chemistry as a series of trimers.

These analyses revealed that at least one set of three adjacent nucleotides formed a unique trimer in 13 of the 20 isoaccepting groups.

In vitro aminoacylation assays are being used to test the deduced sites and the assumptions made about the nature of recognition elements. The experiments focus on tRNA acceptor stems and the contribution of their base sequences to recognition. These experiments use two different types of substrates. Mini helices (such as those used in Paul Schimmel's laboratory) are being used for studies of recognition of the $tRNA^{Ser}$ acceptor stems by seryl synthetase. We have developed model full length tRNAs for detailed studies of acceptor stem recognition for synthetases that require information in both the acceptor stem and anticodon. These model substrates are composed of a wild type anticodon, a standard body and interchangeable acceptor stems.

PETER B. DERVAN

Protein-DNA Recognition

The major accomplishments this past year (1990) are:

1. We completed our characterization of the structural motif of the DNA binding domain of the transcriptional activator GCN4 (222-281) using the affinity cleaving method. In our model, the positive dipole of the N-termini of the α-helical leucine zipper domains interact with the phosphate backbone of the DNA as has been suggested previously. The monomers then separate into two arms which follow the major groove of the DNA in opposite directions, placing the N-termini of the DNA binding domain in successive major grooves on the opposite face of the DNA from the zipper region. (*Science* **248**, 847 (1990)).

2. We are constructing hybrid metalloproteins, consisting wholly of naturally occurring α-amino acids, that oxidatively cleave DNA at a single base within 13 base pair recognition sites. (*J. Am. Chem. Soc.* 112, 4604 (1990)). We have found that the tripeptide GGH is a metal-specific structural domain consisting of naturally occurring amino acids that could be incorporated at the NH_2-terminus of a variety of recombinant proteins (such as other DNA binding proteins, receptors, or antibodies) with the function of precise, efficient substrate-directed oxidation, activated in the presence of Ni(II) and peracid (25°C, pH 7.5).

3. We continue our studies of binding constants of DNA binding proteins at discrete sites on restriction fragments using *quantitative affinity cleaving*.

4. In unpublished work, we have developed reagents for attachment of EDTA to discrete internal amino acid positions in synthetic proteins. These will be useful for mapping the structure of protein-DNA complexes by affinity cleaving. (D. Mack, P. B. Dervan)

5. We have initiated studies of protein-EDTA autocleavage as a method to determine protein structure (protein affinity cleavage) in protein-DNA complexes.

DARPA Supported Research

1987 Synthesis of a Sequence Specific DNA Cleaving Peptide. J. Sluka, M. Bruist, S. J. Horvath, M. I. Simon, and P. B. Dervan, *Science* **238**, 1129.

A-1

1988 Design and Chemical Synthesis of a Sequence Specific DNA-Cleaving Protein. D. P. Mack, B. L. Iverson and P. B. Dervan. *J. Am. Chem. Soc.* **110**, 7572.

1990 Nickel-Mediated Sequence Specific Oxidative Cleavage of DNA by a Designed Metalloprotein. D. P. Mack and P. B. Dervan, *J. Am. Chem. Soc.* **112**, 4604.

1990 Structural Motif of the GCN4 DNA Binding Domain Characterized by Affinity Cleaving. M. G. Oakley and P. B. Dervan, *Science* **248**, 847.

1990 The Importance of Minor Groove Contacts for the Recognition of DNA by the DNA Binding Domain of Hin Recombinase. J. P. Sluka, S. J. Horvath, A. C. Glasgow, M. I. Simon and P. B. Dervan, *Biochemistry* **29**, 6551.

1990 Orientation of the Putative Recognition Helix in the DNA Binding Domain of Hin Recombinase Complexed with the Hix Site. D. P. Mack, J. P. Sluka, J. A. Shin, J. H. Griffin, M. I. Simon and P. B. Dervan, *Biochemistry* **29**, 6561.

1990 Reagents and Methods for the Solid Phase Synthesis of Protein-EDTA for Use in Affinity Cleaving. J. P. Sluka, J. H. Griffin, D. P. Mack and P. B. Dervan, *J. Am. Chem. Soc.*, **112**, 6369.

1990 Structural Motif of the DNA Binding Domain of $\gamma\sigma$ Resolvase Characterized by the Affinity Cleaving Method. K. S. Graham and P. B. Dervan, *J. Biol. Chem.* **265**, 16534.

1990 Characterization of Protein:DNA Complexes by Affinity Cleaving. P. B. Dervan, in *Methods in Enzymology*, in press.

LEROY E. HOOD

The "zinc finger" is a repeating protein domain, a structural motif that is responsible for the sequence-specific DNA-binding activities of a variety of eukaryotic transcription factors. The zinc finger model (McLachlan A. D. & Klug, A. (1985) *EMBO J.* 4, 1609-1614; Brown, R. S., Sander, C. & Argos, P. (1985) *FEBS Lett.* **186**, 271-274) based on analysis of *Xenopus* transcription factor TFIIIA predicted that conserved Cys and His residues act as zinc metal ion ligands: conserved aromatic and aliphatic side chains were predicted to interact in a hydrophobic core, stabilizing the domain.

The yeast transcription factor ADR1 contains two adjacent finger domains. In our first approach to test the structural and folding prediction of the model, peptides encompassing the double finger (ADR1c) and each of the single fingers (ADR1a and ADR1b) had been chemically synthesized. The folding and metal binding characteristics of these synthetic peptides had been assessed, as well as the high resolution three-dimensional solution structure of both single finger domains had separately been determined employing 2D NMR spectroscopy (COSY, RELAY, TOCSY, NOESY) and the distance geometry program DSPACE (Hare Research). An experimentally determined model of each of the two single fingers is proposed that is consistent with circular dichroism, one and two-dimensional nuclear magnetic

resonance and visual spectroscopy of the single-finger peptide reconstituted in the presence of zinc. Similar work with the double-finger peptide is in progress.

Our study revealed that in the case of wild-type zinc finger peptides, zinc binding is coincident with the formation of a compact, globular domain containing an amphiphilic α-helix and an extended β-like region. Whereas the two different wild-type zinc fingers folded into one conformational species upon tetrahedral binding of zinc, a synthetic deletion peptide, deleted from a single amino acid residue (del.138-ASN) did not bind zinc with tetrahedral geometry, nor did it fold into a domain. This result was also found when both wild-type peptides underwent thiol alkylation or imidazole participation, and spacing between pairs of ligands (which is conserved) are absolute requirements for tetrahedral binding of metal and subsequent folding into a domain. This provides direct, experimental support for the initial prediction of the model.

Several point-mutant single-finger peptides of ADR1a and ADR1b have also been chemically synthesized. These mutants contain alteration of conserved aromatic and aliphatic residues, which should allow us to determine the structural and functional significance of the hydrophobic core in the folded zinc-finger structure.

Using genetic method several separate alanine point mutations have been introduced in the finger corresponding to ADR1b in the protein and the mutant proteins' ability to bind DNA have been screened. For the first round of mutagenesis experiments we concentrated on the α-helical region of the finger. Then we chemically synthesized 3 selected alanine point mutants of the ADR1b-finger: in the genetic experiment (a) showed decreased DNA-binding, (b) showed increased DNA-binding, (c) showed DNA-binding that of the wild type. The structural features of these and other alanine point mutants, hopefully, will lead us to an understanding of how the zinc fingers are interacting with DNA.

Publications

Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. and Klevit, R. E. (1988) Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* 214, 1489-1492.

Parraga, G., Horvath, S., Hood, L., Young, E. T. and Klevit, R. E. (1990) Spectroscopic studies of wild-type and mutant "zinc finger" peptides: Determinants of domain folding and structure. *Proc. Natl. Acad. Sci. USA* 87, 137-141.

Klevit, R. E., Herriott, J. R. and Horvath, S. J. (1990) Solution structure of a zinc finger domain of yeast ADR1. *Proteins: Structure, Function and Genetics* 7, 215-226.

JOHN H. RICHARDS

Summary

We have made important progress in our recent efforts to characterize the three-dimensional structure of the 52 residue DNA binding domain of Hin recombinase. The peptide can be successfully folded in relatively acidic buffer with a small amount of trifluoroethanol (TFE). The unique folding conformation is now being investigated by NMR. We have also used circular dichroism (CD) spectroscopy to study the secondary and tertiary structure of the peptide in various situations.

The Peptide Folding Studied by NMR

Previous NMR studies in this laboratory have shown that the peptide in the absence of DNA adopts an unstable conformation sensitive to solution conditions. Detailed NMR structure analysis of the peptide could not be done under physiological conditions (20 mM phosphate and NaCl, pH 7.6) because the peptide oligomerizes at the relatively high concentration necessary for NMR studies. The oligomerization can be prevented by adjusting the NMR sample to acidic conditions (pH 3.4) in the appropriate buffers, yet the peptide still did not adopt a unique conformation under a variety of conditions. It was found that a small amount of TFE (12%(v) or 3.4 mol%) greatly stabilizes the peptide conformation under the acidic conditions (Fig. 1). The NMR of the peptide exhibits a single set of resonances with dispersed chemical shifts, an indication of folded and unique peptide conformation. In further NMR studies, we observed continuous stretches of peptide amide proton-amide proton (NH-NH) connectivities (d_{NN}) in the NOESY of the 52 mer peptide in water. 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. We are now in the process of sequence-specific resonance assignments of the 52 mer NMR spectrum.

The Peptide Folding Studied by CD

Under physiological conditions, the helical content of the peptide increases as peptide concentration increases. This phenomenon has been observed for small peptides, and was interpreted as a result of unstable peptide conformation and peptide intermolecular interactions (1). Under acidic conditions, the peptide helical content is independent of its concentration in the range of 0.0001 to 0.2 mM, indicating the absence of intermolecular interactions (Fig. 2) at these concentrations at pH 3.4.

A TFE titration (up to 20%) of the peptide solution under the acidic conditions observed by CD measurements showed that the helical content increases as TFE is added (Fig. 3), a result which is consistent with earlier reports (2). Interestingly the CD spectra exhibits a signal at 275 nm which peaks at approximately 12% TFE (Fig. 4). We believe this signal is caused by a lock-out of the tyrosine rings in the peptide in the process of folding. Therefore the CD result coincides with the NMR observation that the peptide folding occurs at an optimum TFE concentration of 12% in acidic conditions.

TFE has been known to promote α-helical secondary structures for small peptides, but there is so far no report of its stabilizing tertiary structures of a peptide.

Studies of DNA-Protein Binding

In the presence of DNA with the specific base sequences characteristic of the Hin recombinase binding sites, the peptide conformation is greatly stabilized, as indicated by NMR. Interestingly, circular dIchroism (CD) examination of the peptide does not suggest a major change in the composition of secondary structures of the peptide upon binding to the DNA (under physiological conditions, Fig. 5). The α-helical contents are about 25% in both situations. About the same values are obtained for the peptide in the presence of 12% TFE under acidic conditions.

We propose that the secondary structural aspects of the peptide are very similar in both cases; folding into a stable tertiary structure depends more critically on the

solution conditions, and very importantly, binding to DNA with the specific base sequence of the Hix site greatly stabilizes a native tertiary structure.

References

1. Brems, D. N. *et al.* (1987) *Biochemistry* **26**, 7774-7778.
2. Lehrman, S. R. *et al.*, (1990) *Biochemistry* **29**, 5590-5596.
   Nelson, J. S. *et al.*, (1989) *Biochemistry* **28**, 5256-5261.
   Mammi, S. *et al.*, (1988) *Biochemistry* **27**, 1374-1379.

MELVIN I. SIMON

We made a great deal of progress in the past year primarily focusing our attention on the specific interaction between the DNA binding peptide and its site. It has become clear to us from a number of experiments that in order to completely redesign this peptide in a useful fashion we will have to have a precise idea about its three-dimensional interaction with DNA. In order to do this we have begun efforts to get a crystal structure of the protein DNA complex starting with the 52 amino acid peptide that we have synthesized. As part of this project, Dr. Richards has shown that the peptide by itself in solution has very little native structure, however, in the presence of the appropriate DNA sequence the peptide is able to assume NMR characteristics that suggest a distinct three-dimensional structure. Thus, it appears that the correct DNA sequence allows the protein to fold up, to bind appropriately and to assume a defined structure. This explains why initial attempts to crystallize the 52 amino acid peptide by itself were unsuccessful. However, we have now collaborated with Dr. Dickerson and Dr. Reed Johnson at UCLA. We've prepared very pure 52 amino acid peptide and different versions of the DNA half site binding sequence with one and two base pair extensions on either end. In initial experiments when these were set up to crystallize we were able to obtain small crystals that are still too small to diffract. However, they appear to contain protein and we are currently working on conditions to extend these studies and to get crystals large enough to obtain data regarding the three-dimensional structure of the protein DNA complex. We are also proceeding to work on the three-dimensional structure of the entire enzyme bound to DNA. Our collaborator, Dr. Doug Rees in the Chemistry Department here at Caltech, has just hired a postdoctoral fellow who will be working together with Dr. Han Lin in our laboratory to purify large amounts of Hin recombinase for crystalization with the DNA.

Characterization of Binding and Recombination

In order to fully characterize both *in vivo* and *in vitro* the precise requirements for DNA protein binding in this system a set of 40 recombination sites which vary by symmetric pairs of single base substitutions from the HixC consensus recombination site were constructed. This set was used in the *Salmonella* P22 base challenge phage selection to define the DNA sequence determinants for the binding of Hin to DNA *in vivo*. Base pair substitutions at 4 symmetric positions in the HixC site were severely defective in the ability of Hin to bind to these sequences *in vivo* (see the enclosed table). These include two positions in the major groove recognition portion of the HixC sequence and two positions in the minor groove recognition portion. The base substitutions in the major groove recognition which were defective in Hin binding still retained residual binding capability while the base pair substitution affecting the minor groove recognition lost all binding capability. *In vivo* binding assays were carried out with the HixC sites containing three deazaadenine residues substituted for adenine residues at positions 5 and 6 of the HixC site. In one construct the

apparent binding constant was reduced by 2½-fold when three deazaadenine residues were incorporated compared to the identical substrate containing adenine residues. These results suggest that hydrogen bond contacts to bases in the minor groove of the Hix DNA is not a significant contributor of binding specificity; rather Hin probably recognizes the shape of the minor groove and interacts through van der Waals and phosphate contacts. We will be continuing to do studies on the construction of DNA binding sites with modified bases to define more clearly the contribution that each base and its ability to form hydrogen bonds with the protein plays in the direction of the protein with DNA.

We are continuing our work on the modification of the peptide and the appropriate site to engineer new activities into this particular polypeptide. Dave Mack and Peter Dervan, as part of our group collaboration in the DARPA project, showed that a tripeptide, Gly Gly His, when added to the N-terminus of the Hin 52 amino acid binding peptide, was able to bind nickel very tightly. When the peptide bound to its appropriate DNA site the nickel can be activated by the addition of an oxidizing agent and the nickel then causes a break in the DNA minor groove. Cleavage by nickel is very efficient and highly localized in the minor groove binding region. We have synthesized the gene that has the methionine codon followed by Gly Gly His and the 52 mer sequence. We will try to produce this in *E. coli*. Under these circumstances, the methionine residue may be removed. As there is independent evidence that when a glycine follows methionine the methionine residue is efficiently removed from the polypeptide by endogenous enzymes in *E. coli*. We will overproduce this 55 amino acid fragment and assay for its ability to bind nickel, to bind to Hix DNA and to cleave Hix DNA.

Publications During this Period

Mack, D. *et al.* (1990) *Biochemistry* **29**, 6564.

Mack D. *et al.* (1990) *Biochemistry* **29**, 655.

Hughes, K. T., Lin, A. H., and Simon, M. I. Genetic studies of the requirement for Hin recombinase binding. Studies on symmetric binding sites. Manuscript in preparation.

Glascow, A., Johnson, R. and Simon, M. I. DNA bending associated with the Hin recombination enhancer function. Manuscript in preparation.

Figure 1. NMR spectra of random coiled and folded Hin peptide (52 mer), 1 mM in D_2O, 25°C. **A.** Salt free, pH 3. **B.** 50 mM phosphate, pH 3, 12%(v) TFE.

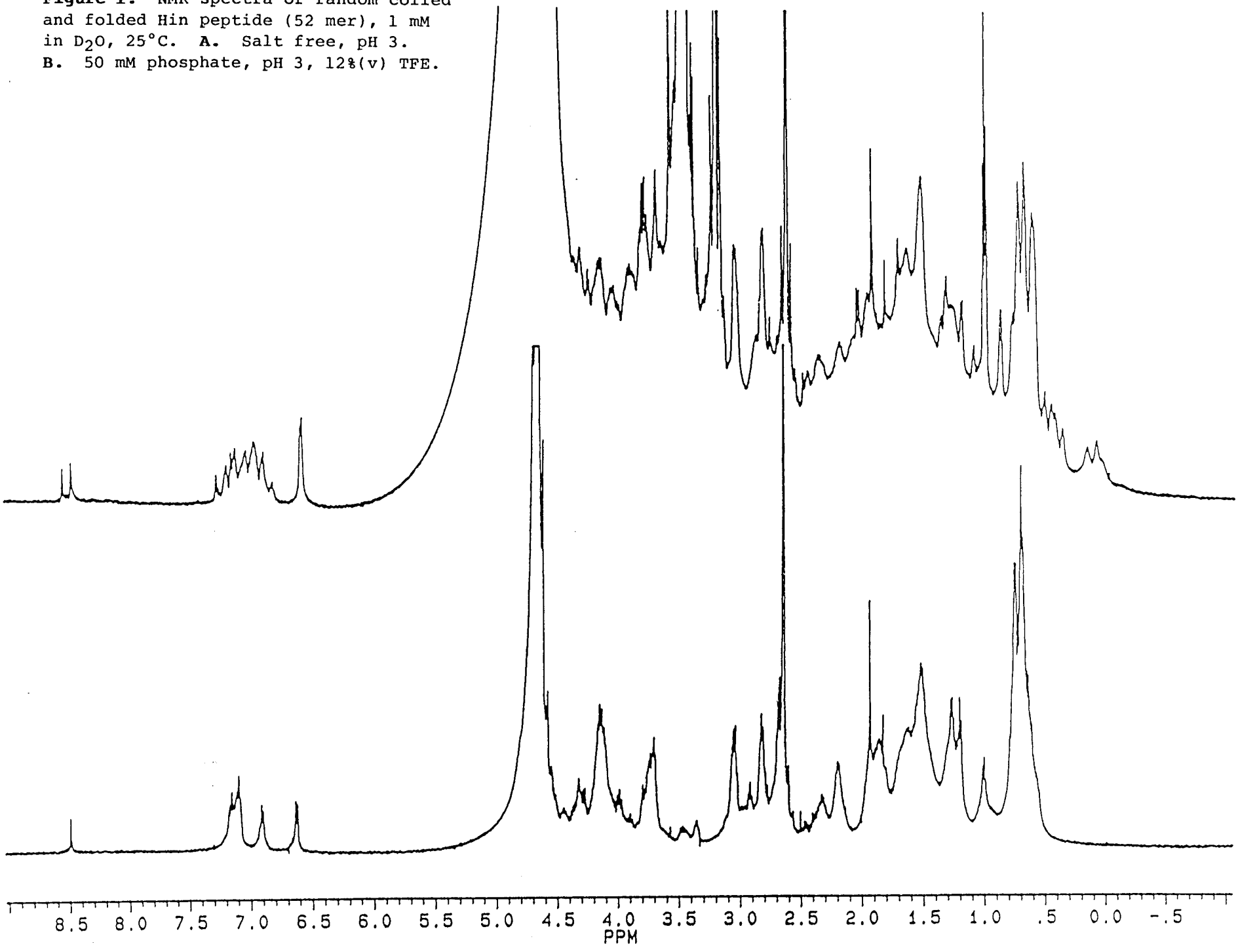


Figure 2. Concentration dependence of the helical contents of the peptide. **Solid line: A.** pH = 7.6, 20 mM phosphate and NaCl. **Broken lines:** pH = 3.4, 50 mM phosphate; **B.** 0% TFE, **C.** 5% TFE, **D.** 12% TFE, **E.** 20% TFE.

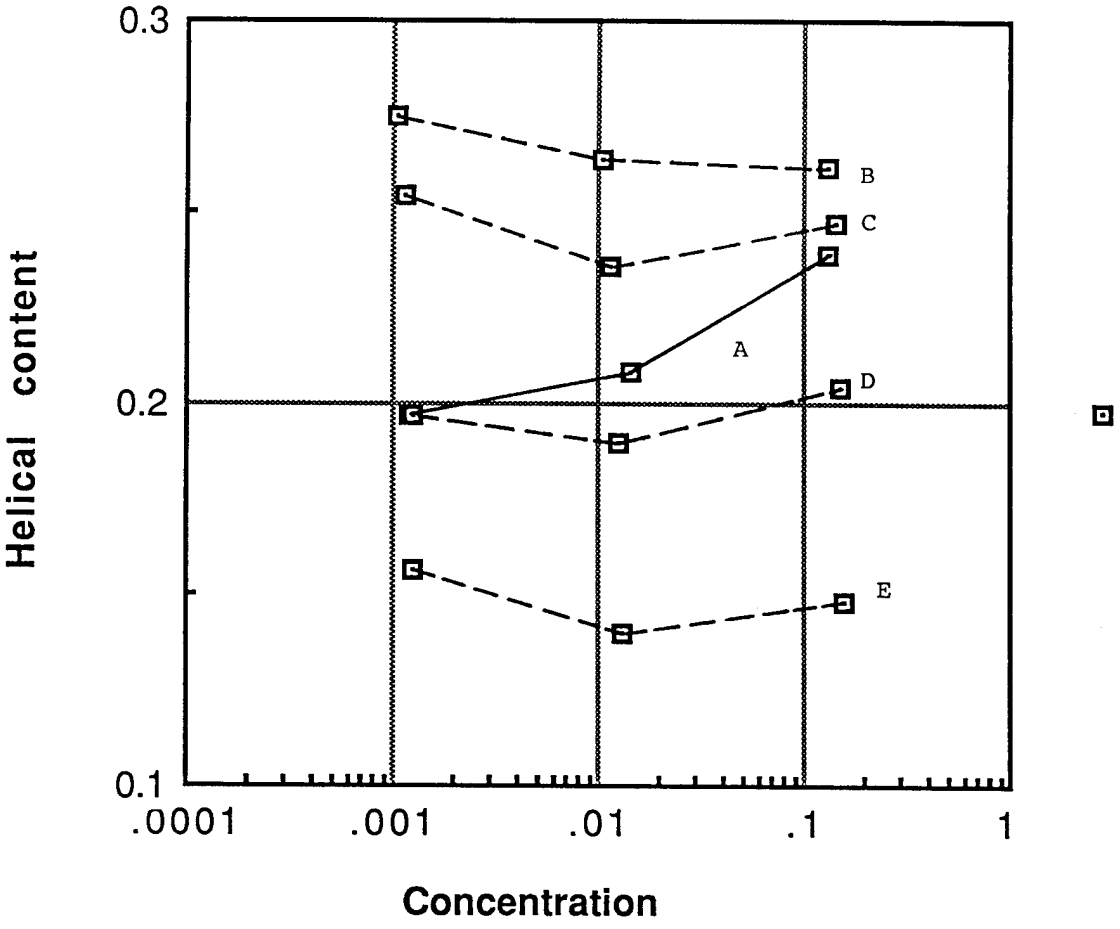


Figure 3. Circular dichroism spectra of the peptide in buffer of pH = 3.4, 50 mM phosphate. **A.** 0% TFE, **B.** 5% TFE, **C.** 12% TFE, **D.** 20% TFE. The helical content of the peptide increases as percentage of TFE increases.

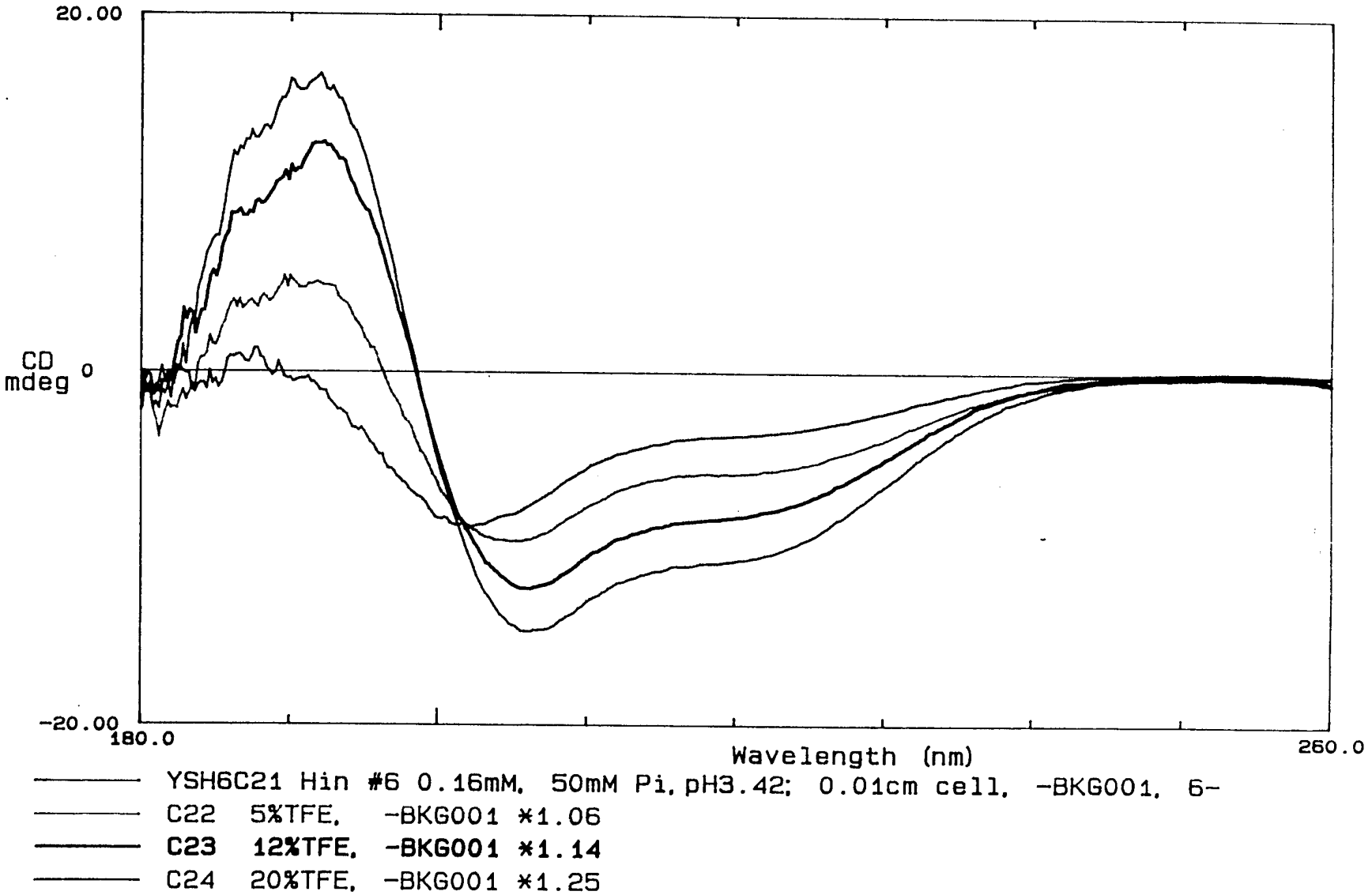


Figure 4. CD of the peptide in the range of 250 nm to 300 nm. **A.** 0% TFE, **B.** 5% TFE, **C.** 12% TFE, **D.** 20% TFE. The signal at 275 nm peaks at 12% TFE (C).

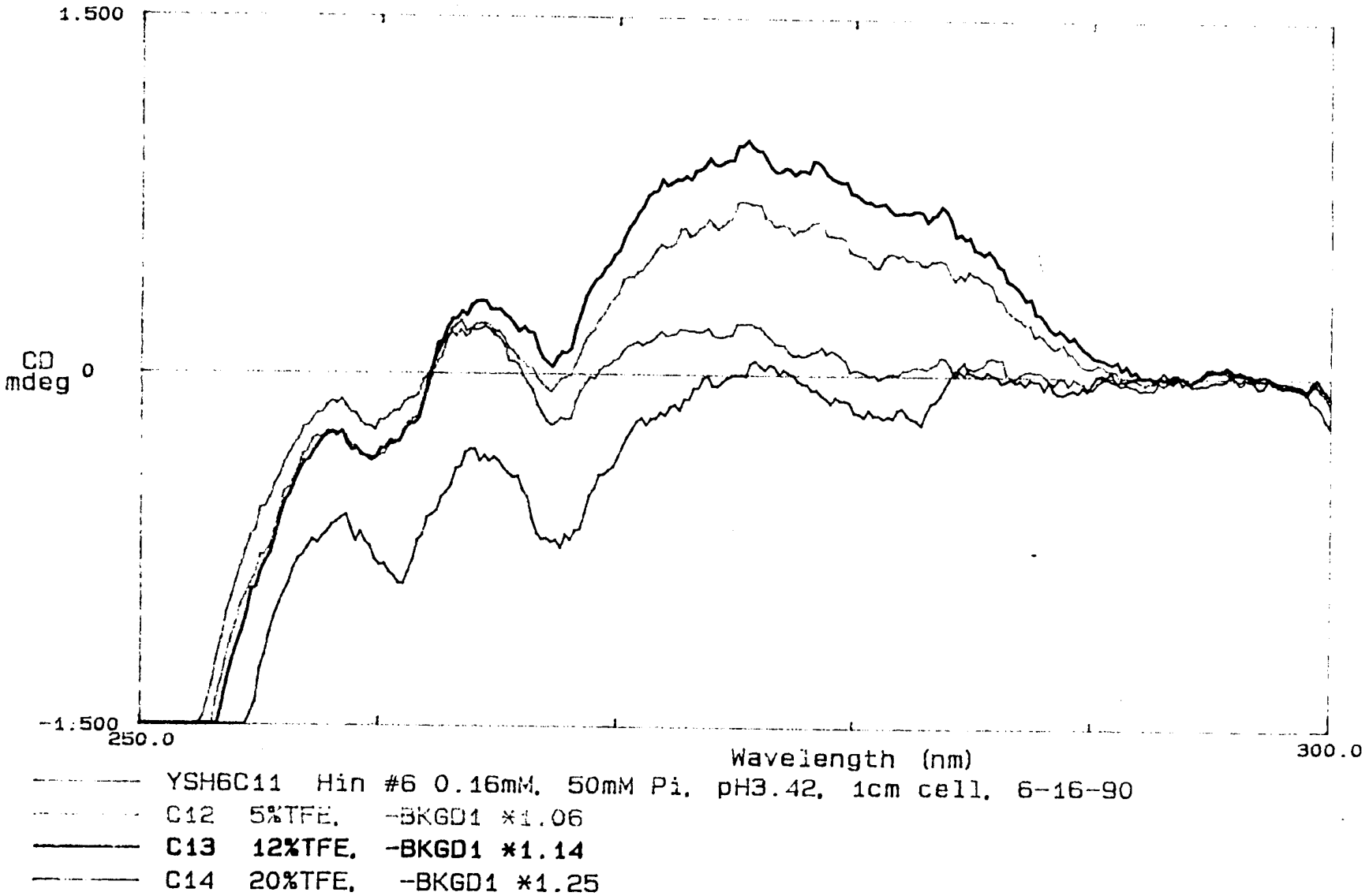
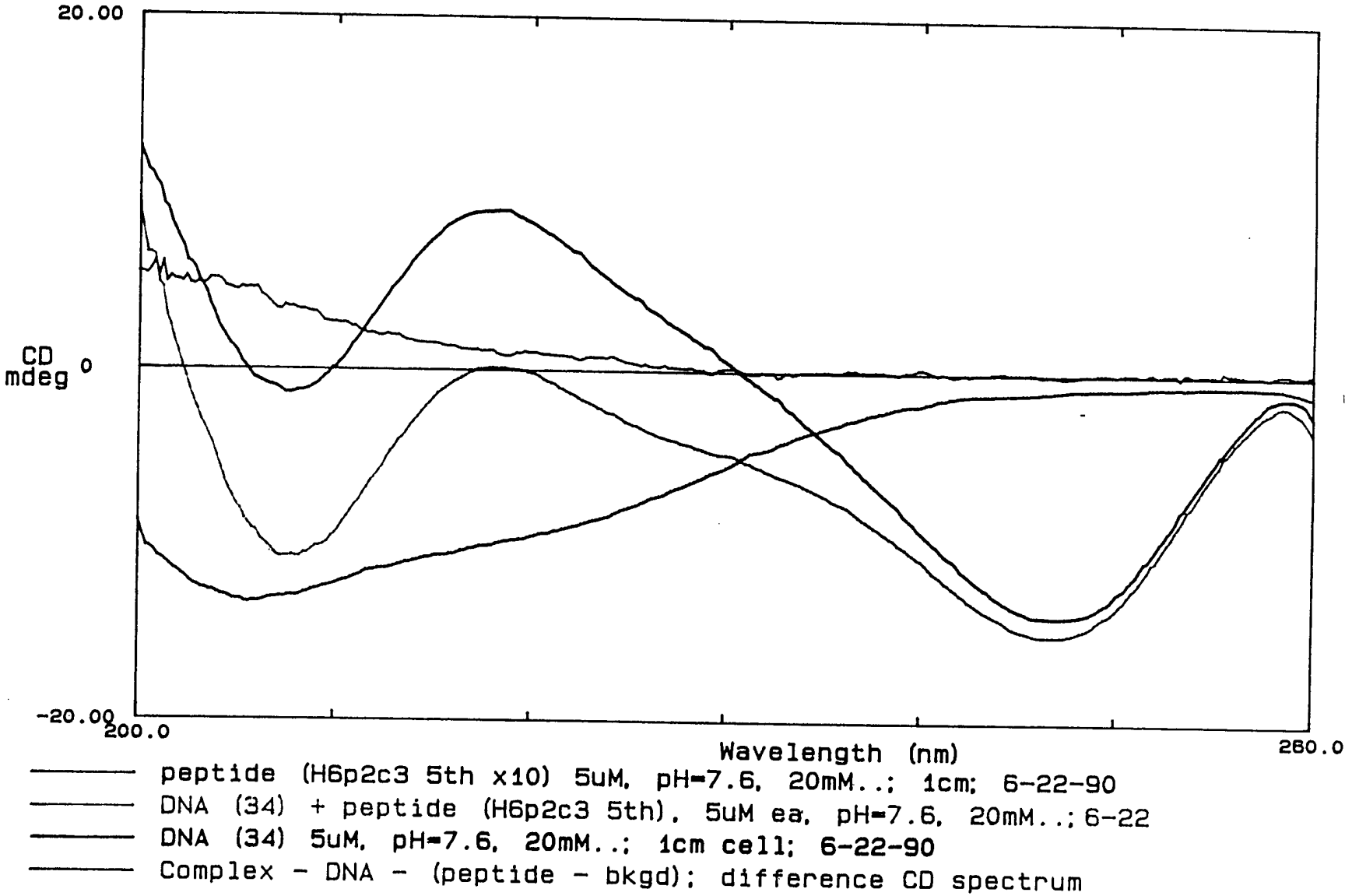


Figure 5. CD of the peptide (A), the DNA (B), the peptide and the DNA mixture (C), and the difference spectrum of the three (D). The change in secondary structure of the peptide upon binding is small. Samples are measured in 20 mM phosphate and NaCl, pH 7.6 aqueous solution.



OPERATOR	SEQUENCE	[IPTG] 0μM	10μM	100μM	1,000μM
hixC WT	TTATCAAAAACCATGGTTTTTGATAA	1×10^{-5}	2×10^{-5}	0.4	0.4
-1C+1G	TTATCAAAAACCGGGTTTTTGATAA	9×10^{-2}	0.1	0.4	0.4
-1G+1C	TTATCAAAAACCCGGTTTTTGATAA	8×10^{-8}	2×10^{-5}	0.4	0.3
-1T+1A	TTATCAAAAACCTAGGTTTTTGATAA	7×10^{-6}	2×10^{-5}	0.4	0.4
-2A+2T	TTATCAAAAACATTGTTTTTGATAA	7×10^{-6}	1×10^{-5}	0.4	0.3
-2G+2C	TTATCAAAAACGATCGTTTTTGATAA	9×10^{-6}	1×10^{-5}	0.4	0.4
-2T+2A	TTATCAAAAACTATAGTTTTTGATAA	7×10^{-6}	1×10^{-5}	0.4	0.3
-3A+3T	TTATCAAAAAACATGTTTTTTGATAA	5×10^{-8}	2×10^{-6}	5×10^{-4}	0.4
-3G+3C	TTATCAAAAAGCATGCTTTTTGATAA	5×10^{-6}	7×10^{-6}	0.2	0.3
-3T+3A	TTATCAAAAATCATGATTTTTGATAA	7×10^{-6}	1×10^{-5}	0.4	0.3
-4C+4G	TTATCAAAACCCATGGGTTTTGATAA	7×10^{-6}	9×10^{-6}	0.2	0.4
-4G+4C	TTATCAAAAGCCATGGCTTTTGATAA	3×10^{-8}	6×10^{-6}	2×10^{-4}	0.4
-4T+4A	TTATCAAAATCCATGGATTTTGATAA	8×10^{-6}	1×10^{-5}	0.3	0.4
-5C+5G	TTATCAAACACCATGGTGTTTGATAA	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$
-5G+5C	TTATCAAAGACCATGGTCTTTGATAA	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$
-5T+5A	TTATCAAATACCATGGTATTTGATAA	0.1	0.1	0.4	0.3
-6C+6G	TTATCAACAACCATGGTTGTTGATAA	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$
-6G+6C	TTATCAAGAACCATGGTTCTTGATAA	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$
-6T+6A	TTATCAATAACCATGGTTATTGATAA	8×10^{-6}	1×10^{-5}	0.3	0.3
-7C+7G	TTATCACAAACCATGGTTTGTGATAA	3×10^{-7}	2×10^{-6}	4×10^{-2}	0.3
-7G+7C	TTATCAGAAACCATGGTTTCTGATAA	4×10^{-6}	3×10^{-6}	0.3	0.3
-7T+7A	TTATCATAAACCATGGTTTATGATAA	5×10^{-6}	9×10^{-6}	0.4	0.4
-8C+8G	TTATCCAAAACCATGGTTTTGGATAA	3×10^{-6}	3×10^{-6}	0.4	0.3
-8G+8C	TTATCGAAAACCATGGTTTTCGATAA	3×10^{-6}	5×10^{-6}	0.4	0.4
-8T+8A	TTATCTAAAACCATGGTTTTAGATAA	6×10^{-6}	9×10^{-6}	0.4	0.4
-9A+9T	TTATAAAAAACCATGGTTTTTTATAA	0.1	0.2	0.4	0.4
-9G+9C	TTATGAAAAACCATGGTTTTTCATAA	2×10^{-5}	3×10^{-5}	4×10^{-3}	0.3
-9T+9A	TTATTAAAAACCATGGTTTTTAATAA	6×10^{-4}	5×10^{-4}	0.4	0.3
-10A+10T	TTAGCAAAAACCATGGTTTTTGTTAA	7×10^{-7}	6×10^{-7}	1×10^{-6}	2×10^{-6}
-10C+10G	TTACCAAAAACCATGGTTTTTGGTAA	1×10^{-2}	0.1	0.4	0.4
-10G+10C	TTAGCAAAAACCATGGTTTTTGCTAA	1×10^{-7}	4×10^{-7}	1×10^{-6}	5×10^{-5}
-11C+11G	TTCTCAAAAACCATGGTTTTTGAGAA	1×10^{-5}	5×10^{-5}	0.4	0.4
-11G+11C	TTGTCAAAAACCATGGTTTTTGACAA	5×10^{-7}	2×10^{-6}	4×10^{-4}	0.2
-11T+11A	TTTTCAAAAACCATGGTTTTTGAAAA	3×10^{-6}	1×10^{-5}	0.4	0.3
-12A+12T	TAATCAAAAACCATGGTTTTTGATTA	2×10^{-6}	3×10^{-6}	0.2	0.3
-12C+12G	TCATCAAAAACCATGGTTTTTGATGA	7×10^{-8}	2×10^{-5}	0.4	0.3
-12G+12C	TGATCAAAAACCATGGTTTTTGATCA	8×10^{-8}	5×10^{-8}	2×10^{-8}	2×10^{-7}
-13A+13T	ATATCAAAAACCATGGTTTTTGATAT	8×10^{-5}	1×10^{-4}	0.3	0.4
-13C+13G	CTATCAAAAACCATGGTTTTTGATAG	5×10^{-5}	1×10^{-4}	0.3	0.4
-13G+13C	GTATCAAAAACCATGGTTTTTGATAC	4×10^{-5}	8×10^{-5}	0.3	0.3