1-1-1997

Reversible gelation of genetically engineered macromolecules.

Wendy A. Petka
University of Massachusetts Amherst


2 2020 h920 99027E

REVERSIBLE GELATION OF GENETICALLY ENGINEERED MACROMOLECULES

A Dissertation Presented

by

WENDY A. PETKA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1997

Polymer Science & Engineering

REVERISBLE GELATION OF GENETICALLY ENGINEERED MACROMOLECULES

A Dissertation Presented

by

WENDY A. PETKA

Approved as to style and content by:

David A. Tirrell, Chair

Thomas P. Russell, Member

Lynmarie K. Thompson, Member

Richard J. Farris, Department Head
Polymer Science and Engineering Department

DEDICATION

I would like to dedicate this work to my mother whose perseverance of education and justice will always be remembered.

ACKNOWLEDGMENTS

As scientists, we learn that all things change and evolve with time and people are no exception to this rule. Before I move onto a new phase in my career and life, I would like to acknowledge those who have invested in me during my tenure at UMASS. The first and foremost is my advisor, David Tirrell. I would like to thank him for giving me the opportunity to explore a field of polymer chemistry that I would have otherwise missed and for providing me with a scientific foundation upon which to build. I would also like to thank my committee members Lynmarie Thompson and Thomas Russell for taking part in my dissertation work. In addition, I would like to thank Murugappan Muthukumar for participating in my dissertation work in its early stages.

There have been many people who have contributed to my education, my research, and my well-being in the past five years. The first person I would like to acknowledge is Kevin McGrath. He was a mentor for my project in the first few years and subsequently became a collaborator. In addition, I would like to thank James Harden for becoming an integral part of my gelation research. I would like to thank Brian Price for being a friend and for the use of his analyses programs not only in my research but in other group members research as well. I am also indebted to Max Prokopy for his assistance in preparing some of my materials. I would like to thank Alan Waddon for helping me with X-ray, Charlie Dickinson for NMR assistance and Wolfgang Wedler for assisting me with rheological measurements. A special thanks goes to those who make our group and the department run smoothly, namely Linda Strzegowski, Eileen Besse, Eleanor Thorpe, and Sophie Michelson.

I have been very fortunate to have gained many lasting friendships over the past few years. These friends have been an important part of the whole graduate school experience not only on a scientific basis but on a personal level too. I would like to thank the gang Mike, Darrin, Bert, Darius, and Shalabh for all of the good times and memories. I would also like to thank Meredith for her valuable friendship and support over these few years. I would like to thank Vipavee for being a good friend and for updating my wardrobe with the most fashionable Thai styles and to thank Regina for all of the artwork that hangs in my apartment.

In our careers, we will encounter one of two kinds of people, those which are associates and those which are "true" friends. I would like to thank Carissa and Dong for their sincerity and for belonging to this latter group. Thanks to Eric W. and Mark for their fruitful discussions and for their comedy in the lab and elsewhere. I would also like to thank Laurie, Geni, Susan, Tim, Vince, Eiichiro, Bo, Mike, Jeff, Wendy, Jim, Lynore, Ajay, Seenu, Eric C., and Nandita for their friendships and their insightful conversations.

I would also like to thank my high school, college friends, and non-polymer people for supporting, emailing, calling, and visiting me over the past five years. Thanks to Sheila, Erica, Alison, Erin, Kelly, Cindy, Josh, Tara, and Chris for reminding me that a whole other world is out there. These friendships do not fade with time or distance.

A special thanks goes to my boyfriend, Elbert, for being a great person. I am very lucky to have found someone as funny, smart, kind, and generous as he. Finally, I would like to acknowledge my family members for their unconditional support. My grandfather, uncle, and brother have supported me not only mentally but financially for most of my life. I know they are very proud of me but I am equally as proud of them.

ABSTRACT

REVERSIBLE GELATION OF GENETICALLY ENGINEERED MACROMOLECULES

SEPTEMBER 1997

WENDY A. PETKA, B.S., CHATHAM COLLEGE

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor David A. Tirrell

Genetic engineering of protein-based polymers offers distinct advantages over conventional synthesis of polymers. Microorganisms can synthesize high molecular weight materials, in relatively large quantities, that are inherently stereoregular, monodisperse, and of controlled sequence. In addition, specific secondary and higher order structures are determined by this protein sequence. As a result, scientists can design polymers to have unique structural features found in natural protein materials and functional properties that are inherent in certain peptide sequences.

For this reason, genetic engineering principles were used to create a set of artificial genes that encode twelve macromolecules having both α-helical and disordered coil protein sequences with the last amino acid being cysteine (cys) or tryptophan (trp). Triblock copolymer sequences having coiled-coil protein ends, A or B, where A and B represent α-helical acidic and basic leucine zipper proteins, separated by a water soluble flexible spacer coil protein, C, where C represents $[(AG)_3PEG]_n$ (n=10 or 28), showed reversible physical gelation behavior. This behavior is believed to result from the

aggregation of two or more helices that form physical crosslinks with the disordered coil domain retaining solvent and preventing precipitation of the chain.

Diffusing wave spectroscopy was used to investigate the gelation behavior of $AC_{10}Acys$ in buffer when environmental conditions such as pH, temperature, and concentration were varied. The dynamic intensity autocorrelation function recorded over time for 5% (w/v) $AC_{10}Acys$ showed that the protein behaved as a gel at pH 6.7-8.0 and that the melting point was between 40 °C and 48 °C.

In addition to the triblock results, the incorporation of 5',5',5'-trifluoroleucine (Tfl) in place of leucine (Leu) in the A and B blocks was accomplished by synthesizing proteins in bacterial hosts auxotrophic for Leu. The substitution of Tfl for Leu in A and B was confirmed by electrospray mass spectrometry. Amino acid analyses performed on purified Tfl A and Tfl B populations suggested 66% and 38% levels of Tfl substitution, respectively. Thermal denaturation temperatures measured by circular dichroism of the Tfl containing helices were higher than those of the corresponding Leu containing helices by 8 °C and 13 °C for A and B respectively.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Model Biological Polymer Systems

Nature provides us with many unique material systems that are important in both organization and function. For instance, some chemical and mechanical properties found in natural fibers such as wool and silk that are derived from animal origins, or those of cotton, jute, and flax derived from vegetable origins have not yet been duplicated in synthetic fibers. Wool, for example, contains a high content of certain nonpolar amino acids such as glycine and proline which are thought to make woolen fabrics both moisture removing and heat retaining materials (1,2). Silk contains certain amino acids which impart both strength and toughness in fibers, whereas the amino acid sequences of mammalian elastin fibers provide flexibility. These specific properties, along with many others, are attractive from a materials standpoint, and syntheses of natural assemblies would be a first step in elucidating the chemical compositions and conformations which give rise to specific structures and functions in materials.

Repeating patterns of amino acids, as in fibrous proteins, are one group of constituents in naturally occurring materials. One method of synthesizing these polypeptides is by genetic engineering or recombinant DNA technology. This technology utilizes the same machinery that naturally produces protein for the survival of the host organism. The general process of making genetically synthesized materials begins with inserting a chemically synthesized piece of DNA, which encodes a specific

pattern of amino acids, into a circular cloning vector through a series of cutting and ligating reactions. Once the proper DNA sequence is verified, this piece of DNA is cut from the cloning vector and inserted into an expression vector which allows for protein production in a host microorganism such as *E. coli*. *E. coli* cells are then grown to some optimal density and production of the protein is initiated. After an allotted amount of time, the cells are lysed and the proteins removed for purification and characterization. The advantages of this protein polymer technology (3) are that microorganisms can synthesize high molecular weight materials, in relatively large quantities, that are inherently stereoregular, monodisperse, and of controlled sequence.

The most important advantage of designing and synthesizing polymeric materials by this biosynthetic route is the ability to control the structure of a material on microscopic and macroscopic levels. It is rare that both stereoregularity and monodispersity are achieved by conventional methods of polymerization such as step, chain, ring-opening, or coordination methods. Biosynthetic polymers, on the other hand, have both of these characteristics built in naturally and in many cases, the degree of structural control extends to the secondary, tertiary, and quaternary levels. In turn, the physical properties associated with these biomaterials can be developed on the bases of shape, hydrophilic/hydrophobic character, and charge placement.

The design of structurally and functionally interesting biological macromolecules requires an understanding of protein structure and how it relates to amino acid sequence. The different levels of structural complexity of a polypeptide in solution or in the solid state are directly related to the chemical nature of both macromolecule and its environment. Variables such as pH, temperature, and concentration will influence the

types of conformations that polypeptides adopt in solution. Favorable or unfavorable solvent-polymer interactions will contribute to the overall dimensions of the polymer. For instance, a good solvent results in expanded conformations from preferred segment-solvent interactions and a poor solvent tends to force chains to contract to exclude solvent and to aggregate and precipitate (4). The correlation between conformations existing in natural polypeptides and their amino acid periodicity form the basis for rational materials design.

Proteins are generally classified as either globular or fibrous and have various conformations within each of these groups (5). Fibrous proteins, consist of repeated sequences of amino acids which have short range order and in some cases, long range three-dimensional order. The secondary structures of fibrous proteins are defined in terms of bond geometry, electron delocalization, sequence symmetry, polar and/or ionic bond character and hydrogen bonding propensity (6). The geometry of a polypeptide backbone is described by angles ϕ_i, Ψ_i, and ω_i, for N_i-$C_{\alpha i}$, $C_{\alpha i}$-C_i, and C_i-N_{i+1} bonds respectively. Electron delocalization of the C-N bond directly influences the ω_i rotation since the C-N bond distance in a peptide is 1.32 Å compared to 1.47Å for a normal C-N bond (6). As a result, sterics and electron configuration will force six atoms in the peptide chain to be planar and fixes ω_i at or near 0 °C or 180 °C; the latter being more energetically preferred.

Fibrous proteins are formed from a repeated peptide sequence which in turn leads to repeated ϕ, Ψ, and ω bond rotational sequences. On the other hand, globular proteins will have very few peptide repeats or bond rotational symmetries. The ϕ, Ψ, and ω angles of fibrous proteins therefore will depend on van der Waals radii or the overlap

potential of neighboring atoms, ionic or polar forces acting over longer distances than the van der Waals interactions, and intrachain or interchain H-bonding that exists in secondary or tertiary structures. Characteristic ϕ, Ψ, and ω angles have been calculated for known α-helical, β-turn, and random coil protein sequences (7).

The recombinant DNA process previously outlined is being used to synthesize structurally and functionally unique copolymers to be described herein. Both α-helical and random coil fibrous proteins are combined in one chain to impart both rigidity through the "hard" helical segment and flexibility through the "soft" coiled segment. The general design of the triblock copolymer is a helix-coil-helix motif where the helix belongs to a special class of coiled-coil proteins, leucine zippers, and the random coil sequence belongs to a class of repetitive alanylglycine proteins, $[(AG)_3PEG]_n$ (n = 10, 28) (8, 9). The structural motif includes helical end blocks that will recognize and associate with each other through interchain hydrophobic and electrostatic interactions separated by random coil domains that will act as flexible water soluble spacers inserted between these ends.

1.2 Molecular Recognition and Association of Engineered Macromolecules

The helical elements are unique because they can be classified as part of a group of materials that spontaneously self-organize into well-defined supramolecular arrays under a given set of conditions (10). A range of potential applications that exist for molecular recognition of macromolecules include reversible gelling systems and adhesives (11), thermoreversible networks (12, 13, 14), non-covalent fiber networks,

and recyclable materials. Natural self-assembling protein materials that have reversible structural and physical properties given the appropriate pH, temperature, and solvent conditions can be modeled and designed after naturally occurring proteins found in both the body and the environment. For this reason, proteins modeled after a class of self-assembling α-helices called coiled-coils are being used in this research.

1.2.1 Coiled-Coil and Leucine Zipper Proteins

Polypeptides such as keratin and myosin are known to adopt α-helical conformations which were originally thought to pack into parallel orientations with straight axes (15). Not long after this proposal, Pauling and Corey observed that the actual packing of α-keratin was a long range phenomenon and that the axes of individual helices pursued a helical course to form multistranded cables (16). The origin of deformation resulted from what was believed to be repeating sequences of amino acids; every fourth and seventh amino acid was repeated. At the same time, Crick (17) proposed that very long chains of amino acids could indeed form what was termed "coiled-coils". It was in this work (18) that the calculations of the predicted X-ray diffraction pattern for double and triple stranded ropes were generated. More importantly though was that the calculated equatorial and meridional reflection were very similar to the experimental results of MacArthur (19) and Perutz (20). The results showed meridional arcs at spacings of approximately 5.1 Å and 1.5 Å and a group of reflections on and near the equator at spacings approximately 10.5 Å.

These investigations led to the initial studies of how, why and when α-helices recognize each other and pack into a favorable conformational state. It was not until 1991 that Alber and co-workers (21) determined the X-ray structure of a 33-residue leucine zipper portion of the yeast transcription factor, GCN4, a special type of coiled coil protein that belongs to a class of transcription factors that regulate the expression of many different genes in both simple and complex organisms. Since 1991, others have determined high resolution X-ray crystal structures of similar two stranded coiled-coils (22, 23, 24).

Prior to solving the crystal structures of this family of eukaryotic transcription factors referred to as the basic-region-leucine-zipper (bZIP motif), other structural features of the leucine zipper had been determined. The first amino acid sequence of a coiled-coil, tropomyosin, indicated that the stabilization of the super coiled helices resulted from a 3-4 hydrophobic repeat (25, 26). It was suggested that a seven amino acid repeat, $(a\ b\ c\ d\ e\ f\ g)_n$ where *a* and *d* amino acids were hydrophobic residues, defined a coiled-coil. It was Hodges *et al.* who synthesized and characterized the first synthetic coiled-coil analog of tropomyosin (27). While investigating short coiled-coil domains of DNA binding proteins, Landshulz *et al.* named this class of proteins "leucine zippers" for what was believed to be leucine sidechains interdigitating like the teeth of a zipper at the *d* position on adjacent helices (28). However, it was later confirmed by the X-ray crystallographic data previously mentioned that the leucine sidechains pack in a side-to-side arrangement on adjacent chains.

In addition to the hydrophobic core, which is formed by the side-to-side packing of the *a* and *d* residues that are aligned on the complementary faces of two 7_2 α-helices,

it has been shown that the formation of coiled-coils is further modulated by interactions between regularly spaced charged groups at specific positions in the general heptad repeat (29). The stabilization of the super coiled helices is enhanced or diminished by positioning attractive or repulsive charged groups at *e* and *g* positions (30). It is then possible to design polypeptides that have the propensity to form parallel or antiparallel interchain attractive coiled-coils based on charge-charge interactions at the *e* and *g* positions on adjacent helices (31).

Parallel and antiparallel arrangements of two leucine zippers are best illustrated by considering the general positions of residues that are in proximity to each other. In a coiled-coil structure, positions *a* and *d* of two 7_2 helices (two turns per seven amino acids with a pitch of 5.2 Å) are associated at the interface and intertwine to form a super coiled structure with a slight left-handed twist. The frequency of finding hydrophobic residues, mainly leucines, at these two positions in coiled-coils has been calculated by Lupas and co-workers (32). A general representation of parallel and anti-parallel dimerizations of two α-helical leucine zippers is shown in Figure 1.1 where one chain is denoted by a general heptad repeat $(a\ b\ c\ d\ e\ f\ g)_n$ and the other by $(a'\ b'\ c'\ d'\ e'\ f'\ g')_n$. In addition to the dimer state, leucine zippers are also known to form trimeric (33, 34, 35, 36, 37, 38) and tetrameric aggregates (39, 40, 41). These higher order aggregate states are referred to as multiple-helix bundles instead of coiled-coils. This distinction has been made because of the difference in packing between hydrophobic faces; coiled-coils contain a narrow hydrophobic surface whereas multiple helix bundles have a wide surface. The latter surface involves additional hydrophobes in the packing arrangement. It has been suggested that the conformation of the core amino acid side chains determine

whether a two-, three-, and four-stranded oligomerization state will predominate (42, 43).

1.2.2 Original Design of Proteins A1 and B1

A set of coiled-coil proteins, A1 and B1, designed by McGrath *et al.* (44) is being used as associating α-helical proteins (Figure 1.2). Both A1 and B1 proteins contain a 42 amino acid sequence (six internal heptad repeats, *abcdefg*) which constitutes the coiled-coil region. The choice of residues at positions *a* and *d* was based on the *a* / *d* residue patterns see in the *Jun* oncogene product (45), which forms a weakly stabilized homodimer. An algorithm developed by Lupas *et al.* (32), which identifies the most probable amino acids at positions *a*-*g*, was used to select residues occupying *b*, *c*, and *f* positions in coiled-coils. Residues that would increase the solubility of the coiled-coil complex in water were chosen first for these positions. Acidic or basic groups, glutamic acid and lysine, respectively, were placed at the *e* and *g* positions to provide intermolecular attraction or repulsion between chains. The placement of oppositely charged groups at *e* and *g* positions in A1 and B1 proteins was designed to maximize heterodimer formation over either homodimer when solutions equimolar in A1 and B1 are mixed. The underlined regions of the *Helix* in Figure 1.2 emphasize this charge pattern at positions *e* and *g*. Helical wheels representing parallel and antiparallel arrangements of A1 and B1 homodimers and heterodimers are shown in Figures 1.3 and 1.4.

As stated previously, the stabilization of coiled-coils can be modulated by both intra- and interhelical electrostatic interactions (29). Eleven nearest neighbor intrachain electrostatic interactions exist on both A1 and B1 helices. In addition, ten *e*/*g* interchain interactions exist for the dimerization of either A1 or B1 proteins. The intrachain interactions are described and listed in Figure 1.5. Attractive or repulsive interactions along the helix will significantly affect the stability of an individual α-helix as well as state of association of many helices. It has been suggested that both intra- and interhelical interactions are cooperative processes in coiled-coil formation (46).

1.2.3 Evidence for Helical Interactions of Coiled-Coil Proteins

Prior to the construction of new DNA encoding the A1 and B1 helical elements, the biological syntheses of A1 and B1 proteins were done initially to test the feasibility of forming α-helical leucine zipper macromolecules and to see if they would form homodimer and/or heterodimer coiled-coils in solution. The resulting studies have focused on determining the association characteristics of A1-A1, B1-B1, A1-B1 mixtures in solution when environmental factors such as temperature, solvent, and pH are varied.

1.2.3.1 Physical Properties

The physical properties of A1 and B1 complexes are listed in Tables 1.1 and 1.2. The charge pattern at the *e* and *g* positions along the heptad repeat distinguishes A1

from B1; the polar charged groups at these positions determine whether the protein chain is largely acidic or largely basic. Complexation studies for both proteins were performed with sedimentation equilibrium centrifugation analysis (47, 48). These data suggest that homodimer complexes exhibit a concentration-dependent dimer-tetramer equilibrium in solution. Values of K_d are significantly different for A1 (K_d = 76 μM) and for B1 (K_d = 123 μM). In both A1 and B1, tetramer to octamer equilibria constants were observed at concentrations greater than 200 μM. The mixture of A1-B1 gave ambiguous results which are not included in the table.

Although A1 and B1 proteins were originally designed to form parallel coiled-coils, both parallel and antiparallel alignments have to be considered here because the exact orientation is not known. Only one orientation can exist for a particular coiled-coil and therefore the dimerization is either parallel or antiparallel (46). The orientation can be determined with circular dichroism by monitoring the thermal denaturation at 222 nm as a function of concentration for helical sequences containing one cysteine end and for those that do not. O'Shea *et al.* (29) have shown that solutions of GCN4-p1 containing helices that lack cysteine end-groups exhibit concentration-dependent thermal transitions. The same solutions show no concentration-dependent thermal transitions over the concentration range of 0-35 μM for sequences that form parallel homodimers by N-termini linking. A concentration dependence would arise if the molecules did not favor parallel orientations and were free to bind to other helices. Similar results were obtained for parallel orientations of Fos, Jun, and Fos/Jun peptides at concentrations of 1-20 μM (49) and for favorable antiparallel arrangements of coiled coil stem loop CCSL peptide at concentrations of 0.1-1000 μM (50). Both A1 and B1 proteins have been modified to

include a single cysteine at the C-terminus for probing the orientation of A1 and B1. The synthesis of these molecules is described in Chapter 2.

1.2.3.2 Helix Content

Circular dichroism spectroscopy (CD) was used to study the secondary structure formation and energetics of interactions between A1 and B1 protein chains in buffered aqueous solutions. The α-helical conformations of these proteins give rise to a characteristic CD spectrum. The three amide transition CD bands of interest are $-[\theta]_{222}$ ($n \rightarrow \pi^*$), $-[\theta]_{208}$ ($\pi \rightarrow \pi^*$, parallel), and $+[\theta]_{190}$ ($\pi \rightarrow \pi^*$, perpendicular) where $[\theta]_x$ is the molar ellipticity measured at x wavelength in nm. The % helix content can be calculated by the method developed by Chen *et al.* (51) where eight basis spectra of helical proteins were used to determine 100% helix content of a protein x amino acids in length at any wavelength in the 185-250 nm regime. Estimation of helicity (% helix) is calculated by the following equation:

$$\% \text{ helix} = [\theta]_{222} / [\theta]_{max} \qquad \{1\}$$

where $[\theta]_{max} = -39500\,[1-(2.57/n)]$ (deg cm^2 dmol^{-1}) and

where n = 42 for A1 and B1, then

$$[\theta]_{max} = -37082.97 \text{ (deg cm}^2 \text{ dmol}^{-1}\text{)}.$$

In the design of the A1 and B1 proteins, 42 out of 74 amino acids (57 %) comprise the α-helical leucine zipper region. The molar ellipticities at 222 nm of A1 and B1 measured

at 25°C in 50 mM sodium phosphate, 100 mM NaCl pH 7.4 at a total protein concentration of 3 μM were reported as -24390 and -24669 respectively (52). These data correspond to helix contents of 65.8% for A1 and 66.5% for B1.

1.2.3.3 Thermodynamic Analysis

Thermal and urea denaturation studies were done with 3 μM solutions of A1, B1, and A1-B1 complexes and the results are shown in Figures 1.6 and 1.7 (52). The temperature curves show that the thermal transition temperature, T_m, from a folded to an unfolded state is higher for the A1-B1 mixture than either of the A1-A1 or B1-B1 solutions. These results are similar to the urea denaturation profiles which show a higher $[urea]_{½}$ (transition where 50% of the helix is unfolded at 20°C) for the unfolding process of A1-B1 than for the A1-A1 and B1-B1. This means that at pH 7.4 the heterodimer complexes are more stable than either of the homodimer complexes. Table 1.1 compares the data collected and calculated from these curves for A1-A1, B1-B1, and A1-B1 interactions. These data show that the transition temperature for the heterodimer complex is higher than that of either homodimer solution. In addition to this, the free energy of unfolding, ΔG_u, and the difference in the free energies of unfolding, $\Delta\Delta G_u$, also increase upon mixing of equimolar solutions of A1 and B1. Note that ΔG_u is obtained from plotting $\ln K_{d,u}$ (equilibrium constant for dissociation of dimer to monomers) against [denaturant] and extrapolating back to zero denaturant concentration (53, 54) whereas $\Delta\Delta G_u$ can be determined by comparing free energy values of proteins that are calculated when 50% of the protein has unfolded in the presence of denaturant (55, 56). The latter

method minimizes any errors that are introduced in the measurement of the slope by extrapolation back to zero.

If the arrangement (parallel or antiparallel) is known then it is possible to calculate the Kcal/mole stabilization for each attractive interaction going from a homodimer complex to a heterodimer complex. This has been calculated by many groups (45, 56, 57) and the stabilization values range from 0.16-0.20 Kcal/mole per ionic interaction. The exact orientation of A1 and B1 needs to be determined prior to this calculation as the number of attractive interactions differs in the parallel and antiparallel orientations for the homodimer as well as for the heterodimer complexes (refer to Table 1.2).

Thermal denaturation curves as a function of pH were recorded with 12 μM solutions of A1, B1, and the A1-B1 mixture (Figures 1.8, 1.9, 1.10). At lower pH values, the thermal transition occurs at higher temperatures for A1. This might be expected since A1 chains contain carboxylic acid groups at the *e* and *g* positions in the heptad repeat and lower pH values would promote the formation of coiled coils as the sidechains remain protonated. At higher pH values, the negatively charged carboxylate groups might cause interchain electrostatic repulsion and reduce the stability of the coiled coil structure. The opposite trend is observed for B1, and is consistent with this argument because B1 chains contain basic lysine sidechains at the *e* and *g* positions. The data for the equimolar mixture of A1 and B1 shows that at every pH the transition shifted to a higher temperature.

Thermal reversibility studies were carried out with A1, B1, and A1-B1 solutions (12 μM) at pH 7.2 with 30 minute equilibration times between scans. CD spectra were

recorded for each solution after one thermal cycling. A helical structure exists for all three protein solutions at 0°C as shown in Figures 1.11-1.13. However, scans of the solutions at temperatures above the transition temperature for A1 and above but close to the transition temperatures for B1 and A1-B1 indicate that a more randomly coiled structure exists in solution. The reduction of [θ] at 222 nm after thermal cycling was found to be less than 5% for each protein. The interpretation is that the transition from folded to unfolded states are thermally reversible for A1, B1, and A1-B1 solutions.

Table 1.1 Thermodynamic analysis of A1-A1, B1-B1, and A1-B1 interactions.

Dimer Complex	T_m in PBS (°C)[a]	[Urea] ½ (M)[b]	slope m[b]	ΔG_u (Kcal/mole)[c]	$\Delta\Delta G_u$ (Kcal/mole)
A1-A1	58.3	3.57	-1.295	11.54	0.00
B1-B1	75.5	3.66	-1.512	12.55	0.13
A1-B1	86.3	5.76	-1.165	13.76	2.75

[a] T_m = temperature at which 50% of the helix is unfolded in 50 mM sodium phosphate, 100 mM NaCl, pH 7.4.
[b] Concentration of urea at which 50% of the helix is unfolded at 20°C.
[c] ΔG_u = molar free energy of folding calculated from the urea denaturation profiles.

1.3 Summary

The reversibilty of the helical folding of A1 and B1 proteins with pH and temperature forms the basis for the helical-random-helical block copolymer design. The idea is to create an architecture where the non-covalent crosslinks, helical ends, are precisely controlled with pH and temperature. To be described herein is the genetic

construction as well as the protein synthesis of the triblock copolymers, the reversible gelation studies of one of the triblocks synthesized, and the modifications of the original A1 and B1 proteins at the hydrophobic interface.

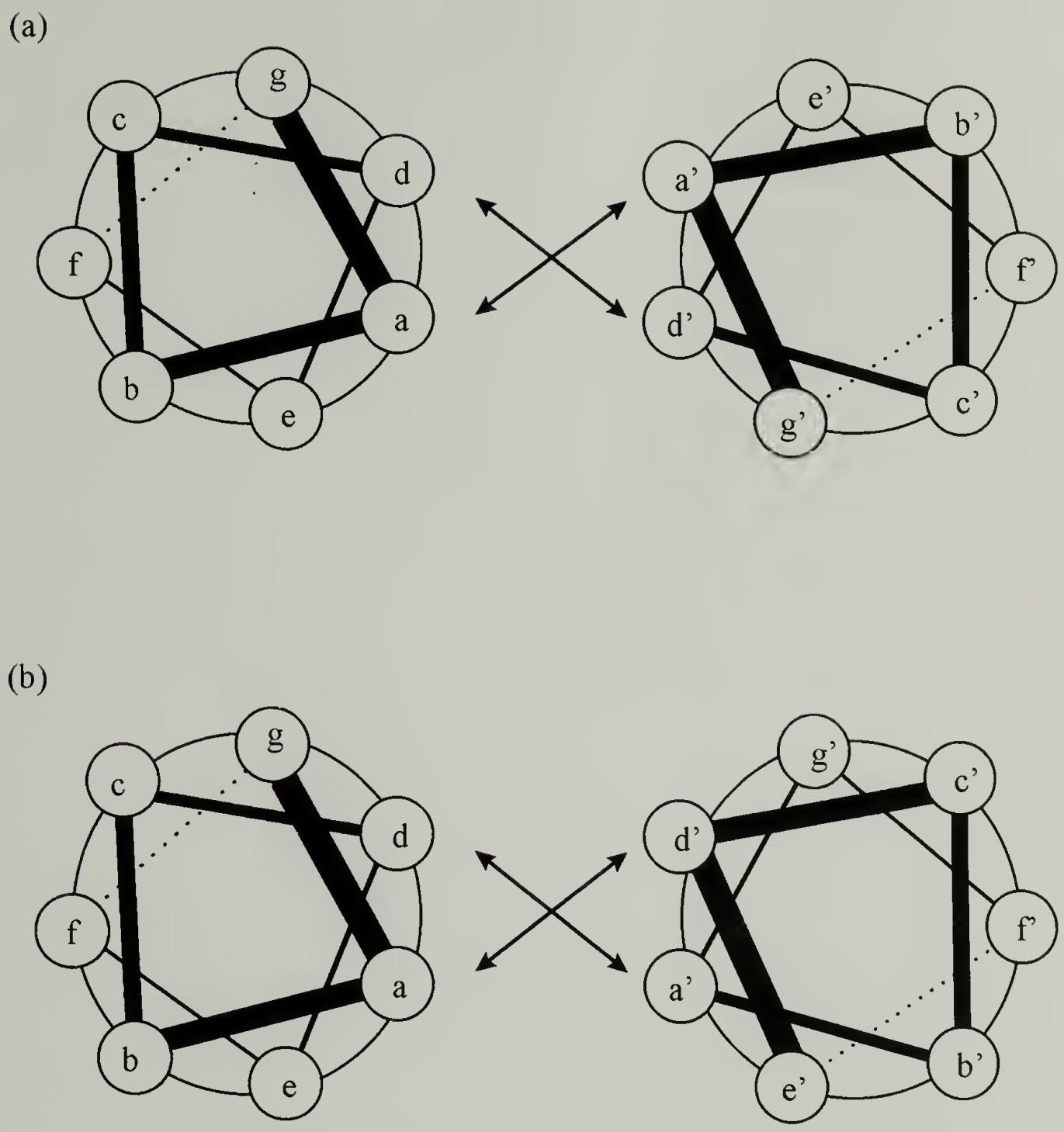
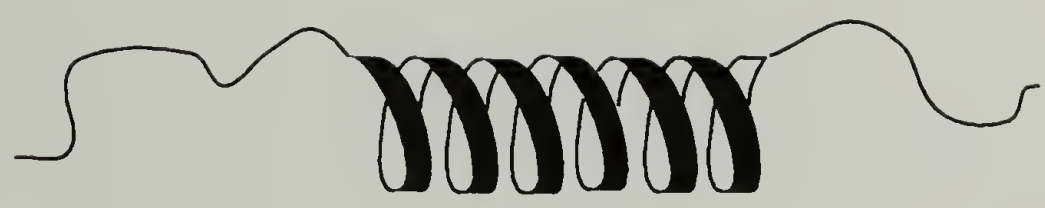


Figure 1.1 Helical wheel representation of coiled-coil structure. (a) Representation of a parallel coiled-coil viewed down the helical axes from the NH_2-termini and (b) representation of an antiparallel coiled-coil viewed down the helical axes from the NH_2- and the COOH-termini. Interhelical hydrophobic interactions are present between the a, a', d, and d' groups whereas the e, e', g, and g' form the electrostatic interhelical interactions.



N' MRGSHHHHHHGSMA - *Helix* - IGDLNNTSGIRRPAAKLN C'

Helix = SGDLENEVAQLEREVRSLEDEAAELEQKVSRLKNEIEDLKAE = A1

Helix = SGDLKNKVAQLKRKVRSLKDKAAELKQEVSRLENEIEDLKAK = B1

Figure 1.2 Amino acid sequences for recombinant leucine zipper proteins, A1 and B1. Charge patterns in the *e* and *g* positions of the *Helix* heptad repeat, $(\,a\,b\,c\,d\,e\,f\,g\,)_6$, are underlined and eight leucine positions in each chain are bolded. Abbreviations for the amino acids are : A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and W, Trp.

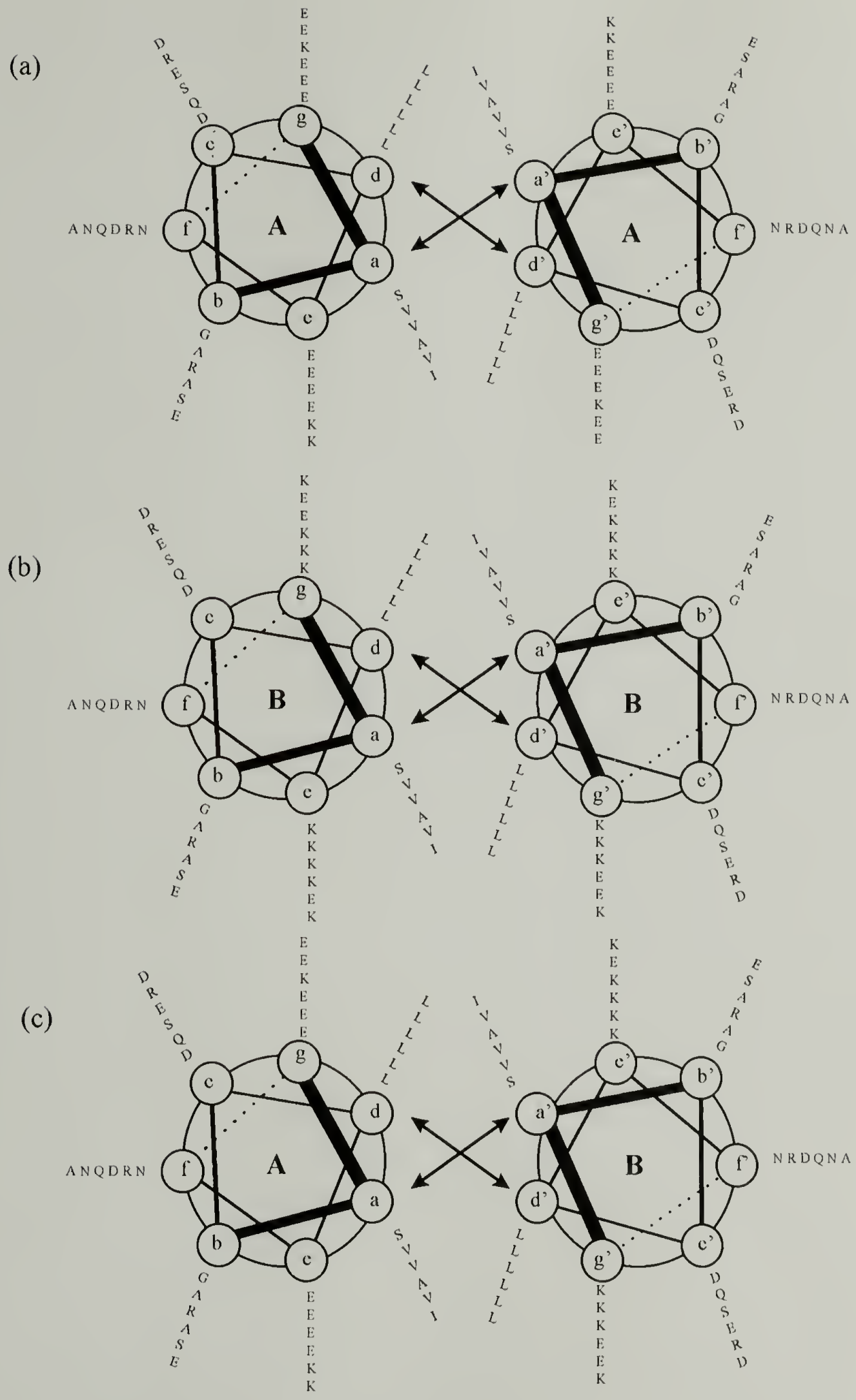


Figure 1.3 (a) Parallel coiled-coil of A1 homdimer (b) parallel coiled-coil of B1 homodimer and (c) parallel coiled-coil of A1-B1 heterodimer.

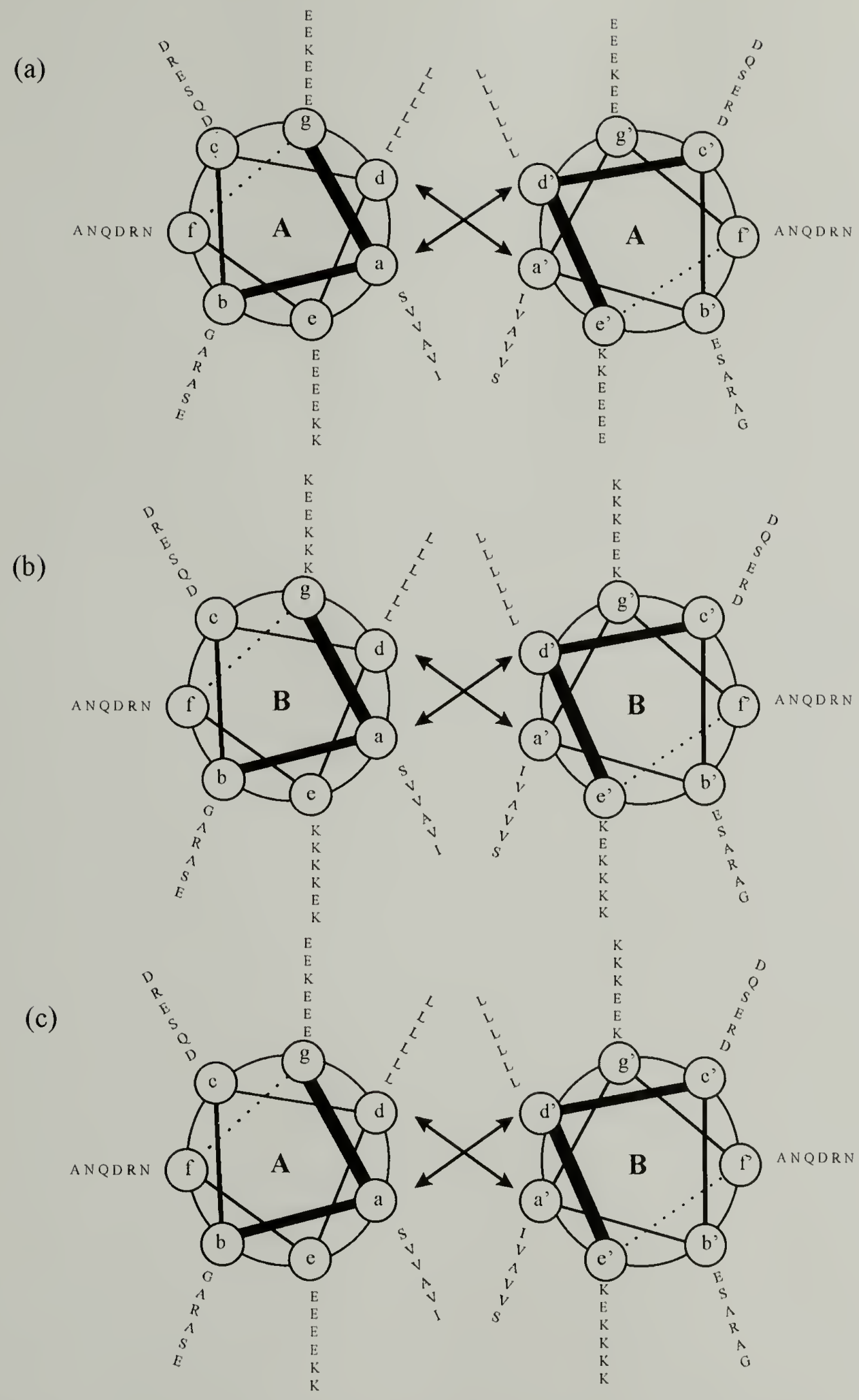


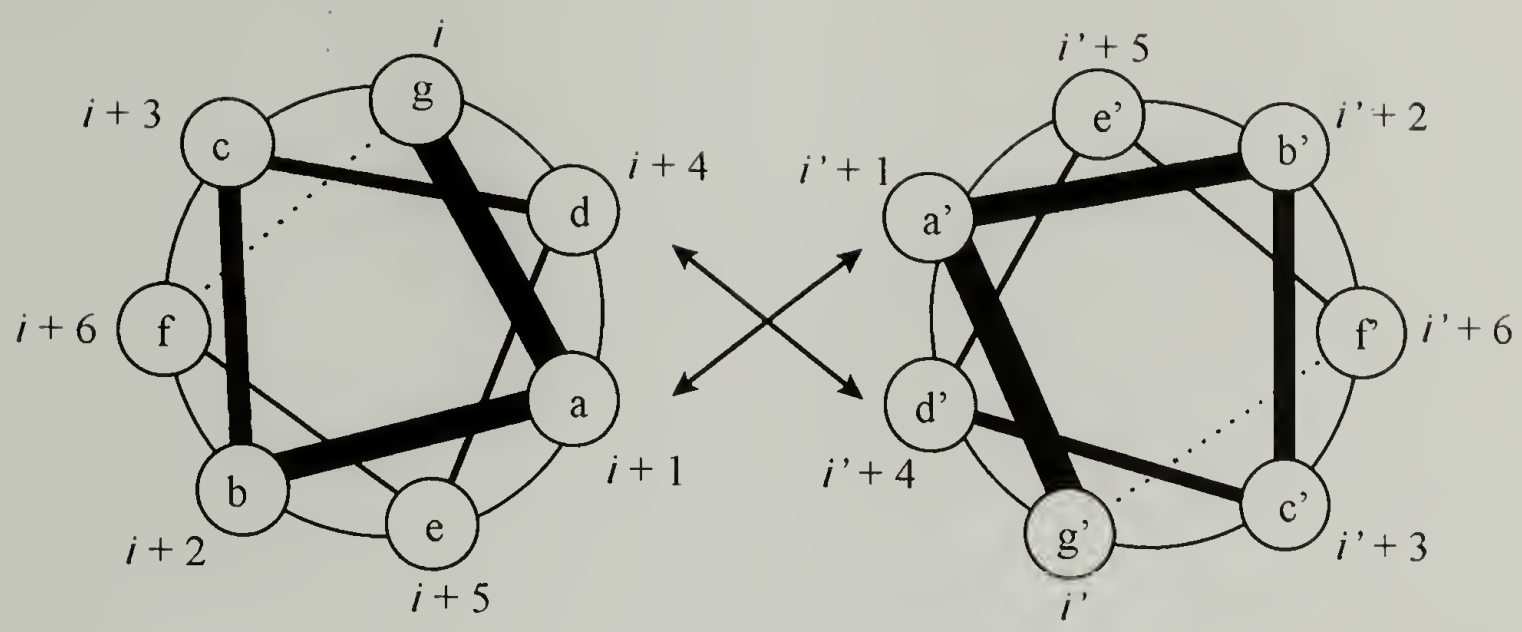
Figure 1.4 (a) Antiparallel coiled-coil of A1 homodimer (b) antiparallel coiled-coil of B1 homodimer and (c) antiparallel coiled-coil of A1-B1 heterodimer.

Table 1.2 Physical properties of synthetic coiled-coils.

Protein	**e/g** charge pattern	Parallel Interchain Interactions[a]	Antiparallel Interchain Interactions[a]	Intrachain Interactions[a]	Monomer Avg. MW[b]	Dimer-Tetramer K_d (in μM)[c]
A1	EEEEEEEKKEKE	4A 6R	2A 8R	12A 16R	8306	76 (50-112)
B1	KKKKKKKEEEKK	6A 4R	4A 6R	12A 16R	8300	123 (1-6)
A1-B1	---	5A 5R	7A 3R	12A 16R	---	---

[a] A = attractive charge interaction (acid-base interaction); R = repulsive charge interaction (includes both acid-acid and base-base interactions). Calculated interactions include **e-g'** and **e'-g** for parallel interhelical dimer interaction and **e'-e** and **g-g'** for antiparallel interhelical dimer interaction. Intrachain interactions include **b-f** (i to i+4), **b-e** (i to i+3), **c-g** (i to i+4), **e-b** (i to i+4), **f-c** (i to i+4), **f-b** (i to i+3) and **g-c** (i to i+3). [b] Average molecular weight of protein sequence calculated by DNA Strider™ 1.1. [c] Analytical ultracentrifugation data measured for protein solutions (24 to 220 μM). 95% confidence limits are shown in parentheses.

(a)



(b)

i to *i*+4		*i* to *i*+3		*i* to *i*+4		*i* to *i*+3			
b	f	b	e	c	g	c	f		
R	D	R	E/K	D	E/K	none			
		E	E/K	E	E/K				
				R	E/K				
				D	E/K				

i to *i*+4		*i* to *i*+4		*i* to *i*+3		*i* to *i*+4		*i* to *i*+3	
e	b	f	c	f	b	g	d	g	c
E/K	R	D	E	R	R	none		E/K	E
E/K	E							E/K	R
								E/K	D

Figure 1.5 Intrachain electrostatic interactions. (**a**) Parallel coiled-coil positions *a*-*g* denoted *i* to *i*+6 and *i'* to *i'*+6. (**b**) Intrachain electrostatic interactions of positions *b*, *c*, *e*, and *f* where *i* represents the position being considered along the helix.

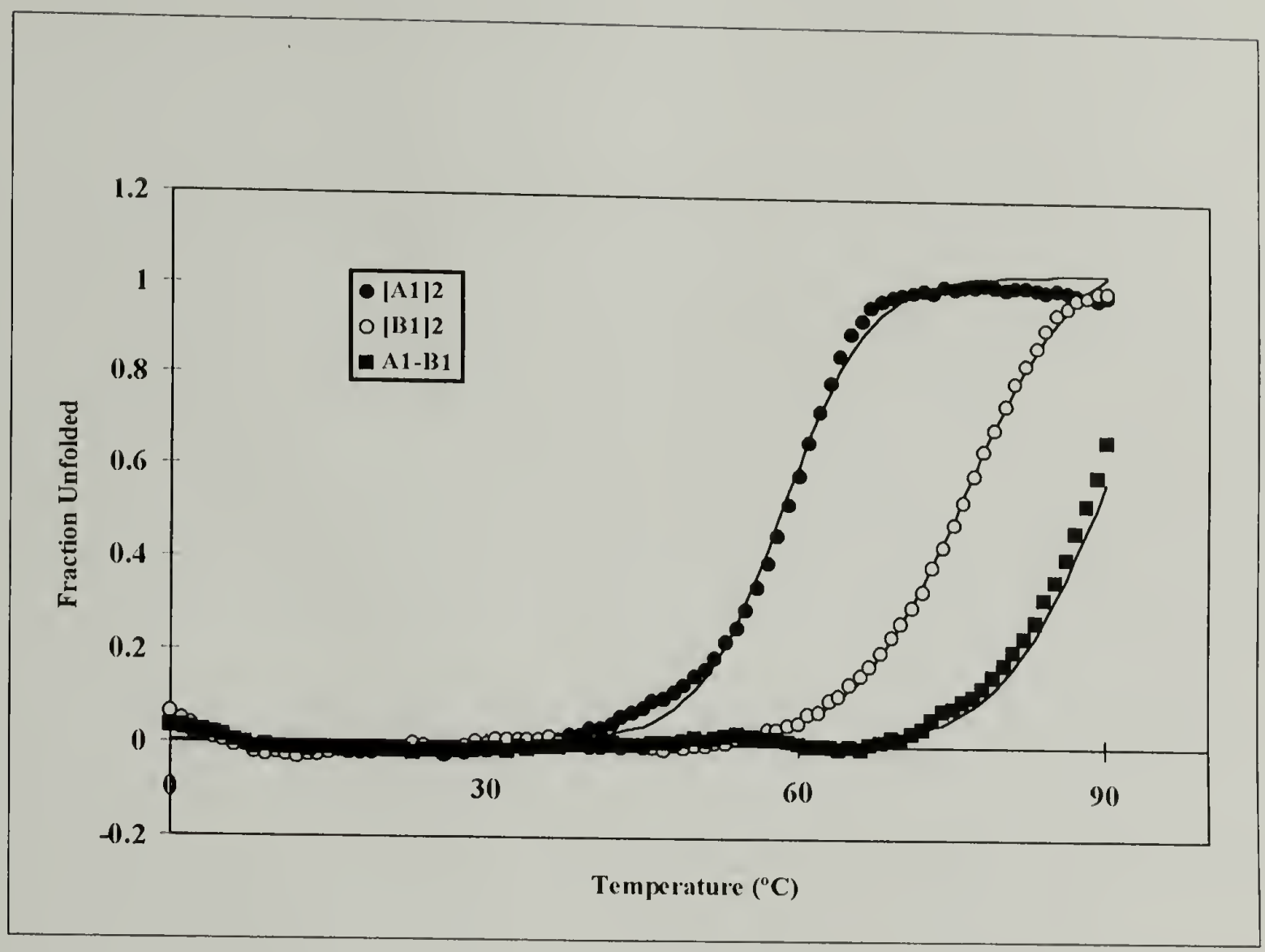


Figure 1.6 Thermal melting profiles of A1-A1, B1-B1, A1-B1 dimers. The fraction of unfolded peptide was determined by monitoring CD spectra at 222 nm (3 μM protein in 50 mM NaH_2PO_4, 100 mM NaCl, pH 7.4).

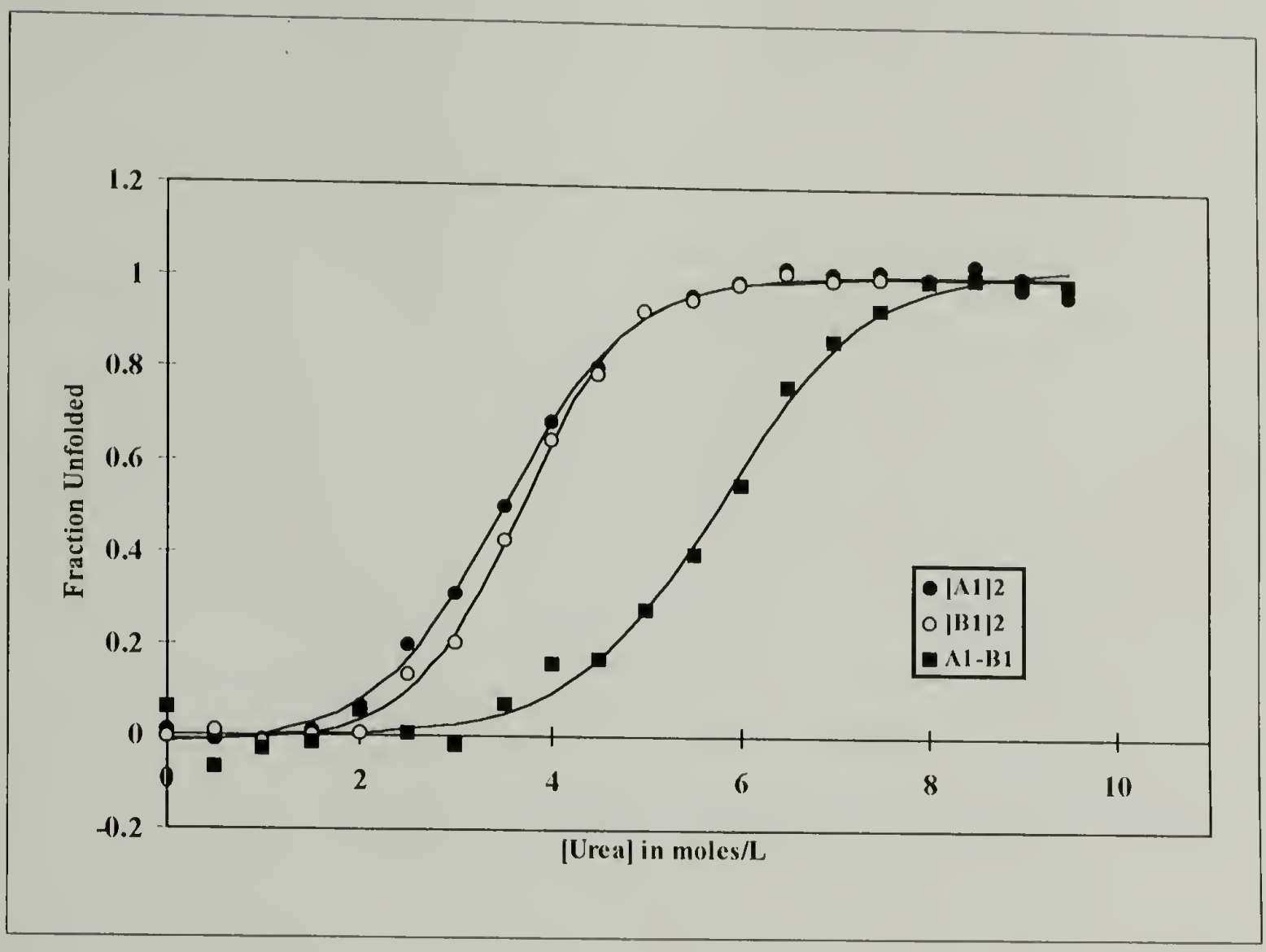


Figure 1.7 Urea denaturation profiles of A1-A1, B1-B1, and A1-B1 dimers. The fraction of unfolded peptide was determined by monitoring CD spectra at 222 nm as a function of [urea] (0-9.5 M) (3 mM protein concentration in 50 mM NaH_2PO_4, 100 mM NaCl, pH 7.4).

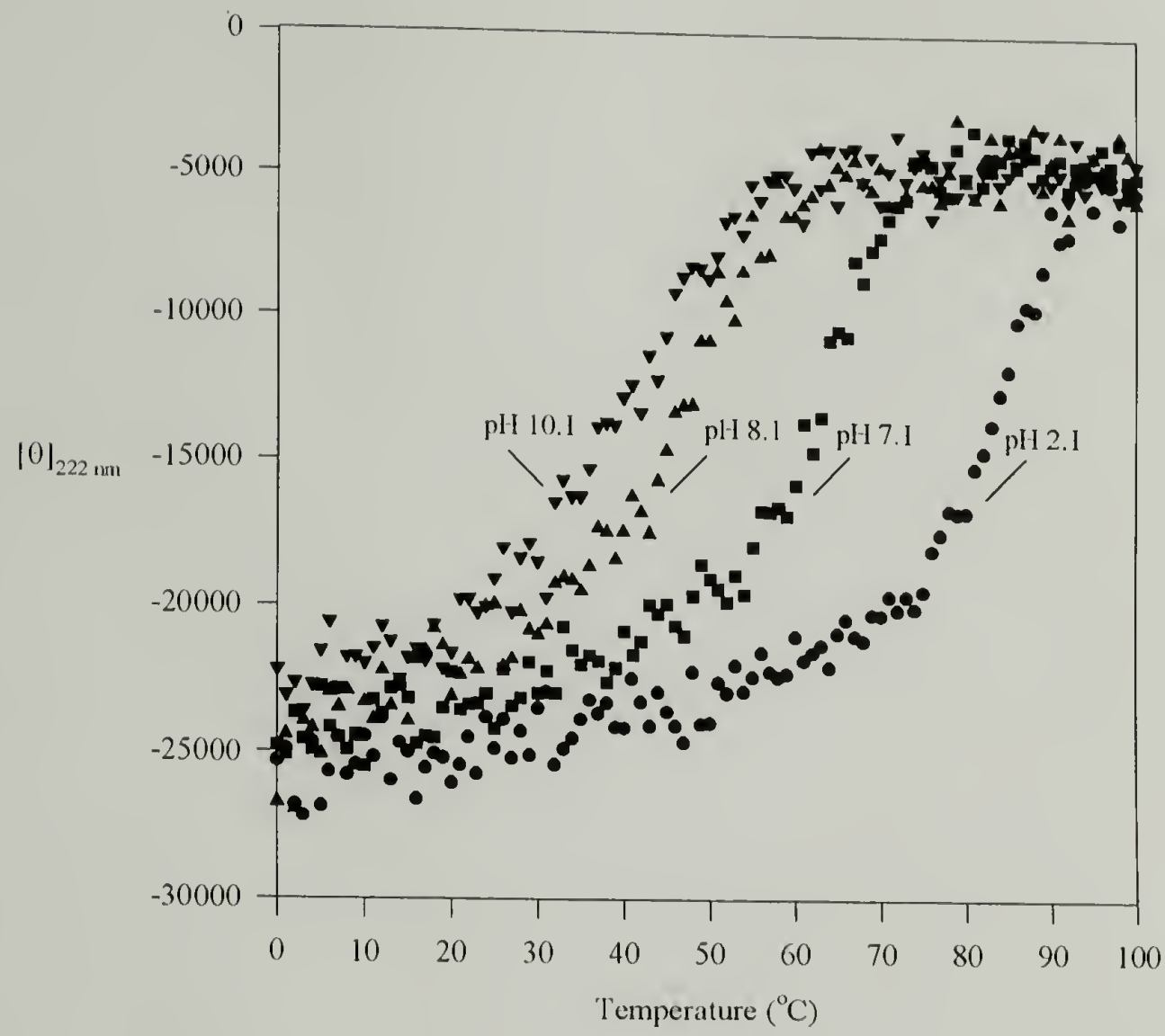


Figure 1.8 Thermal denaturation curves recorded at 222 nm for A1 (12 μM protein in 10 mM NaH_2PO_4) at pH 2.1, 7.1, 8.1, and 10.1.

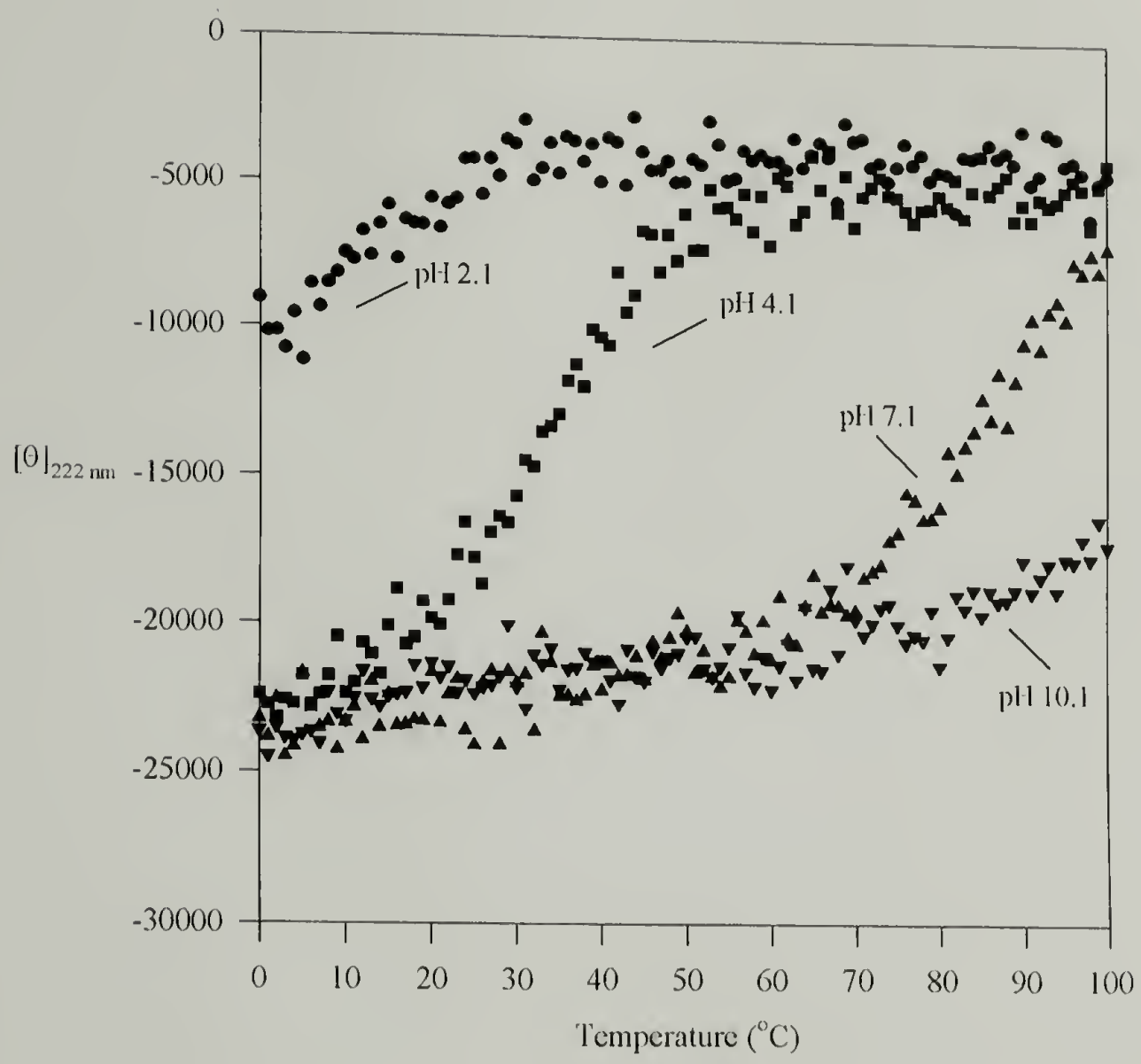


Figure 1.9 Thermal denaturation curves recorded at 222 nm for B1 (12 μM protein in 10 mM NaH_2PO_4) at pH 2.1, 4.1, 7.1, and 10.1.

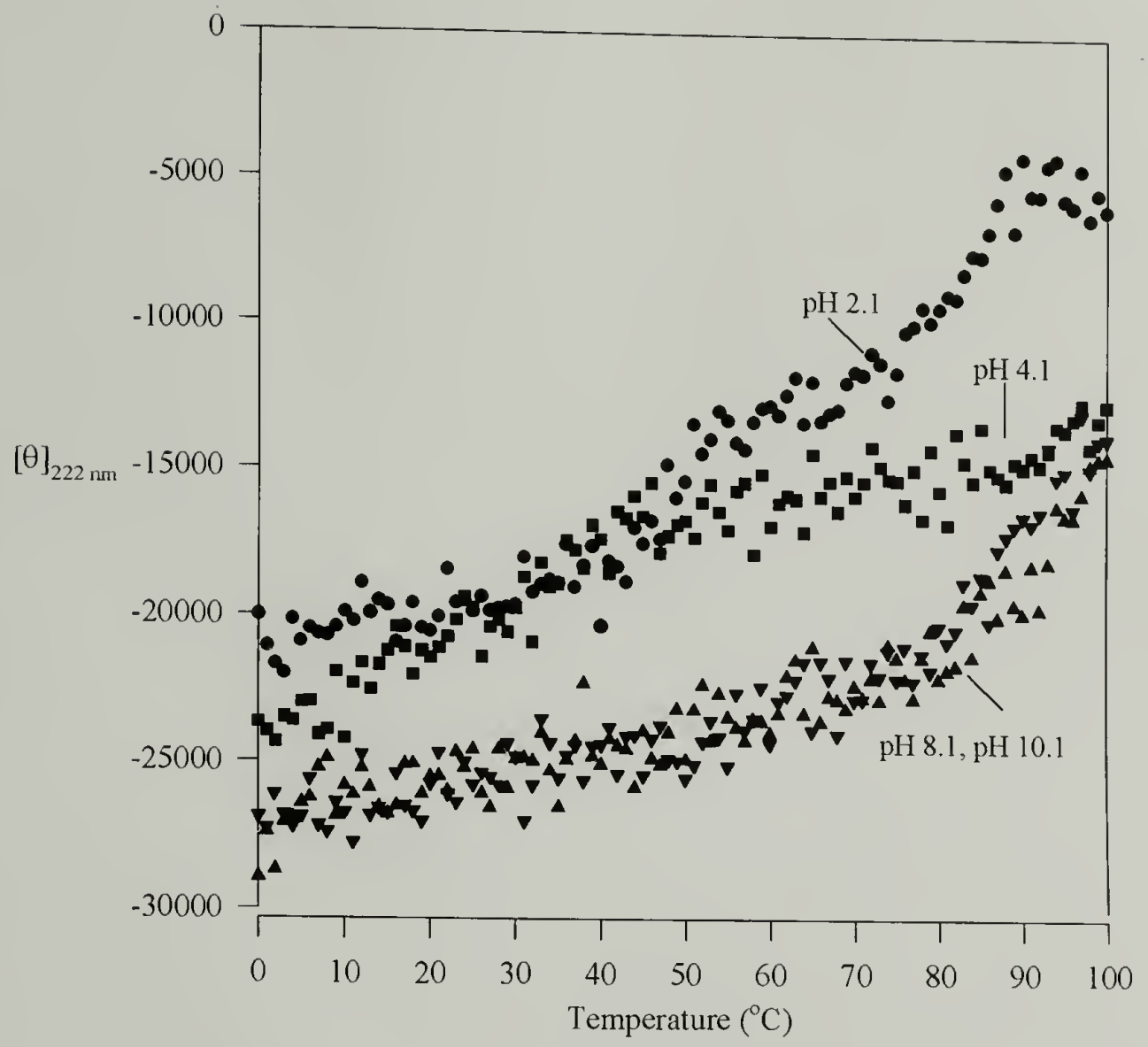


Figure 1.10 Thermal denaturation curves recorded at 222 nm for equimolar A1-B1 (12 μM protein in 10 mM NaH_2PO_4) at pH 2.1, 4.1, 8.1, and 10.1.

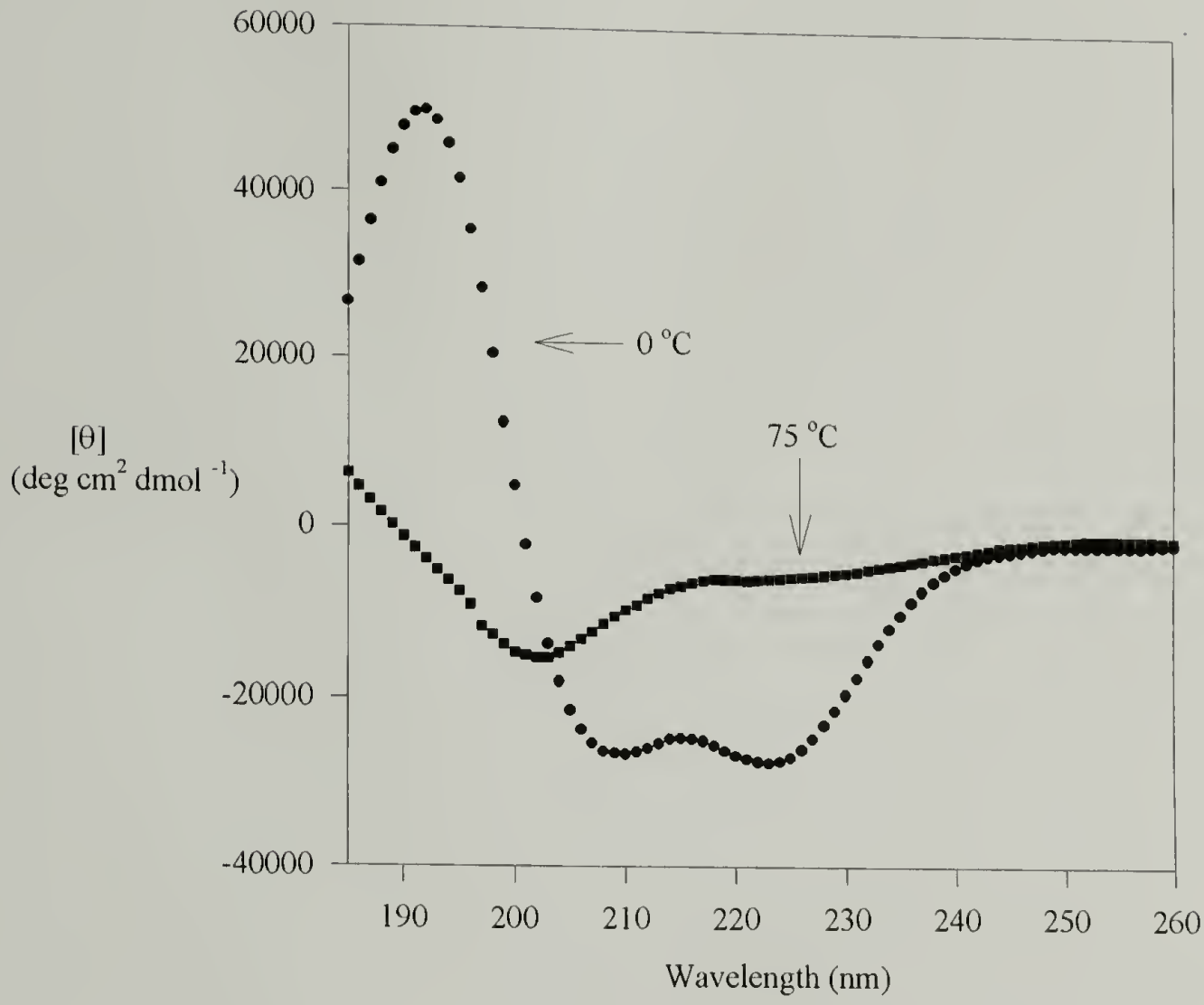


Figure 1.11 Molar ellipticity, [θ], as a function of wavelength recorded for A1 (12 μM protein in 10 mM NaH_2PO_4, pH 7.2) at temperatures of 0 and 75 °C.

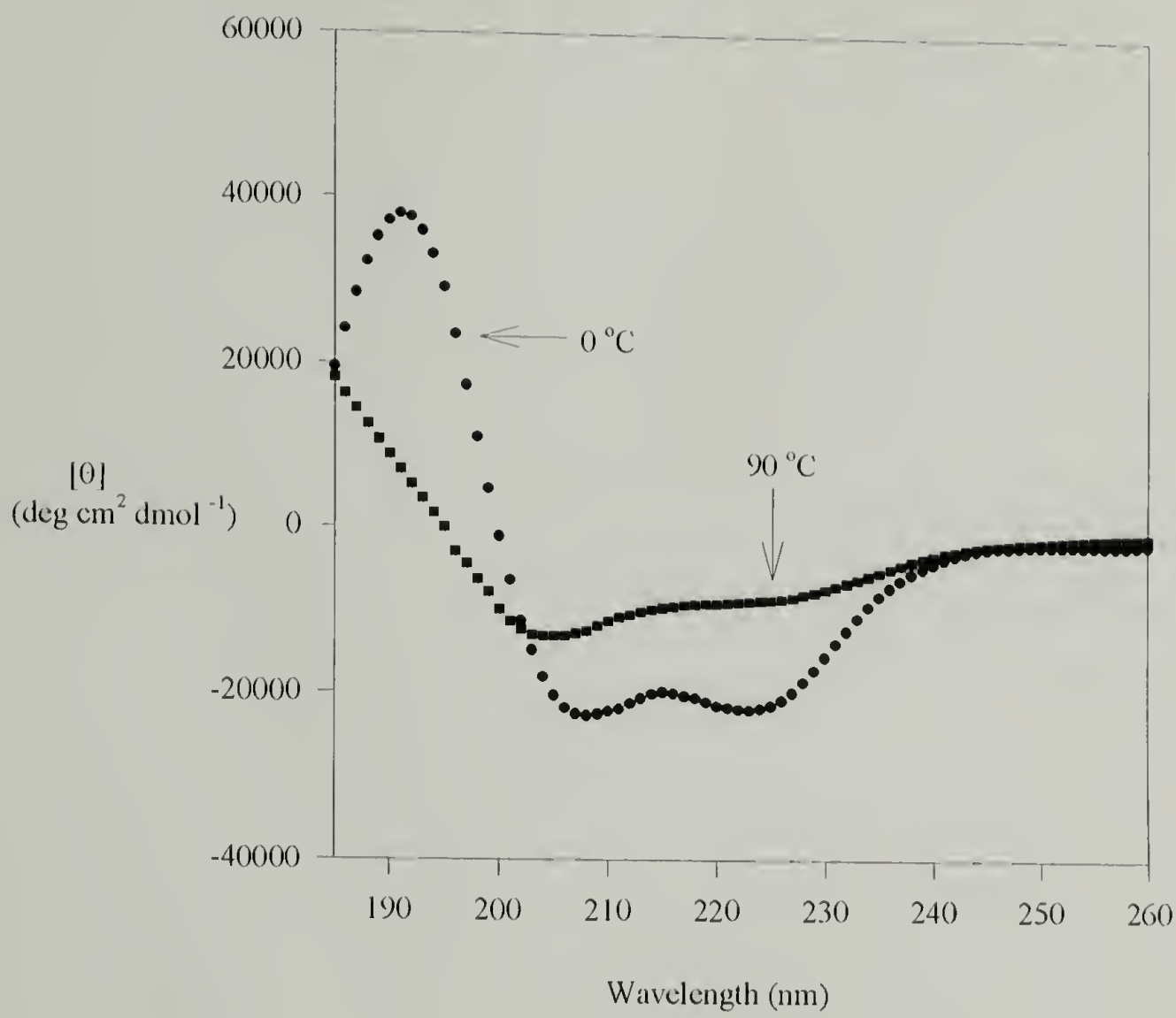


Figure 1.12 Molar ellipticity, [θ], as a function of wavelength recorded for B1 (12 μM protein in 10 mM NaH_2PO_4, pH 7.2) at temperatures of 0 and 90 °C.

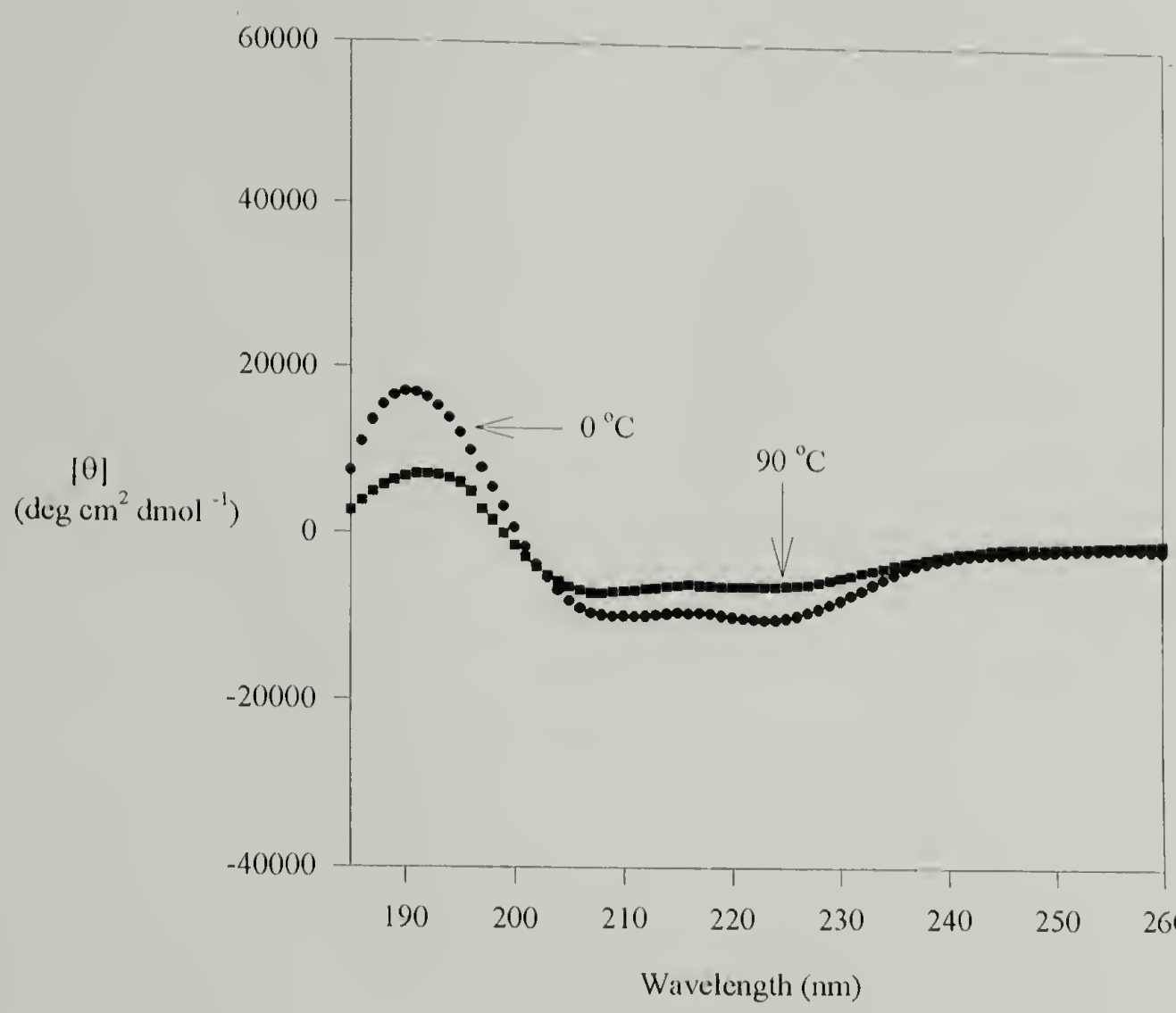


Figure 1.13 Molar ellipticity, [θ], as a function of wavelength recorded for equimolar A1-B1 (12 μM protein in 10 mM NaH_2PO_4, pH 7.2) at temperatures of 0 and 90 °C.

1.4 References

(1) Asquith, R. S. *Chemistry of Natural Protein Fibers*; Plenum Press: New York, 1977.

(2) Chum, H. L. *Polymers from Biobased Materials*; Noyles Data Corporation: New Jersey, 1991.

(3) Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198.

(4) Creighton, T. E. *Protein Folding*; W.H. Freeman and Co.: New York, 1992.

(5) Jirgensons, B. *Optical Activity of Proteins and Other Macromolecules*; Springer-Verlag: New York, 1973.

(6) Walton, A. G. *Polypeptides and Protein Structure*; Elsevier: New York, 1981.

(7) Ramachandran, G. N. *Advan. Protein Chem.* **1968**, *23*, 283.

(8) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.

(9) Creel, H. *Ph.D. Dissertation* **1994**, University of Massachusetts, Amherst.

(10) Lehn, J.-M. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1304.

(11) St. Pourcain, C. G.; Griffin, A. C. *Macromolecules* **1995**, *28*, 4116.

(12) Hilger, C.; Drager, M.; Stadler, R. *Macromolecules* **1992**, *25*, 2498.

(13) Guenet, J.-M. *Thermoreversible Gelation of Polymers and Biopolymers*; Academic Press Inc.: San Deigo, 1992.

(14) Stockmayer, W. H. *Macromolecules* **1991**, *24*, 6367.

(15) Astbury, W. T.; Haggith, J. W. *Biochim. Biophys. Acta* **1953**, *10*, 483.

(16) Pauling, L.; Corey, R. B. *Nature* **1953**, *171*, 59.

(17) Crick, F. H. C. *Nature* **1952**, *170*, 882-883.

(18) Crick, F. H. C. *Acta Cryst.* **1953**, *6*, 689-697.

(19) MacArthur, I. *Nature* **1943**, *152*, 38.

(20) Perutz, M. F. *Nature* **1951**, *167*, 1053.

(21) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

(22) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.

(23) König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139.

(24) Glover, J. N. M.; Harrison, S. C. *Nature* **1995**, *373*, 257.

(25) Hodges, R. S.; Sodek, J.; Smillie, L. B.; Jurasek, L. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *37*, 299.

(26) Sodek, J.; Hodges, R. S.; Smillie, L. B.; Jurasek, L. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 3800.

(27) Hodges, R. S.; Saund, A. K.; Chong, P. C. S.; St.-Pierre, S. A.; Reid, R. E. *J. Biol. Chem.* **1981**, *256*, 1214.

(28) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759.

(29) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538.

(30) Graddis, T. J.; Myszka, D. G.; Chaiken, I. M. *Biochemistry* **1993**, *32*, 12664.

(31) Monera, O. D.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1993**, *268*, 19218.

(32) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.

(33) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. *Nature* **1981**, *289*, 366.

(34) Sorger, P. K.; Nelson, H. C. M. *Cell* **1989**, *59*, 807.

(35) Weis, W. I.; Brunger, A. T.; Skehel, J. J.; Wiley, D. C. *J. Mol. Biol.* **1990**, *212*, 737.

(36) Conway, J. F.; Parry, D. A. D. *Int. J. Biol. Macromol.* **1991**, *13*, 14.

(37) Peteranderl, R.; Nelson, H. C. M. *Biochemistry* **1992**, *31*, 12272.

(38) Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F.; Eisenberg, D. *Science* **1993**, *259*, 1288.

(39) Weber, P. C.; Salemme, F. R. *Nature* **1980**, *287*, 82.

(40) Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, *196*, 657.

(41) DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622.

(42) Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *371*, 80.

(43) Monera, O. D.; Zhou, N. E.; Lavigne, P.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1996**, *271*, 3995.

(44) McGrath, K. P.; Kaplan, D. L. *Mat. Res. Soc. Symp. Proc.* **1993**, *292*, 83.

(45) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Cell* **1992**, *68*, 699.

(46) Hodges, R. S. *Biochem. Cell Biol.* **1995**, *74*, 133.

(47) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Royal Society of Chemistry: London, 1992.

(48) Laue, T. M. *Meth. Enzymol.* **1995**, *259*, 427.

(49) O'Shea, E. K.; Rutkowski, R.; Stafford, W. F.; Kim, P. S. *Science* **1989**, *245*, 646.

(50) Myszka, D. G.; Chaiken, I. M. *Biochemistry* **1994**, *33*, 2363.

(51) Chen, Y.-H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350.

(52) McGrath, K. P.; Butler, M. M.; DiGirolamo, C. M.; Petka, W. A.; Kaplan, D. L.; Laue, T. M. **1997**, *submitted for publication.*

(53) Pace, C. N. *Meth. Enzymol.* **1986**, *131*, 266.

(54) Green, S. M.; Meeker, A. K.; Shortle, D. *Biochemistry* **1992**, *31*, 5717.

(55) Sali, D.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1991**, *220*, 779.

(56) Monera, O. D.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1994**, *3*, 1984.

(57) Lumb, K. J.; Kim, P. S. *Science* **1995**, *268*, 436.

CHAPTER 2

SYNTHESIS OF GENETICALLY ENGINEERED MONODISPERSE BLOCK COPOLYMERS WITH HELIX-COIL-HELIX DOMAINS

2.1 Introduction and Objectives

Genetic engineering principles were used to create a new class of artificial proteins that combine the unique properties of coiled-coil leucine zipper proteins with those of a flexible water soluble random coil protein, $[(AG)_3PEG]_n$ (n = 10 and 28). Triblock proteins having α-helical end blocks separated by random coil domains are being synthesized for potential use as reversible three-dimensional networks where physical aggregates of two or more helical chains form the junction points and the random coil domains control swelling of the gel. Reported herein are the details of the genetic design and DNA cloning strategy for assembling the helix-coil-helix domains as well as the corresponding bacterial syntheses and characterization data of these macromolecules.

2.1.1 Design of End-Associated Polymers

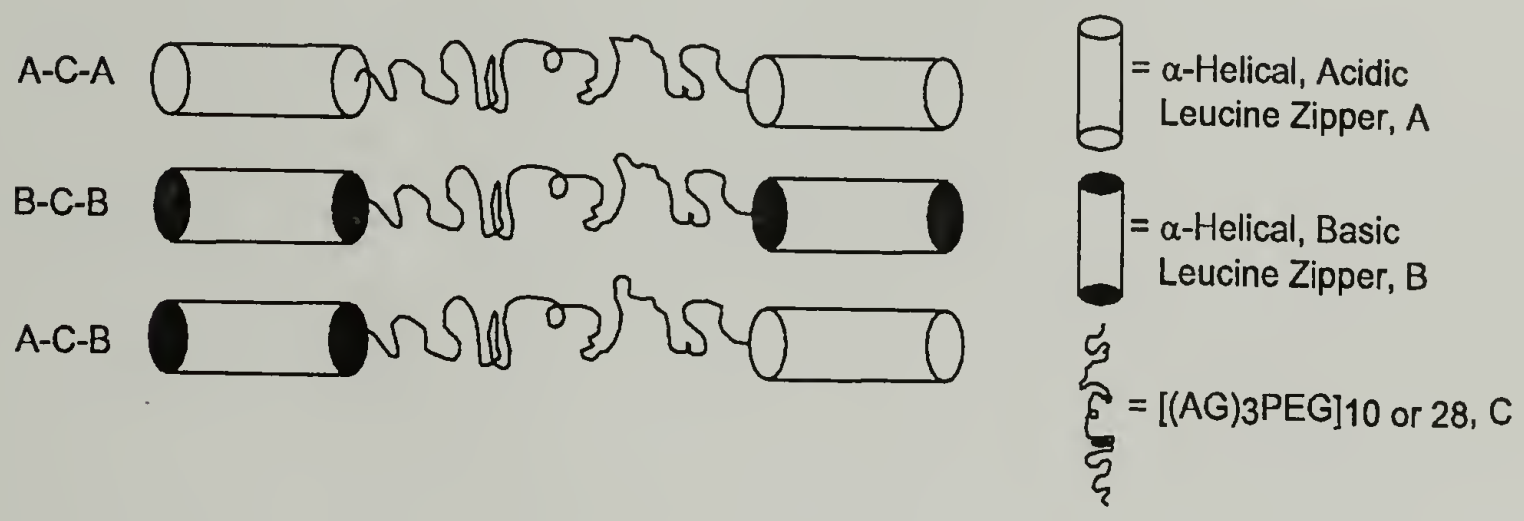
Protein engineering of polymers provides an efficient and controlled synthetic route to the production of *de novo* protein-based materials that have chemical and mechanical properties that are similar to or exceed those of natural materials. Our research interests have involved the synthesis of polymers with precisely controlled

architectures (1, 2) and with unusual chemical and biological function (3, 4). It is our goal to design and create entirely new polymeric systems where precise molecular weight, stereoregularity, sequence, and function are concurrently controlled. To this end, monodisperse triblock copolymers with helix-coil-helix domains were designed and synthesized by genetic engineering principles for the study of reversible physical gelation processes. Both structural and functional elements are being combined through α-helical ends, coiled-coils, and random coil domains, $[(AG)_3PEG]_n$ (n = 10, 28).

The genetic strategy for the synthesis of materials that contain naturally occurring helical protein ends (designated A and B) separated by an artificial random coil protein (C) is described herein. The end blocks are representative of a class of coiled-coil proteins called leucine zippers that are known to associate through interchain hydrophobic and electrostatic interactions (5). Association of coiled-coil leucine zipper proteins can produce dimers, trimers, tetramers, and higher order aggregates, depending on the details of the sequence (6).

The α-helix in the triblock design considered here is a 42 amino acid leucine zipper protein that forms stable water soluble coiled-coils (7). Leucine zipper proteins have a characteristic seven amino acid repeat, $(a\ b\ c\ d\ e\ f\ g)_n$, where the *a* and *d* positions are occupied by hydrophobic amino acids and the *e* and *g* positions are populated by polar acidic or basic amino acids. A hydrophobic interface is formed from the side-to-side packing of the *a* and *d* residues that are aligned on the complementary faces of two helices (8). The *a* and *d* positions of the helices in this work were modeled after those amino acids found in the *Jun* oncogene product, with leucines at every seventh position. Two different helices were designed such that the charge patterns at

the *e* and *g* positions are largely acidic (A) or largely basic (B). Both of these peptide domains have the propensity to form stable interchain homodimers in solution, and when mixed in equimolar amounts, stable heterodimers (9). The three types of triblock copolymers that were successfully prepared by biosynthetic methods are given below.



2.2 Experimental Section

2.2.1 Materials and Methods

Synthesis and Purification of Single Stranded DNA.

Oligonucleotides were synthesized on a Biosearch Model 8700 DNA synthesizer by phosphoramidite chemistry developed by McBride and Caruthers (10) and cleaved from polymer supports with 30% NH_4OH at 65 °C. The glass beads were removed by centrifugation and the decanted supernatant was dried in a speed vacuum. The remaining pellets were then resuspended in 1 mL of deionized distilled water (ddH_2O)

and centrifuged at 15,800*g* (25 °C) for 10 min to remove any insoluble materials. The remaining aqueous solution of DNA was adjusted to a final concentration of 10 mM $MgCl_2$. Ethanol (100% equilibrated at 25 °C) was added to the DNA solution in the ratio of 3:1. The mixtures were then cooled at -70 °C for 10 min and pelleted at 4 °C (15,800*g*) for 10 min. The supernatant was decanted and the pellet dried in the speed vacuum. The resulting pellets were resuspended in 1 mL of ddH_2O and single stranded DNA was quantitated based on optical absorbance measurements at 260 nm. Polyacrylamide gel electrophoresis (10%, 8 M urea) was used to purify the single stranded DNA. Coding and noncoding strands (100 μg) of Linkers 1 (**1**) and 2 (**2**) shown in Figure 2.1 were mixed with 2x formamide loading buffer and heated to 95 °C for 5 min before loading onto the preheated gel. These samples were electrophoresed at constant voltage (350 V) for approximately 2 hrs. Purified bands of the DNA strands were visualized by ethidium bromide and excised from the gel. Gel slices were crushed in 1 mL of elution buffer (500 mM NH_4OAc, 0.1% sodium dodecyl sulfate (SDS), 10 mM $MgOAc_2$, and 1 mM ethylenediaminetetraacetic acid (EDTA), dipotassium salt dihydrate (pH 8.0)) and incubated overnight at 37 °C on a spin wheel. Gel residue was removed by centrifugation at 4 °C in filter vials and the supernatants were first extracted with 1-butanol to remove ethidium bromide and then treated with 10 μL of Type II oyster glycogen (10 μg/mL), NaCl (250 mM), and 0.9 mL 100% EtOH. Mixtures were precipitated at -20°C overnight and centrifuged at 4°C for 30 min. DNA pellets were washed once with 500 μL of cold EtOH (70%) and redissolved in 50 μL of ddH_2O.

Phosphorylation and Annealing of Single Stranded DNA.

Purified oligonucleotides (**1**: coding = 5 μg, noncoding = 5 μg and **2**: coding = 5 μg, noncoding = 5 μg) were suspended in 20-30 μL of ddH_2O, 5 μL of 10x T4 polynucleotide kinase buffer (80 mM Tris-HCl pH 7.5, 20 mM dithiothreitol (DTT), 10 mM $MgCl_2$, and 1 mM ATP), 2 μL of ATP (100 mM), and 1 μL of T4 polynucleotide kinase (10 Richardson units (11)), to make a final volume of 50 μL. These tubes were then incubated at 37°C for 45 min and 1 μL of NaCl (5 M) was added to each vial to make a final concentration of 100 mM. The mixtures were then placed in a 95 °C water bath and allowed to equilibrate to room temperature over a 10 hr period. The annealed duplexes were then extracted with 50:50 phenol:chloroform and washed once with 100% chloroform. Isopropanol (100 μL) and NaOAc (20 μL, 3 M) were added to each vial and these were placed in the -70 °C freezer for 30 min. DNA duplexes were pelleted at 4°C (15,800*g*) for 25 min and dried in the speed vacuum. Pellets were resuspended in 100 μL of ddH_2O.

*Ligation of **1** and **2** Duplexes into pUC-18.*

Pharmacia™ pUC-18 cloning vector (2.5 μg) (12) was digested with EcoRI (2 units; 1 unit of restriction endonuclease activity is the amount of enzyme required to completely digest 1 μg of substrate DNA in a total reaction volume of 50 μL) and HindIII (2 units) restriction enzymes. DNA was visualized on a 1% agarose gel with ethidium bromide and the linearized pUC-18 band was cut and purified by *Bio-Rad*™ gel extraction kit. A 33 molar excess of linker to vector (equal volumes) was combined with

2 μL of 10x ligase buffer (50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 μg/mL bovine serum albumin (BSA)), 0.5 μL T4 ligase (15 Weiss units (13), and 7.6 μL of water to make a total volume of 20 μL. The mixtures were placed in a 12-15 °C refrigerator for 21 hrs.

Transformation of pUC-18 Containing 1 and 2 into DH5αF' Competent Cells.

DH5αF' cells were grown to an optical density (OD_{550}) of 0.573, placed on ice for 10 min, and collected by centrifugation at 2000*g* for 15 min at 4 °C. The cells were then resuspended in 2 mL of TFB1 Buffer (10 mM morpholino ethanesulfonic acid (pH 6.2) (MES), 100 mM $RbCl_2$, 10 mM $CaCl_2{\cdot}2H_2O$, and 50 mM $MnCl_2{\cdot}4H_2O$) by gentle tilting of the tube with the final volume being adjusted to 16 mL. This suspension was left on ice for 15 min before the cells were pelleted at 2000*g* for 15 min and resuspended in TFB2 buffer (10 mM 3-(N-morpholino)propane sulfonic acid (MOPS), 75 mM $CaCl_2{\cdot}2H_2O$, 10 mM $RbCl_2$, and 15% glycerol). The tube was left on ice for 15 min before 200 μL aliquots were dispensed into vials to be stored at -70 °C. Competent cells were then thawed on ice and 50 μL of the cells were added to 10 μL of the ligation mixtures. The vials were then placed on ice for 2 hrs and heat shocked at 42 °C for 3 min before placing them on ice for 5 min. 2xYT medium (500 μL of 16 g Casein-hydrolysate, 10 g yeast extract, 5 g NaCl per liter) was added to the mixtures and incubated at 37 °C for 45 min. The cells were then spread onto agar plates containing 200 μg/ml ampicillin and grown for 17 hrs.

*Mini-preps of pUC-18 Containing **1** and **2** Transformants.*

pUC-18 cloning plasmids containing DNA **1** and **2** were designated as pWAP-L1 and pWAP-L2 respectively (see Figure 2.2). Eight colonies of pWAP-L1 and twelve colonies of pWAP-L2 were grown overnight in 2xYT medium containing ampicillin (200 μg/mL). The DNA from these cell cultures was purified by the following miniprep protocol. A 1.5 mL saturated culture was spun at 13,600*g* for 2 min and the supernatant poured off. The cells were resuspended in 200 μL of GTE buffer (50 mM glucose, 25 mM Tris-HCl (pH 8), and 1 mM EDTA (pH 8.0)) by repeated pipetting. NaOAc (200 μL, 3 M, pH 4.8) was added to neutralize the solution and the mixture was incubated on ice for 5 min. The white precipitate was collected by centrifugation at 13,600*g* for 5 min. The clarified supernatant was transferred to a new 1.5 mL tube containing 2 μL of RNAse A (10 mg/mL) and incubated at 50 °C for 30 min. Phenol:chloroform (50:50, 500 μL) was added and the mixture was spun at 13,600*g* for 5 min. The aqueous layer was decanted and washed once with chloroform. Isopropyl alcohol (500 μL) was added. DNA was precipitated at -20 °C for 30 min and pelleted at 4 °C, 13,600*g* for 30 min. Supernatant was discarded and the DNA was resuspended into 100 μL of ddH_2O before 50 μL of 25% polyethylene glycol (PEG M.Wt. 8000)/2.5 M NaCl was added. This mixture was incubated at -20°C for 30 min, centrifuged at 13,600*g* for 30 min, and resuspended in 50 μL of ddH_2O.

*Screening for Inserts **1** and **2** by Enzymatic Digestions.*

To check for proper DNA insertion of **1** and **2** into pUC-18, restriction digests of pWAP-L1 and pWAP-L2 were done by mixing 15 μL of plasmid DNAs collected from the miniprep above, 5 μL of NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), 1 μL HindIII (20 units), 1 μL NheI (5 units), and 33 μL of ddH_2O to give a final volume of 50 μL. These mixtures were then heated at 37 °C for 4 hrs. The digested DNA was visualized by ethidium bromide on a 1% agarose gel and colonies with the proper inserts were saved for DNA sequencing. In addition to this digestion, individual digestions for pWAP-L1 and pWAP-L2 were done as further checks for insertion. pWAP-L1 (7 μL) was digested with 0.5 μL of NruI (4 units), 2.5 μL NEBuffer NruI (100 mM KCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, pH 7.7), and 15 μL of ddH_2O for a final volume of 25 μL. pWAP-L2 was digested with 0.5 μL of XhoI (10 units), 2.5 μL of NEB2 buffer, and 15 μL of ddH_2O. Again digested DNA was visualized on a 1% agarose gel with ethidium bromide. pUC-18 was used as a positive control since it contains neither NruI nor XhoI restriction sites.

DNA Sequencing of pWAP-L1 and pWAP-L2 Inserted into pUC-18 Cloning Vector.

DNA sequences pWAP-L1 and pWAP-L2 were confirmed by the Sanger (14) dideoxy method using *USB Sequenase Version 2.0.1*™ sequencing protocol (15). Both universal (M13/pUC Sequencing Primer (-20)17mer) and reverse (M13/pUC Sequencing Primer (-47)24mer) were used to read coding anticoding DNA strands respectively.

Digestion and Removal of Phosphate Groups of pWAP-L1 and pWAP-L2.

pWAP-L1 (88 μL or approximately 220 ng) was first digested with 2.0 μL of NruI (50.8 units) and 10 μL of NruI buffer overnight at 37 °C. DNA was precipitated and isolated before being digested with 1.0 μL of SphI enzyme (5 units), 5 μL of NEB2 buffer, and 44 μL of ddH_2O over 5 hrs at 37 °C. pWAP-L2 (88 μL or approximately 1 μg) was digested with 2.0 μL BstEII (20 units) and 10 μL NEB3 buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DDT, pH 7.9) at 60 °C overnight. Both pWAP-L1 and pWAP-L2 mixtures were heat inactivated for 20 min at 65 °C before 1 μL of calf intestinal alkaline phosphatase (CIP) (10 units) was added to the pWAP-L1 digestion mixture and 8 μL of CIP (80 units) was added to the pWAP-L2 digestion mixture. EDTA was added to each solution to give a final concentration of 5 mM and the reaction mixtures were heated to 65 °C for 1 hr. The DNA was extracted with phenol and precipitated with 20 μL of NaOAc and 120 μL of 100% EtOH. Dried DNA was then resuspended in 40 μL of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) before being run on a 1% agarose gel. Gel slices were excised and purified by *Bio-Rad* DNA purification method.

Insertion and Verification of Random Coil and Helical DNA Sequences into Linearized pWAP-L1 and pWAP-L2.

DNA encoding $[(AG)_3PEG]_{10}$ (**3**) (Figure 2.3) and $[(AG)_3PEG]_{28}$ (**4**) (Figure 2.4) were ligated into linearized pWAP-L1 at SphI/NruI sites. In addition, DNA encoding α-helical acidic leucine zipper (**5**) (Figure 2.5a) and basic leucine zipper (**6**) (Figure 2.5b) were ligated into linearized pWAP-L2 at the BstEII site. Prior to ligation, DNA **3** and **4**

were obtained from pET3-5 and pET3-14 plasmids (16) respectively by cutting with SphI and NruI. Likewise, DNA **5** and **6** were digested from pUC-LINKA1 (7) with BstEII enzyme. Purified pieces were ligated into the linear pWAP-L1 or pWAP-L2 by combining 10 μL of insert, 10 μL of vector, 2.5 μL of ligase buffer, 0.5 μL of T4 DNA ligase, and 2.0 μL of ddH_2O. The ligation mixtures were then used to transform DH5αF' cells and the resultant transformants were screened for proper DNA insertion by enzymatic digestions. Plasmids containing **3** or **4** were digested with NdeI and SphI whereas plasmids containing **5** or **6** were digested with NdeI and BglII. All of the combined DNA were confirmed by the Sanger dideoxy sequencing method mentioned previously. Plasmids were designated pWAP-L1C_{10}, pWAP-L1C_{28}, pWAP-L2A, and pWAP-L2B as shown in Figure 2.2.

Enzymatic Digestions of DNA at NheI and SpeI Restriction Sites.

In order to recombine DNA described above to form DNA that encodes the triblocks A-C-A, A-C-B, and B-C-B where A and B are acidic and basic leucine zipper proteins and C is the random coil $[(AG)_3PEG]_{10 \text{ or } 28}$ repeat, a series of digestions and ligations were done at the NheI and SpeI enzymatic restriction sites. Figures 2.6 and 2.7 represent two different routes to the coding sequences of the triblocks. The combination of DNA was achieved through NheI and SpeI sites located outside of the helix and coil regions of the DNA. It is important to note that Nhe and SpeI sites are compatible and upon ligation of a NheI to a SpeI end or vice versa, the site is destroyed. Both restriction sites have the same middle four nucleotide bases which allows them to combine upon ligation (see below).

NheI	5' G C T A G C 3' 3' C G A T C G 5'
SpeI	5' A C T A G T 3' 3' T G A T C A 5'

The frequency at which NheI cut ends combine with SpeI cut ends is approximately 50%. In all cases, NheI and SpeI digestions were done in a total volume of 25 μL; enzymatic buffer, enzyme, ddH_2O, and DNA volumes were adjusted accordingly. Miniprep, transformation, and phosphate removal procedures were described previously. Agarose gels (2%) were used in verification, quantification, and isolation of the various DNA.

DNA encoding twelve different proteins was made through the series of steps outlined. These DNA are abbreviated as pWAP-L2A, pWAP-L2B, pWAP-L1C_{10}, pWAP-L1AC_{10}A, pWAP-L1AC_{10}B, pWAP-L1BC_{10}B, pWAP-L2AC_{10}A, pWAP-L2AC_{10}B, pWAP-L2BC_{10}B, pWAP-L2AC_{10}A, pWAP-L2AC_{10}B, pWAP-L2BC_{10}B, pWAP-L2AC_{28}A, pWAP-L2AC_{28}B, and pWAP-L2BC_{28}B. The difference between L1 and L2 designation is L1 encodes for a single tryptophan at the C-terminus whereas L2 encodes for a single cysteine at this end.

BamHI Digestions of pWAP-L2A, pWAP-L2B, pWAP-L1C_{10}, pWAP-L1AC_{10}A, pWAP-L1AC_{10}B, pWAP-L1BC_{10}B, pWAP-L2AC_{10}A, pWAP-L2AC_{10}B, pWAP-L2BC_{10}B, pWAP-L2AC_{28}A, pWAP-L2AC_{28}B, pWAP-L2BC_{28}B, and pQE9.

pWAP-L2A, pWAP-L2B, pWAP-L1C_{10}, pWAP-L1AC_{10}A, pWAP-L1AC_{10}B, pWAP-L1BC_{10}B, pWAP-L2AC_{10}A, pWAP-L2AC_{10}B, pWAP-L2BC_{10}B, pWAP-L2AC_{28}A, pWAP-L2AC_{28}B, and pWAP-L2BC_{28}B DNA were grown in 100 mL 2xYT

medium with ampicillin (200 μg/mL) and isolated by *Qiagen*™ maxiprep DNA columns. Each plasmid (43 μL or approximately 300 ng) was combined with 5 μL of NEB BamHI buffer (150 mM NaCl, 10 mM $MgCl_2$, 1 mM DDT, pH 7.9) and 2 μL of BamHI (40 units) and incubated overnight at 37 °C (Figures 2.8 and 2.9). *Qiagen*™ pQE9 expression vector (Chatsworth, CA) DNA was also cut with BamHI as above and dephosphorylated by the CIP protocol as previously described (Figures 2.8 and 2.9). The fragments were visualized on 2% agarose gels and excised for purification by *Bio-Rad*™ DNA binding beads. Quantification of the recovered DNA was made on 2% agarose gels.

Ligation of pWAP-L2A, pWAP-L2B, pWAP-L1C$_{10}$, pWAP-L1AC$_{10}$A, pWAP-L1AC$_{10}$B, pWAP-L1BC$_{10}$B, pWAP-L2AC$_{10}$A, pWAP-L2AC$_{10}$B, pWAP-L2BC$_{10}$B, pWAP-L2AC$_{28}$A, pWAP-L2AC$_{28}$B, and pWAP-L2BC$_{28}$B into pQE9.

A 5 molar excess of linker to vector was combined with 2.5 μL ligation buffer, 0.5 μL ligase, and ddH_2O for a total volume of 25 μL. The ligated mixtures were transformed into K-12 *E. coli* strain SG13009 (*Qiagen*™, Chatsworth, CA) containing repressor plasmid, pREP4 (*Qiagen*™, Chatsworth, CA). Mixtures were spread onto 2xYT agar plates containing 200 μg/mL ampicillin and 50 μg/mL kanamycin and incubated for 16 hrs at 37 °C. Single colonies were grown to saturation in 5 mL of 2xYT medium and the DNA isolated. Restriction digests were used to check orientation, size, and digest sites of the transformed colonies. These were accomplished with NheI, EcoRI/HindIII, and BamHI restriction enzymes respectively. Colonies containing the correct inserts were then used for the following bacterial expression.

Protein Expression in pQE9-derived Expression Vectors.

Recombinant vectors (pQE9) containing the targeted sequences were used for protein expression in host cells of strain SG13009 containing repressor plasmid (pREP4) (Figures 2.8-2.9). New plasmid designations are pQE9-L2A, pQE9-L2B, pQE9-L1C_{10}, pQE9-L1AC_{10}A, pQE9-L1AC_{10}B, pQE9-L1BC_{10}B, pQE9-L2AC_{10}A, pQE9-L2AC_{10}B, pQE9-L2BC_{10}B, pQE9-L2AC_{28}A, pQE9-L2AC_{28}B, and pQE9-L2BC_{28}B with the corresponding protein sequences designated **7-18** respectively. Saturated cultures containing the targeted DNA sequences were grown in 25 mL of 2xYT medium with ampicillin (200 μg/mL) and kanamycin (50 μg/mL) for 15 hrs at 37°C. These cultures were then transferred to 2 L flasks each containing 1 L sterilized 2xYT medium and antibiotics. These were incubated for an additional 6 hrs at 37°C with vigorous aeration until optical densities (600 nm, OD_{600}) were in excess of 2. IPTG (1 mM) was added and protein synthesis was induced for a period of 4 hours at 37°C. The cells were then centrifuged (22,100 x g for 30 min), supernatant removed, and 50 mL of 6 M guanidine-HCl buffer containing Na_2HPO_4 (0.1 M, pH 8) was added. The cells were lysed by storing in a -80 °C freezer before the supernatant was collected for protein purification by centrifugation at 22,100*g* for 50 min. The clarified supernatant was loaded onto a nickel-NTA affinity column at pH 8 and the bound target proteins were washed with 8 M urea buffers at pH 8, 7, and 6 before being eluted at pH 5. The appropriate fractions were dialyzed against deionized water before being freeze-dried. The final weights of the purified target proteins were in the range of 50-150 mg per 1 L of cell growth medium (Table 2.1).

Matrix Assisted Laser Desorption Mass Spectrometry with Time-of-Flight Analyzer (TOF).

Matrix assisted laser desorption mass spectrometry was used to analyze the mass of gas-phase protein ions. MALDI-TOF mass spectrometry was performed on a Lasermat 2000™ (*Finnigan Mat*, San Jose,CA) by the Analytical Chemistry and Peptide/DNA Synthesis Facility at Cornell University, Ithaca, NY. The protein samples (1-10 μM) were mixed with a molar excess of matrix solution (3,5-dimethoxy-4-hydroxycinnamic acid (0.05 M) for masses greater than 20,000 and α-cyano-4-hydroxycinnamic acid (0.05 M) for masses less than 20,000 in 30% acetonitrile/water, 0.1% trifluoroacetic acid) before approximately 0.5 μL was dried onto the sample probe. The surface of the probe was irradiated with a nitrogen pulsed laser at 337 nm. Detection of the protein ions was recorded in the positive ion mode with individual spectra being obtained with single laser shots of about 10^6 W/cm^2 irradiance. An accelerating voltage of 20 keV was used.

Quantitative Amino Acid Compositional Analysis.

Amino acid analysis was performed by the Analytical Chemistry and Peptide/DNA Synthesis Facility at Cornell University, Ithaca, NY. Amino acid analysis was carried out on a Pico-Tag Amino Acid Analysis System (17). Duplicate samples were hydrolyzed in constant boiling HCl at 150 °C for 110 min.. Appropriate blanks, controls and standards were hydrolyzed in the same vessel as the batch hydrolysis. The amino acids were derivatized with phenyl isothiocyanate (PITC) and analyzed by reversed phase HPLC (18, 19). The resulting phenylthiocarbamoyl amino acid

derivatives were separated on a 4.6 x 300 mm Nova Pack™ C18 column employing a modified Pico-Tag buffer system and detected at 254 nm.

Polyacrylamide Gel Electrophoresis.

Purified proteins were resolved on a polyacrylamide minigel by the method of Laemmli (20) and visualized by staining with Coomassie Brilliant Blue R-250.

2.3 Results and Discussion

2.3.1 Synthesis and Characterization

The cloning and expression strategies used to assemble DNA encoding the helix, coil, or helix-coil-helix motifs are shown in Figures 2.2 and 2.6-2.9. Two separate DNA, **1** and **2** (Figure 2.1), were designed and synthesized in order to manipulate DNA (**3**, **4**, **5**, and **6**) that encodes the helical or random coil domains (Figures 2.3-2.5). DNAs **1** and **2** were chemically synthesized and verified by sequencing of both coding and noncoding strands in pUC18 cloning plasmid. DNAs **3-6** were previously designed by McGrath *et al.* (7, 20) and verification of these DNA after insertion into the appropriate restriction sites was made by DNA sequencing. The combination of DNA encoding for helix and random coil was accomplished with NheI and SpeI restriction sites. By cutting with both restriction sites, ligating the appropriate DNA fragments, and verifying the orientation of the DNA, helix-coil-helix DNA were isolated by two pathways (Figures 2.6 and 2.7). DNA designated pWAP-L1AC_{10}A, pWAP-L1AC_{10}B, and pWAP-L1BC_{10}B

encodes for a single tryptophan at the C-terminus end (Figure 2.6). On the other hand, DNA designated pWAP-L2AC$_{10}$A, pWAP-L2AC$_{10}$B, pWAP-L2BC$_{10}$B, pWAP-L2AC$_{28}$A, pWAP-L2AC$_{28}$B, and pWAP-L2BC$_{28}$B encodes for a single cysteine at the C-terminus end (Figure 2.7). Finally, all DNA were cut at BamHI restriction sites, purified, and ligated into the single BamHI restriction site in the *Qiagen*™ pQE9 expression vector (Figures 2.8 and 2.9). In the pQE9 plasmid, a sequence of six histidines are placed at the N-terminus of the protein. This sequence can be used for rapid purification of the targeted proteins with Ni^{2+} metal affinity chromatography.

The molar masses of the recombinant proteins were verified by matrix-assisted laser desorption mass spectrometry (MALDI) (21) and purity was assayed by amino acid compositional analysis. The calculated and observed masses of the protonated polymers are shown in Figures 2.10-2.21 and summarized in Table 2.2. The expected and observed molecular masses vary for each protein. The precision of mass determination using RNaseA as the calibration is 0.1% (22). Counter ion association is believed to account for the higher observed masses of Proteins **9-18**. It should be mentioned that previous MALDI experiments done by Beavis *et al.* (23) on $[(AG)_4PEG]_{14}$ revealed two alanine-to-valine substitutions in the expressed proteins. It is possible that this could have occurred in the $[(AG)_3PEG]_n$ containing proteins. However, amino acid analysis does not support this statement and shows that all amino acid analyses were within 5% of the theoretical values (Tables 2.3-2.8). In addition to the $[M+H]^+$ peak, a set of periodic low molecular weight mass fragments is found for **9** (Figure 2.12) in the 5-10K molecular weight range. The mass differences between these small peaks are in the

range of 572 to 678 and they most likely correspond to the ionization of a single alanylglycine repeat unit, $[(AG)_3PEG]$, which has a molecular weight of 667.

Purified proteins were resolved on a polyacrylamide minigel by the method of Laemmli and visualized by staining with Coomassie Brilliant Blue R-250 (Figures 2.22 and 2.23). All of the proteins, with the exception of **7** and **8**, migrate to molecular weights higher than their calculated values. This is believed to result from the acidic $[(AG)_3PEG]_n$ protein for which an anomalously low rate of migration has been seen in previous work (17, 24). One explanation for this is that the charge-to-mass ratios of the SDS-polypeptide complexes are no longer uniform due to the acidic nature of the polypeptides. Normally SDS-bound polypeptides are extended and rodlike in shape whereas the proteins reported here may be more random coil in shape. The reduced binding of SDS to the polypeptides is a result of the acidic groups remaining ionized or negatively charged at pH 9.5, which is the pH of the gel while an external electrical field is applied. Electrophoretic mobility, μ, is defined as the steady-state velocity per unit field and is given by,

$$\mu = v / E, \qquad \{1\}$$

where v is the steady-state velocity of the molecule and E is the electric field. v is also given as

$$v = qE / f, \qquad \{2\}$$

where q is the charge of the molecule, and f is the frictional coefficient which reflects the size and shape of the molecule (25). The results of having many charged acidic groups on a chain may be to reduce SDS binding to the polypeptide chain, to lower the overall charge density on the chain, and to decrease the steady-state velocity.

In conclusion, twelve different macromolecules were synthesized by the cloning strategy outlined. The design strategy is versatile in that it can be used to make a library of proteins that can be easily modified at either the helical domain or the random coil domain. It is our intention to investigate the gelation behavior of the triblock copolymers when changing the charge pattern of the helical domain as well as the size and the charge density of the random coil domain.

(a)

	EcoRI		BamHI							NheI	
5’	AA	□TC	GGA	TCC	GAT	GAC	GAT	GAC	AAA	GCT	AGC
3’		G	CCT	AGG	CTA	CTG	CTA	CTG	TTT	CGA	TCG

	NruI			BstEII			SphI			SpeI	
TAT	CGC	GAT	GGT	GAC	CCG	CGC	ATG	CCG	ACT	AGT	
ATA	GCG	CTA	CCA	CTG	GGC	GCG	TAC	GGC	TGA	TCA	

		BamHI		HindIII		
TGG	TAA	GGA	TCC	A		3’
ACC	ATT	CCT	AGG	TTC	GA	5’

(b)

	EcoRI		BamHI								NheI
5’	AA	TTC	GGA	TCC	GAT	GAC	GAT	GAC	AAA	TGG	GCT
3’		G	CCT	AGG	CTA	CTG	CTA	CTG	TTT	ACC	CGA

	BstEII					XhoI			SpeI	
AGC	GGT	GAC	CAT	GTG	GCG	CCT	CGA	GAC	ACT	AGT
TCG	CCA	CTG	GTA	CAC	CGC	GGA	GCT	CTG	TGA	TCA

					BamHI		HindIII		
ATG	GGT	GGC	TGC	TA□	GGA	TCC	A		3’
TAC	CCA	CCG	ACG	AT□	CCT	AGG	TTC	GA	5’

Figure 2.1 Sequences of coding (5’-3’) and noncoding (3’-5’) DNA for (a) **1** and (b) **2** incorporated into the HindIII and EcoRI restriction sites of cloning vector pUC-18. Important restriction sites are underlined and □ indicates a deletion of a nucleotide.

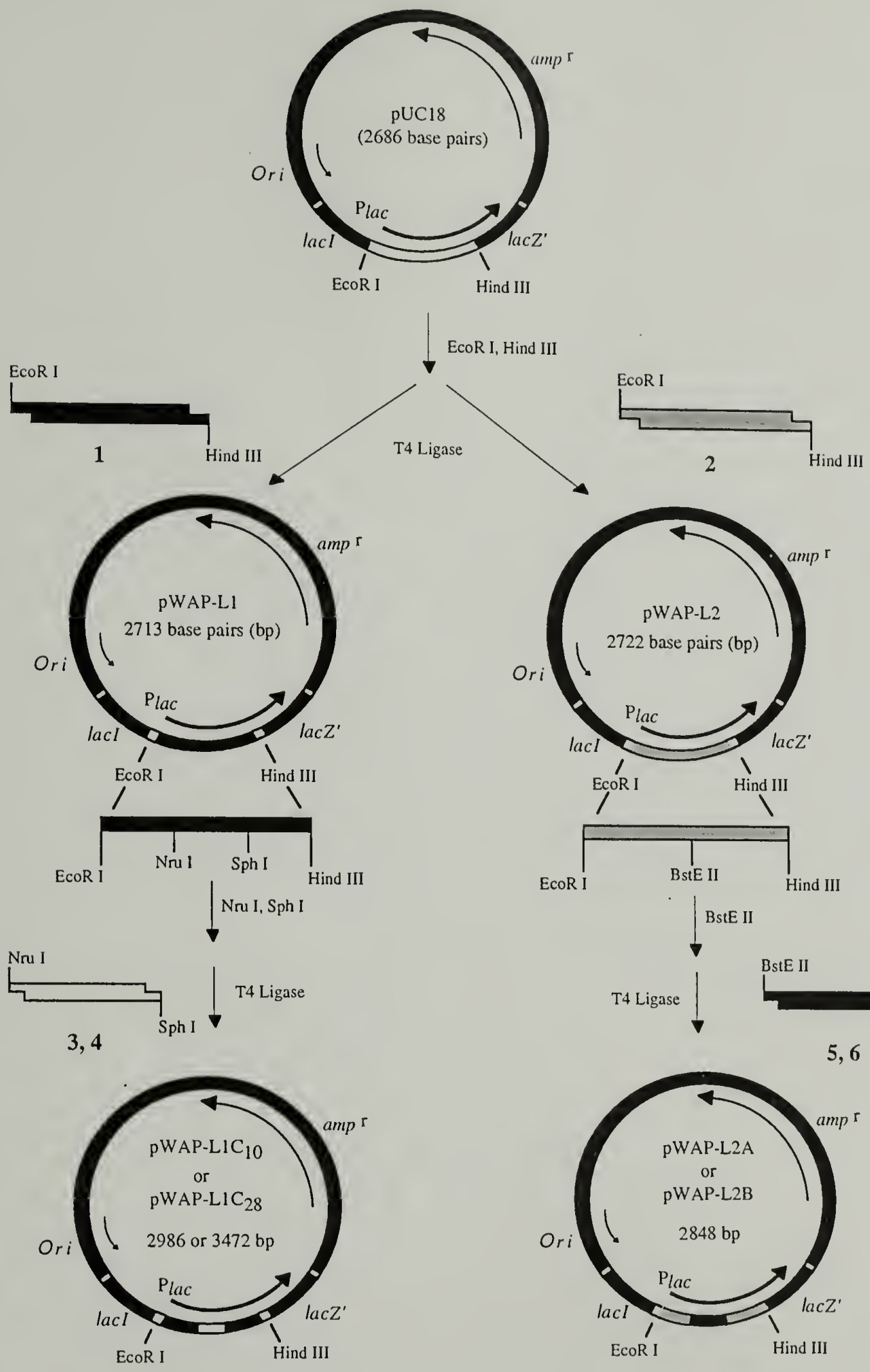


Figure 2.2 Strategy for cloning artificial genes encoding random coil and α-helix into pUC18 plasmid DNA.

NruI BanI

5' C GAT CCG ATG GGT GCC GGC GCT GGT GCG GGC

3' G CTA GGC TAC CCA CGG CCG CGA CCA CGC CCG

CCG GAA GGT GCA GGC GCT GGT GCG GGC CCG GAA

GGC CTT CCA CGT CCG CGA CCA CGC CCG GGC CTT

BanI

GGT GCC GGC GCT GGT GCG GGC GGC GAA GGT GCA

CCA CGG CCG CGA CCA CGC CCG CCG CTT CCA CGT

BanI

GGC GCT GGT GCG GGC CCG GAA GGT GCC GGC GCT

CCG CGA CCA CGC CCG GGC CTT CCA CGG CCG CGA

GGT GCG GGC CCG GAA GGT GCA GGC GCT GGT GCG

CCA CGC CCG GGC CTT CCA CGT CCG CGA CCA CGC

BanI

GGC CCG GAA GGT GCC GGC GCT GGT GCG GGC CCG

CCG GGC CTT CCA CGG CCG CGA CCA CGC CCG GGC

BanI

GAA GGT GCA GGC GCT GGT GCG GGC CCG GAA GGT

CTT CCA CGT CCG CGA CCA CGC CCG GGC CTT CCA

GCC GGC GCT GGT GCG GGC CCG GAA GGT GCA GGC

CGG CCG CGA CCA CGC CCG GGC CTT CCA CGT CCG

BanI SphI

GCT GGT GCG GGC CCG GAA GGT GCC CGC ATG 3'

CGA CCA CGC CCG GGC CTT CCA CGG GC 5'

Figure 2.3 Sequences of synthetic DNA **3** encoding for $[(AG)_3PEG]_{10}$ inserted into the NruI and SphI restriction sites. Important restriction sites are underlined.

NruI BanI

5' C GAT CCG ATG GGT GCC GGC GCT GGT GCG GGC

3' G CTA GGC TAC CCA CGG CCG CGA CCA CGC CCG

CCG GAA GGT GCA GGC GCT GGT GCG GGC CCG GAA

GGC CTT CCA CGT CCG CGA CCA CGC CCG GGC CTT

BanI

GGT GCC GGC GCT GGT GCG GGC GGC GAA GGT GCA

CCA CGG CCG CGA CCA CGC CCG CCG CTT CCA CGT

BanI

GGC GCT GGT GCG GGC CCG GAA GGT GCC GGC GCT

CCG CGA CCA CGC CCG GGC CTT CCA CGG CCG CGA

GGT GCG GGC CCG GAA GGT GCA GGC GCT GGT GCG

CCA CGC CCG GGC CTT CCA CGT CCG CGA CCA CGC

BanI

GGC CCG GAA GGT GCC GGC GCT GGT GCG GGC CCG

CCG GGC CTT CCA CGG CCG CGA CCA CGC CCG GGC

BanI

GAA GGT GCA GGC GCT GGT GCG GGC CCG GAA GGT

CTT CCA CGT CCG CGA CCA CGC CCG GGC CTT CCA

GCC GGC GCT GGT GCG GGC CCG GAA GGT GCA GGC

CGG CCG CGA CCA CGC CCG GGC CTT CCA CGT CCG

BanI

GCT GGT GCG GGC CCG GAA GGT GCC GGC GCT GGT

CGA CCA CGC CCG GGC CTT CCA CGG CCG CGA CCA

GCG GGC CCG GAA GGT GCA GGC GCT GGT GCG GGC

CGC CCG GGC CTT CCA CGT CCG CGA CCA CGC CCG

BanI

CCG GAA GGT GCC GGC GCT GGT GCG GGC CCG GAA

GGC CTT CCA CGG CCG CGA CCA CGC CCG GGC CTT

Continued

Figure 2.4 Sequence of synthetic DNA **4** encoding for $[(AG)_3PEG]_{28}$ inserted into the NruI and SphI restriction sites. Important restriction sites are underlined.

```
                                                                  BanI
GGT GCA GGC GCT GGT GCG GGC CCG GAA GGT GCC
CCA CGT CCG CGA CCA CGC CCG GGC CTT CCA CGG

GGC GCT GGT GCG GGC CCG GAA GGT GCA GGC GCT
CCG CGA CCA CGC CCG GGC CTT CCA CGT CCG CGA
                        BanI
GGT GCG GGC CCG GAA GGT GCC GGC GCT GGT GCG
CCA CGC CCG GGC CTT CCA CGG CCG CGA CCA CGC

GGC CCG GAA GGT GCA GGC GCT GGT GCG GGC CCG
CCG GGC CTT CCA CGT CCG CGA CCA CGC CCG GGC
      BanI
GAA GGT GCC GGC GCT GGT GCG GGC CCG GAA GGT
CTT CCA CGG CCG CGA CCA CGC CCG GGC CTT CCA
                                    BanI
GCA GGC GCT GGT GCG GGC CCG GAA GGT GCC GGC
CGT CCG CGA CCA CGC CCG GGC CTT CCA CGG CCG

GCT GGT GCG GGC CCG GAA GGT GCA GGC GCT GGT
CGA CCA CGC CCG GGC CTT CCA CGT CCG CGA CCA
                        BanI
GCG GGC CCG GAA GGT GCC GGC GCT GGT GCG GGC
CGC CCG GGC CTT CCA CGG CCG CGA CCA CGC CCG

CCG GAA GGT GCA GGC GCT GGT GCG GGC CCG GAA
GGC CTT CCA CGT CCG CGA CCA CGC CCG GGC CTT
BanI
GGT GCC GGC GCT GGT GCG GGC CCG GAA GGT GCA
CCA CGG CCG CGA CCA CGC CCG GGC CTT CCA CGT
                                    BanI
GGC GCT GGT GCG GGC CCG GAA GGT GCC GGC GCT
CCG CGA CCA CGC CCG GGC CTT CCA CGG CCG CGA

GGT GCG GGC CCG GAA GGT GCA GGC GCT GGT GCG
CCA CGC CCG GGC CTT CCA CGT CCG CGA CCA CGC
                  BanI        SphI
GGC CCG GAA GGT GCC CGC ATG 3'
CCG GGC CTT CCA CGG GC      5'
```

(a)

```
          BstEII
5'        GT  GAC  CTG  GAA  AAC  GAA  GTG  GCC  CAG  CTG  GGA
3'             GAC  CTT  TTG  CTT  CAC  CGG  GTC  GAC  CTT
                    BglII
     AGG  GAA  GTT  AGA  TCT  CTG  GAA  GAT  GAA  GCG  GCT
     TCC  CTT  CAA  TCT  AGA  GAC  CTT  CTA  CTT  CGC  CGA
                                   XhoI
     GAA  CTG  GAA  CAA  AAA  GTC  TCG  AGA  CTG  AAA  AAT
     CTT  GAC  CTT  GTT  TTT  CAG  AGC  TCT  GAC  TTT  TTA
                                             BstEII
     GAA  ATC  GAA  GAC  CTG  AAA  GCC  GAA  ATT  G         5'
     CTT  TAG  CTT  CTG  GAC  TTT  CGG  CTT  TAA  CCA CTG  3'
```

(b)

```
          BstEII
5'        GT  GAC  CTG  AAA  AAC  AAA  GTG  GCC  CAG  CTG  AAA
3'             GAC  TTT  TTG  TTT  CAC  CGG  GTC  GAC  TTT
                    BglII
     AGG  AAA  GTT  AGA  TCT  CTG  AAA  GAT  AAA  GCG  GCT
     TCC  TTT  CAA  TCT  AGA  GAC  TTT  CTA  TTT  CGC  CGA
                                   XhoI
     GAA  CTG  AAA  CAA  GAA  GTC  TCG  AGA  CTG  GAA  AAT
     CTT  GAC  TTT  GTT  CTT  CAG  AGC  TCT  GAC  CTT  TTA
                                             BstEII
     GAA  ATC  GAA  GAC  CTG  AAA  GCC  AAA  ATT  G         5'
     CTT  TAG  CTT  CTG  GAC  TTT  CGG  TTT  TAA  CCA CTG  3'
```

Figure 2.5 (a) Sequence of synthetic DNA **5** encoding for acidic leucine zipper and (b) sequence of synthetic DNA **6** encoding for basic leucine zipper inserted into BstEII restriction sites. Important sites are underlined

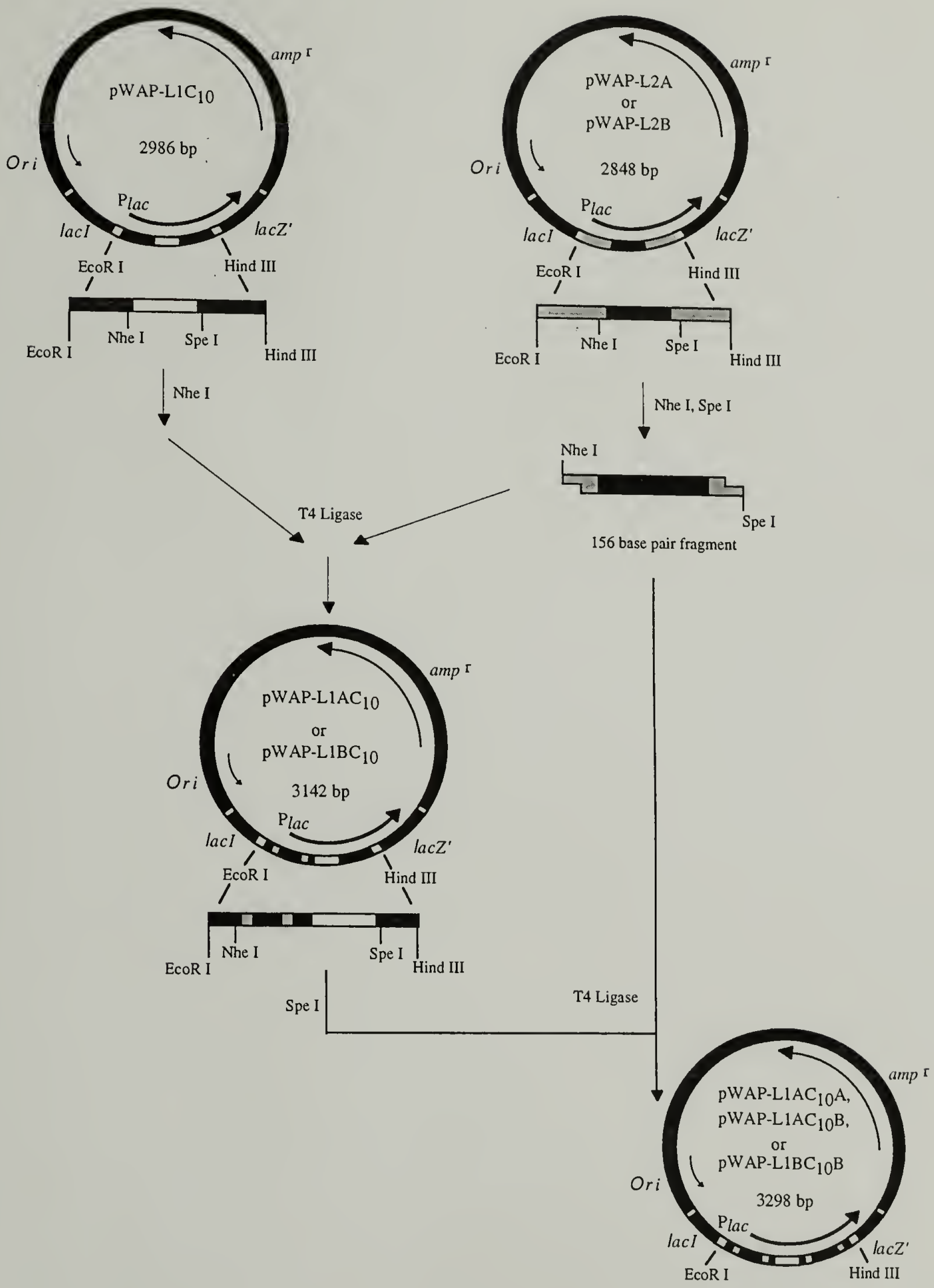


Figure 2.6 Strategy for cloning artificial genes encoding helix-coil-helix into pWAP-L1C_{10}.

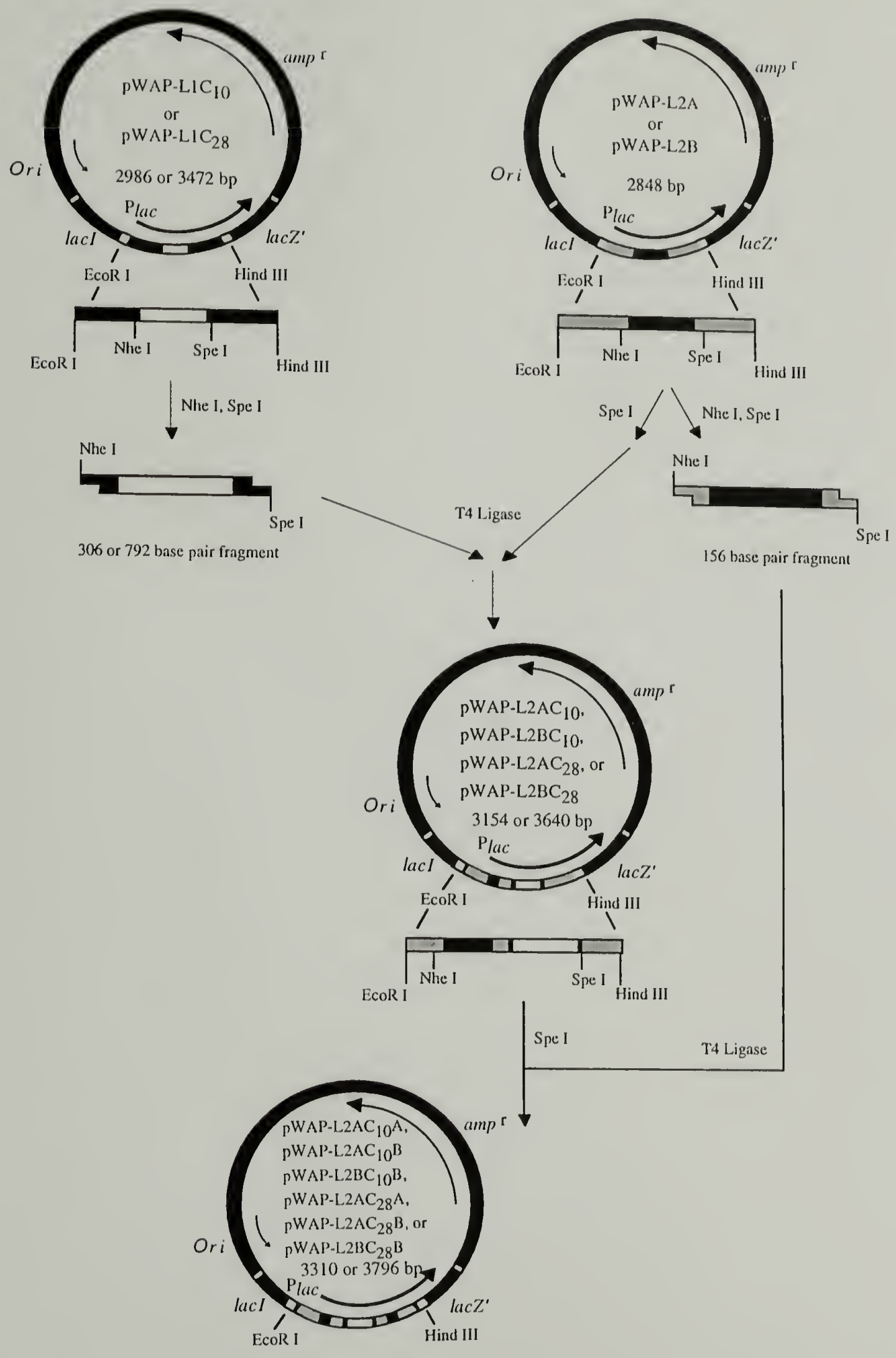


Figure 2.7 Alternate strategy for cloning artificial genes encoding helix-coil-helix into pWAP-L2A and pWAP-L2B.

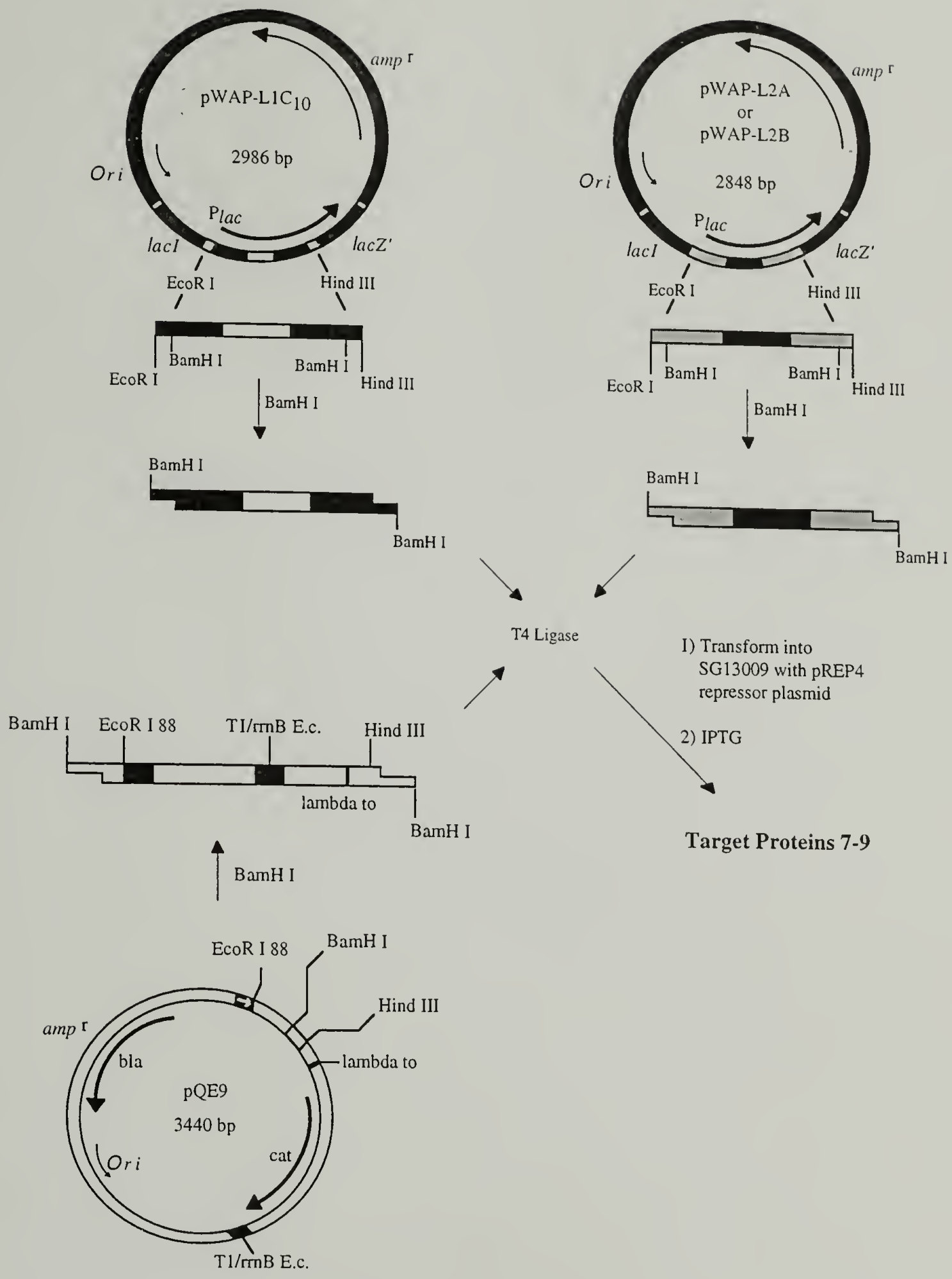


Figure 2.8 Strategy for expressing artificial genes encoding helix or coil in *Qiagen*™ pQE9 plasmid DNA.

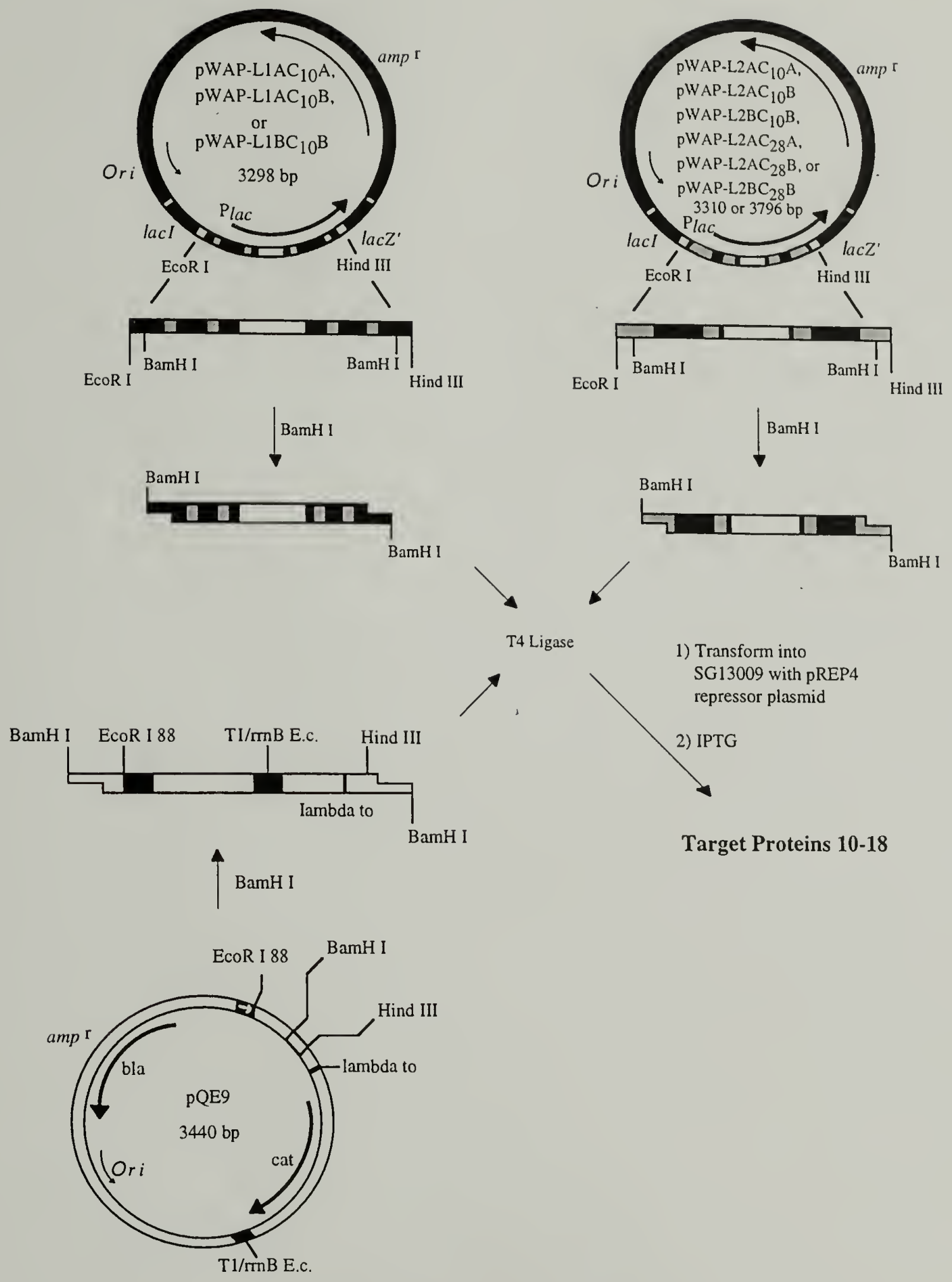


Figure 2.9 Strategy for expressing artificial genes encoding helix-coil-helix in *Qiagen*™ pQE9 plasmid DNA.

Table 2.1 Protein yields per 1 L of cell growth medium after purification with Ni^{2+} metal affinity chromatography.

Protein	Domain Structure	Yield (mg)
7	Acys	122.4
8	Bcys	157.1
9	Ctrp	26.4
10	AC_{10}Atrp	35.0
11	AC_{10}Btrp	137.3
12	BC_{10}Btrp	113.4
13	AC_{10}Acys	56.2
14	AC_{10}Bcys	104.2
15	BC_{10}Bcys	64.2
16	AC_{28}Acys	22.3
17	AC_{28}Bcys	52.4
18	BC_{28}Bcys	26.7

*Yields are an average of 2 (1L) expressions.

(a)

5’
MRGSHHHHHHGSDDDDKWASGDLENEVAQLEREVRSLEDEAAELEQKVSRL
KNEIEDLKAEIGDHVAPRDTSMGGC 3’

(b)

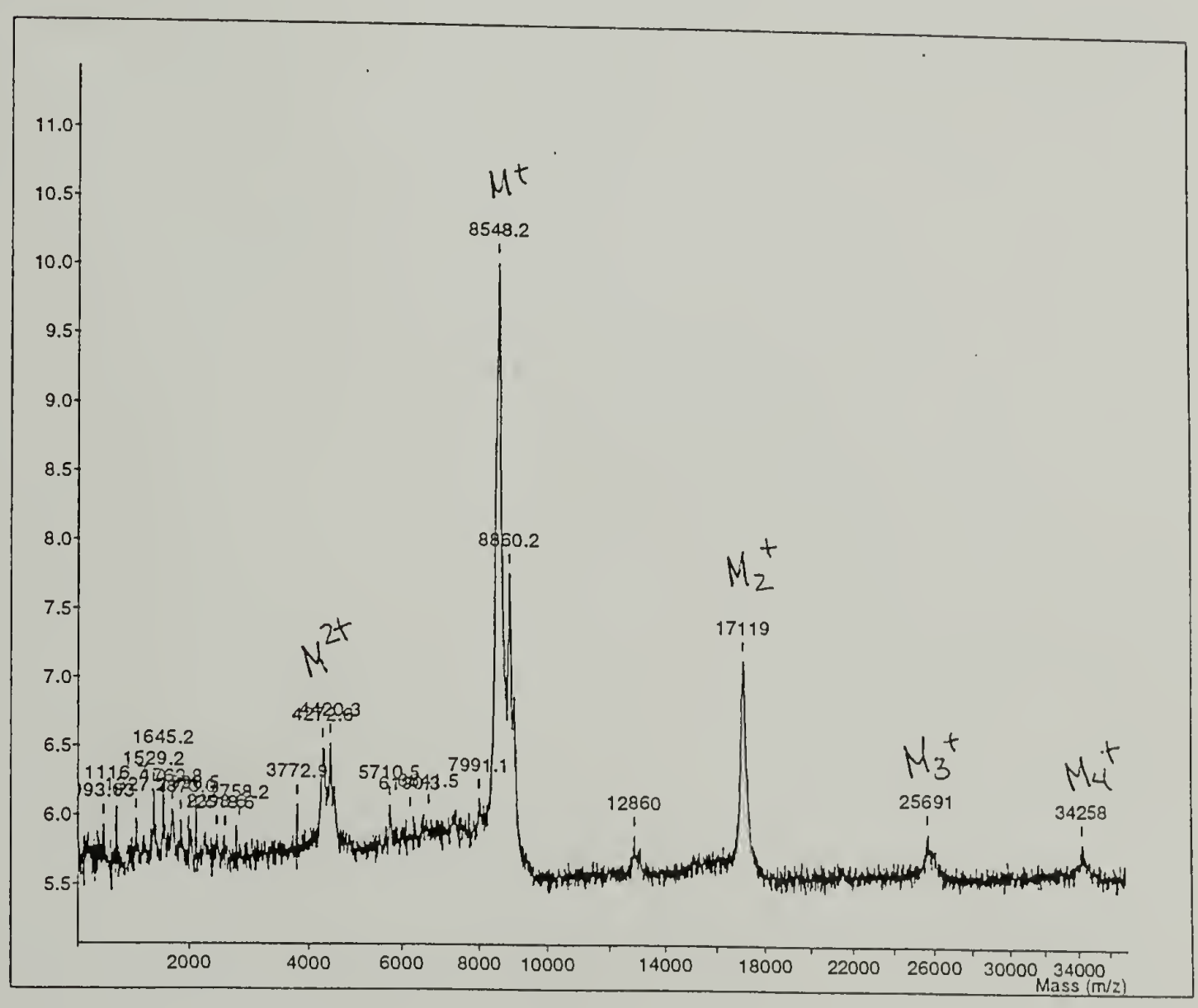


Figure 2.10 (a) Protein sequence 7 and (b) MALDI mass spectrum of 7 in α-cyano-4-hydroxycinnamic acid with calculated molecular weight of 8549.

(a)

5’
MRGSHHHHHHGSDDDDKWASGDLKNKVAQLKRKVRSLKDKAAELKQEVSR
LENEIEDLKAKIGDHVAPRDTSMGGC 3’

(b)

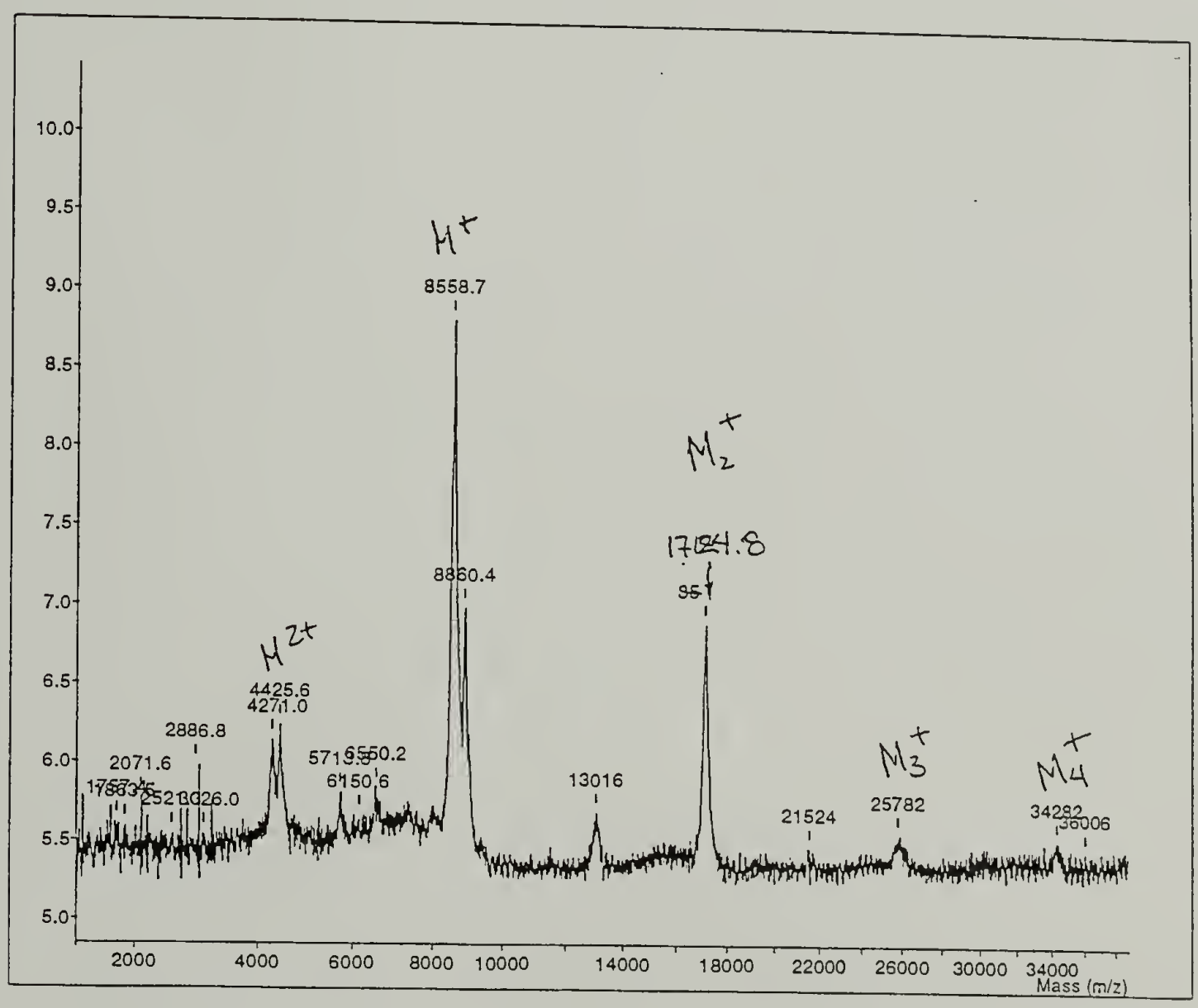


Figure 2.11 (a) Protein sequence **8** and (b) MALDI mass spectrum of **8** in α-cyano-4-hydroxycinnamic acid with calculated molecular weight of 8544.

(a)

5’ MRGSHHHHHHGSDDDDKASYRDPMG**[(AG)$_3$PEG]**$_{10}$ARMPTSW 3’

(b)

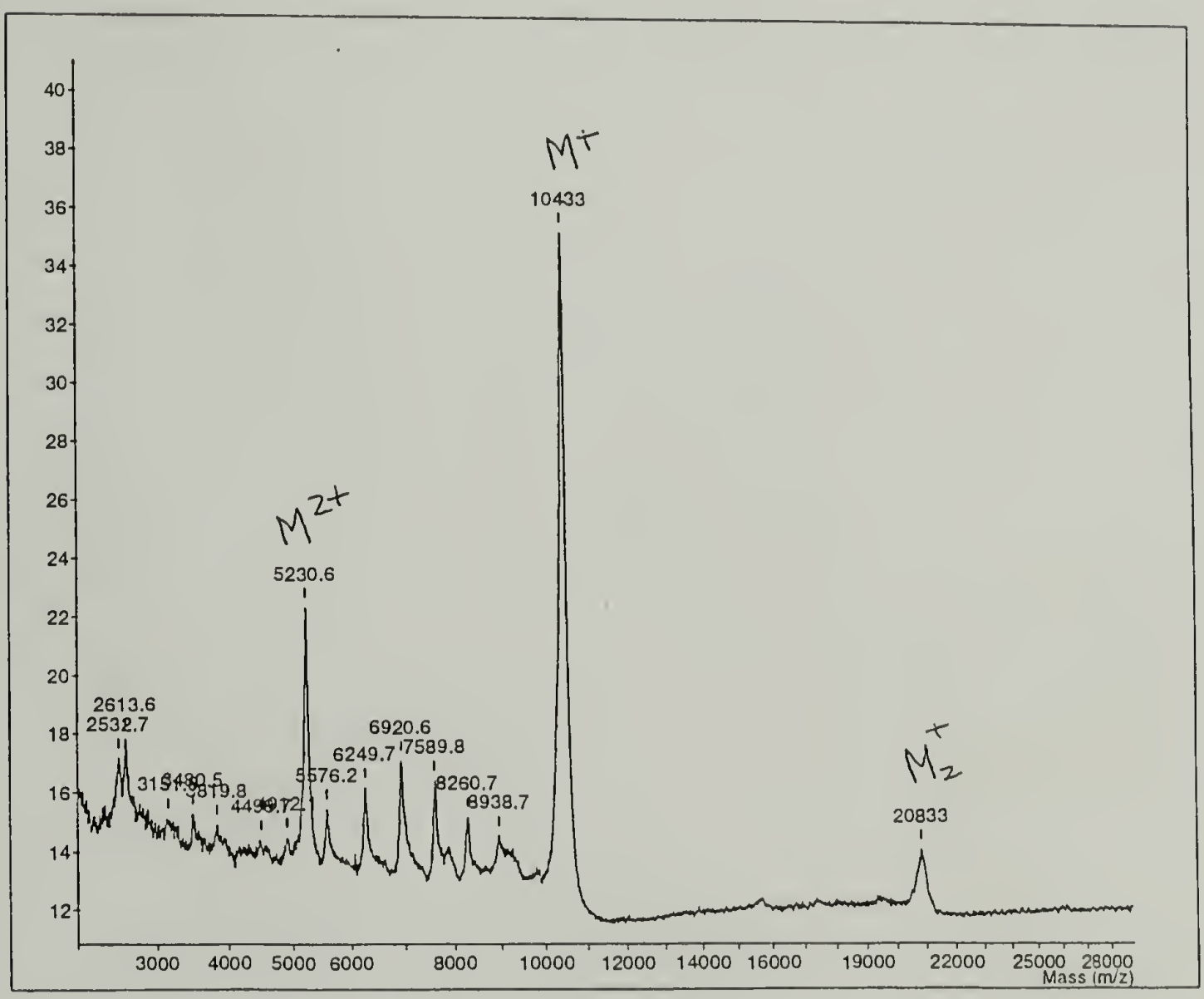


Figure 2.12 (a) Protein sequence **9** and (b) MALDI mass spectrum of **9** in α-cyano-4-hydroxycinnamic acid with calculated molecular weight of 10383.

(a)

5’
MRGSHHHHHHGSDDDDKASGDLENEVAQLEREVRSLEDEAAELEQKVSRLK
NEIEDLKAEIGDHVAPRDTSYRDPMG[**(AG)$_3$PEG**]$_{10}$ARMPTSGDLENEVAQLER
EVRSLEDEAAELEQKVSRLKNEIEDLKAEIGDHVAPRDTSW 3’

(b)

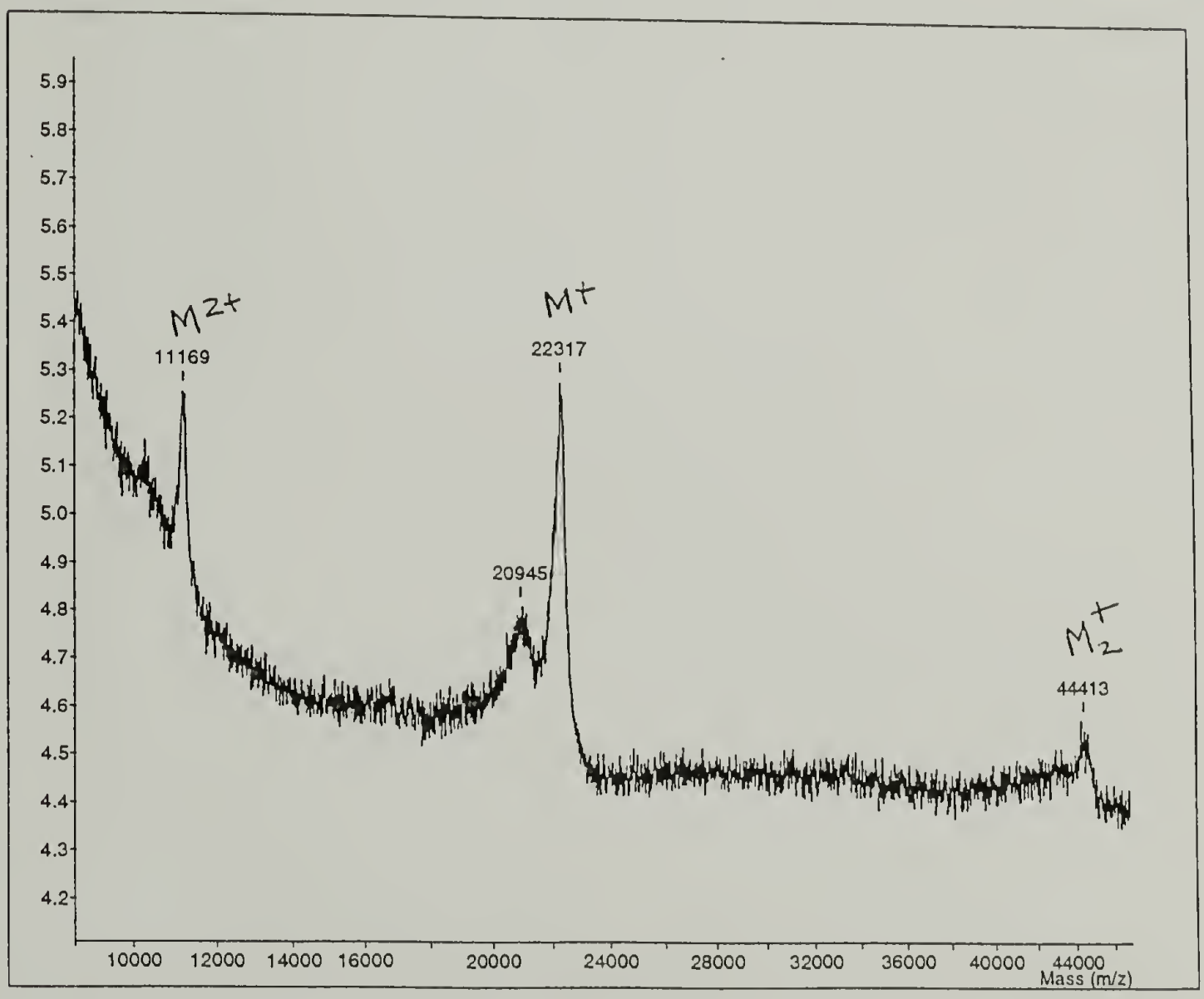


Figure 2.13 (a) Protein sequence **10** and (b) MALDI mass spectrum of **10** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 22091.

(a)

5’
MRGSHHHHHHGSDDDDKASGDLENEVAQLEREVRSLEDEAAELEQKVSRLK
NEIEDLKAEIGDHVAPRDTSYRDPMG[**(AG)$_3$PEG**]$_{10}$ARMPTSGDLKNKVAQLKR
KVRSLKDKAAELKQEVSRLENEIEDLKAKIGDHVAPRDTSW 3’

(b)

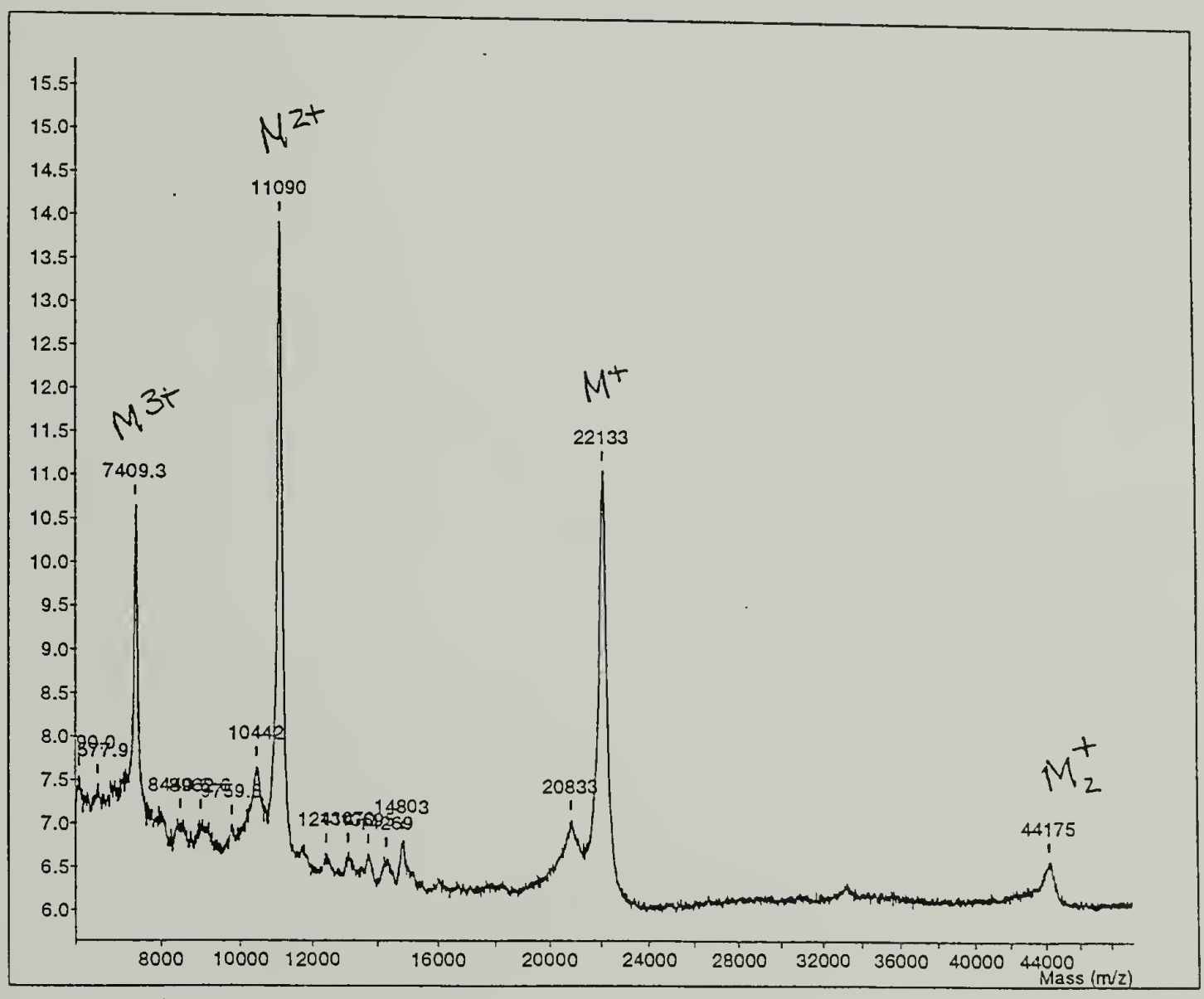


Figure 2.14 (a) Protein sequence **11** and (b) MALDI mass spectrum of **11** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 22085.

(a)

5’
MRGSHHHHHHGSDDDDKASGDLKNKVAQLKRKVRSLKDKAAELKQEVSRL
ENEIEDLKAKIGDHVAPRDTSYRDPMG**[(AG)$_3$PEG]**$_{10}$ARMPTSGDLKNKVAQLK
RKVRSLKDKAAELKQEVSRLENEIEDLKAKIGDHVAPRDTSW 3’

(b)

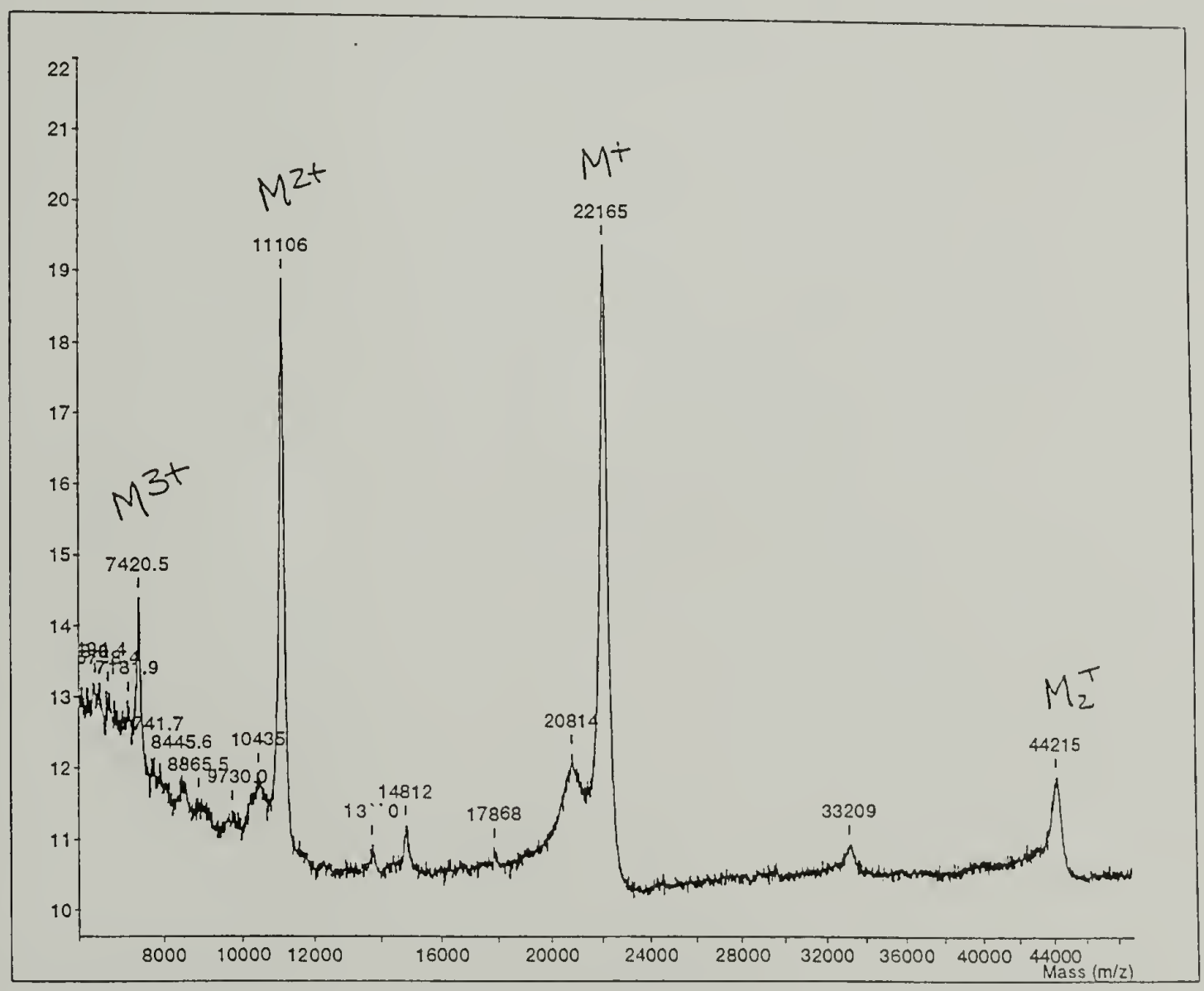


Figure 2.15 (a) Protein sequence **12** and (b) MALDI mass spectrum of **12** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 22079.

(a)

5’ MRGSHHHHHHGSDDDDKWASGDLENEVAQLEREVRSLEDEAAELEQKVSR
LKNEIEDLKAEIGDHVAPRDTSYRDPMG[**(AG)$_3$PEG**]$_{10}$ARMPTSGDLENEVAQL
EREVRSLEDEAAELEQKVSRLKNEIEDLKAEIGDHVAPRDTSMGGC 3’

(b)

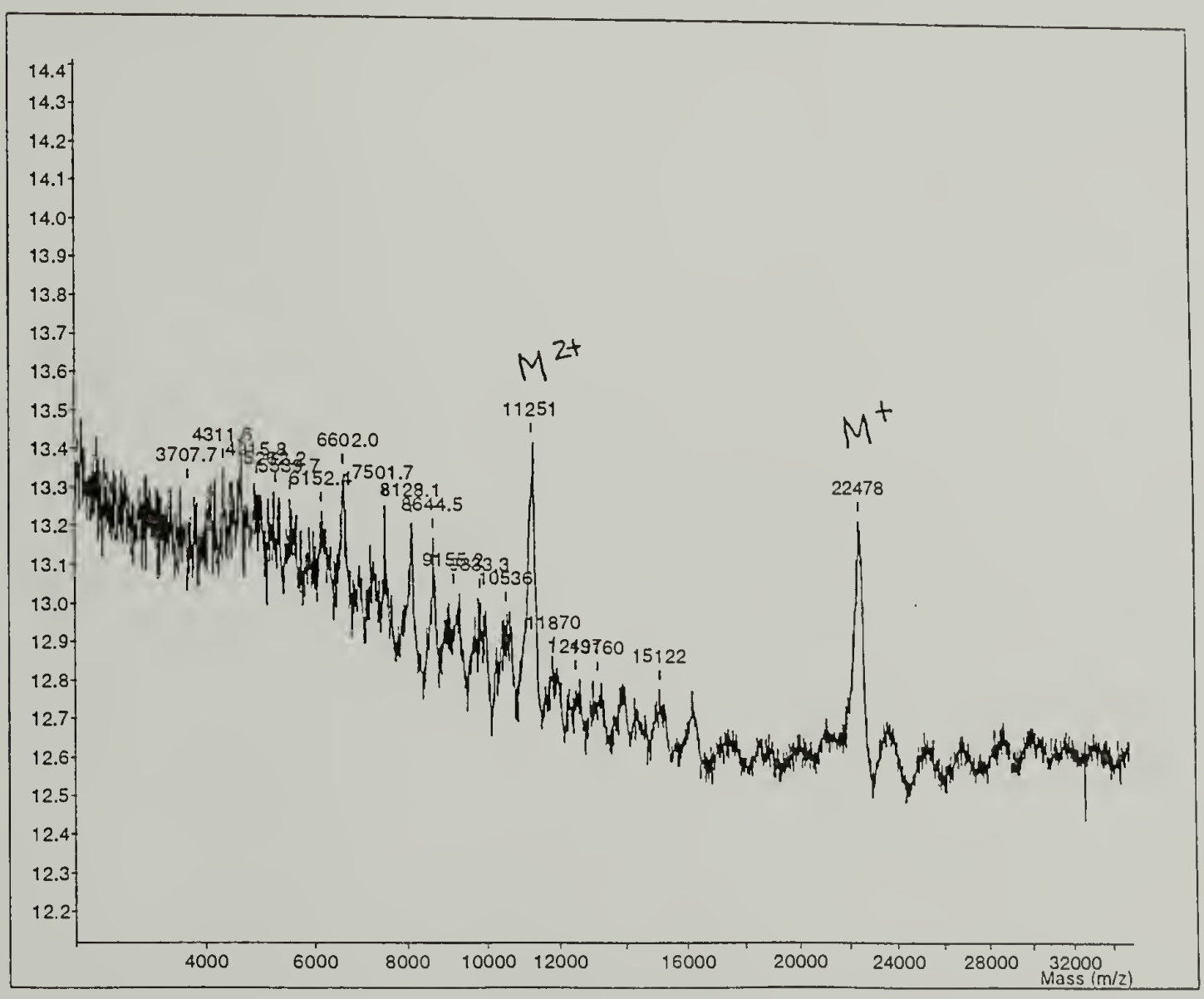


Figure 2.16 (a) Protein sequence **13** and (b) MALDI mass spectrum of **13** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 22439.

(a)

5' MRGSHHHHHHGSDDDDKWASGDLENEVAQLEREVRSLEDEAAELEQKVSR
LKNEIEDLKAEIGDHVAPRDTSYRDPMG**[(AG)$_3$PEG]$_{10}$**ARMPTSGDLKNKVAQL
KRKVRSLKDKAAELKQEVSRLENEIEDLKAKIGDHVAPRDTSMGGC 3'

(b)

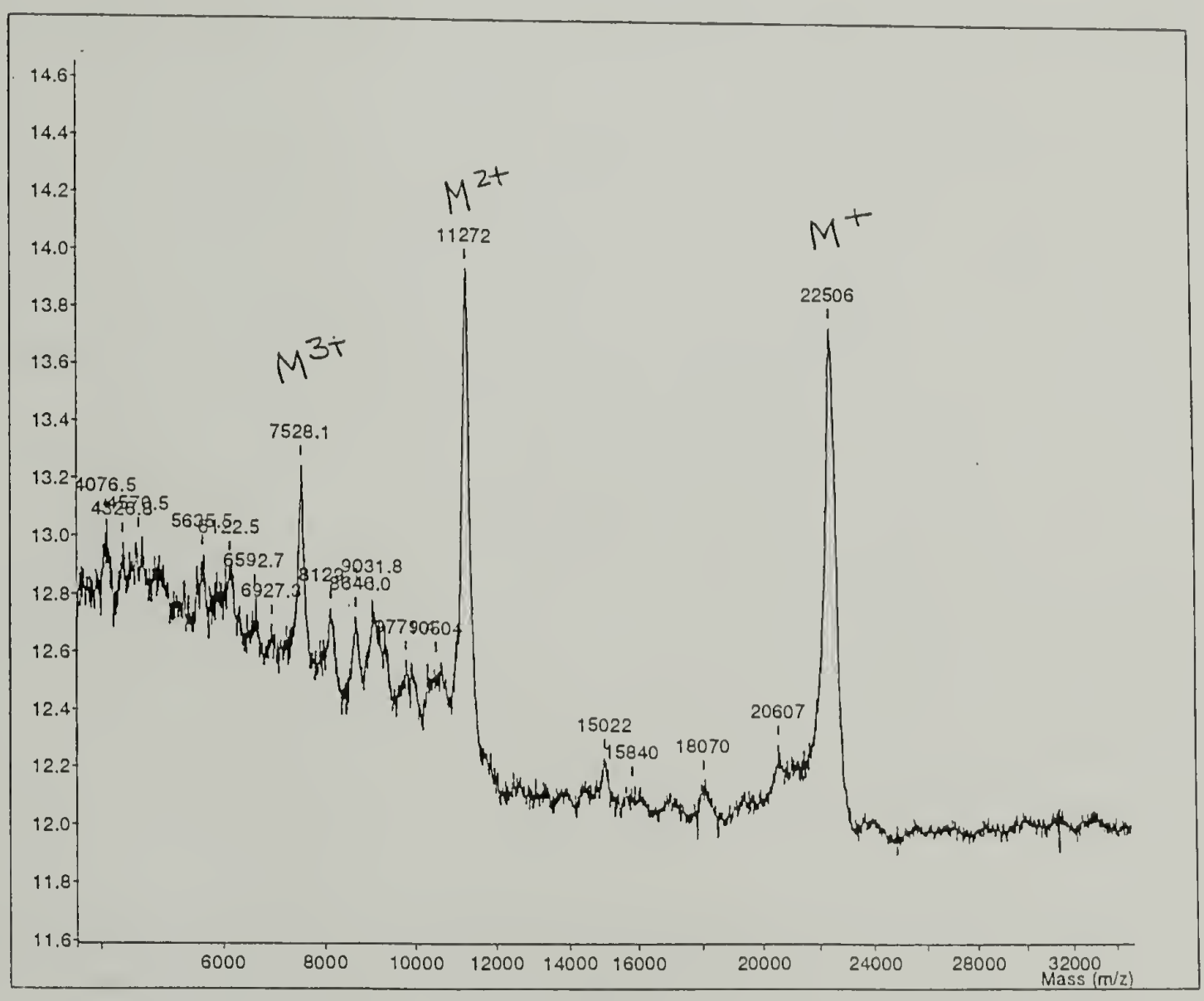


Figure 2.17 (a) Protein sequence **14** and (b) MALDI mass spectrum of **14** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 22433.

(a)

5’

MRGSHHHHHHGSDDDDKWASGDLKNKVAQLKRKVRSLKDKAAELKQEVSR
LENEIEDLKAKIGDHVAPRDTSYRDPMG**[(AG)$_3$PEG]**$_{10}$ARMPTSGDLKNKVAQL
KRKVRSLKDKAAELKQEVSRLENEIEDLKAKIGDHVAPRDTSMGGC 3’

(b)

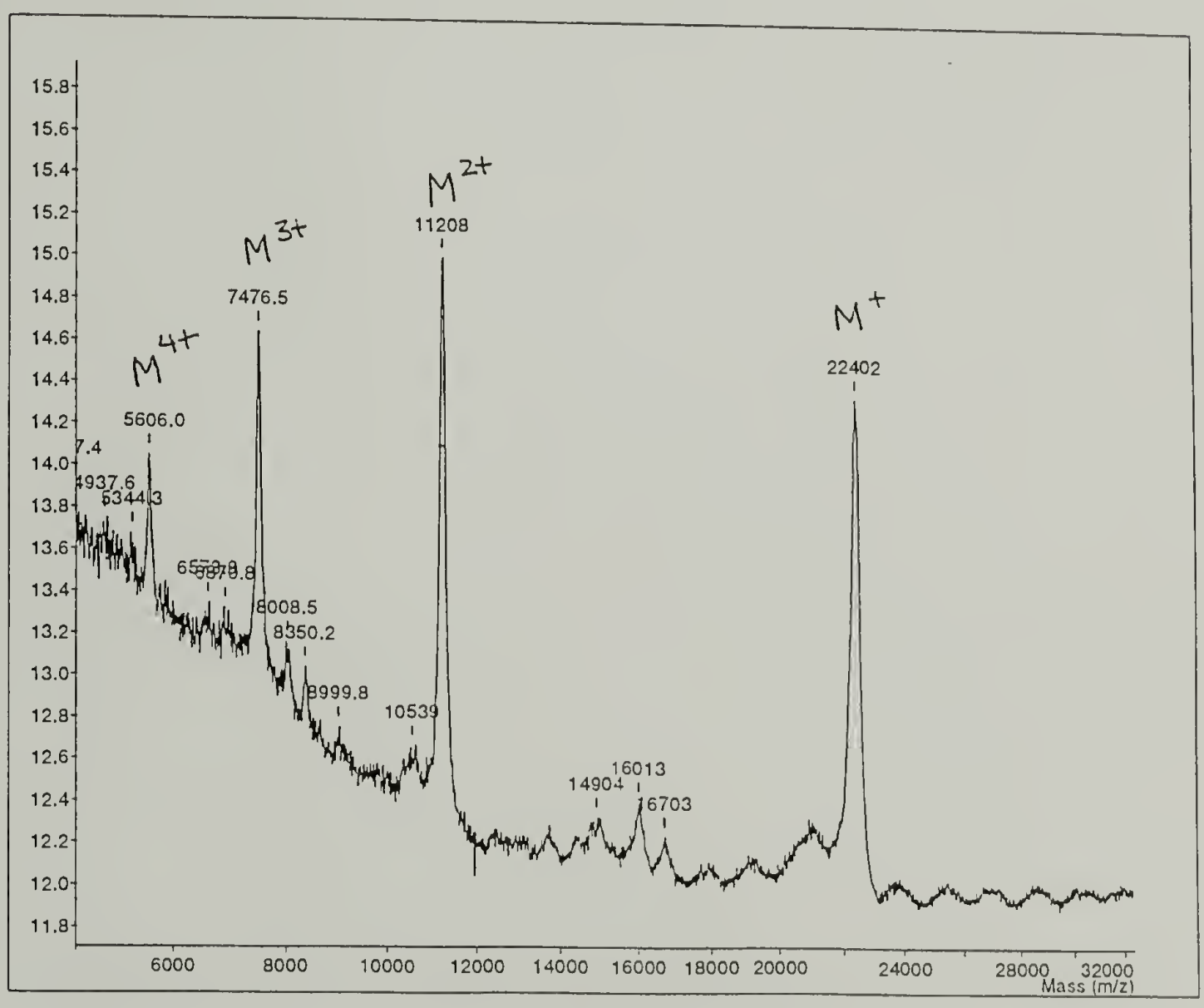


Figure 2.18 (a) Protein sequence **15** and (b) MALDI mass spectrum of **15** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 22428.

(a)

5' MRGSHHHHHHGSDDDDKWASGDLENEVAQLEREVRSLEDEAAELEQKVSR
LKNEIEDLKAEIGDHVAPRDTSYRDPMG[**(AG)$_3$PEG**]$_{28}$ARMPTSGDLENEVAQL
EREVRSLEDEAAELEQKVSRLKNEIEDLKAEIGDHVAPRDTSMGGC 3'

(b)

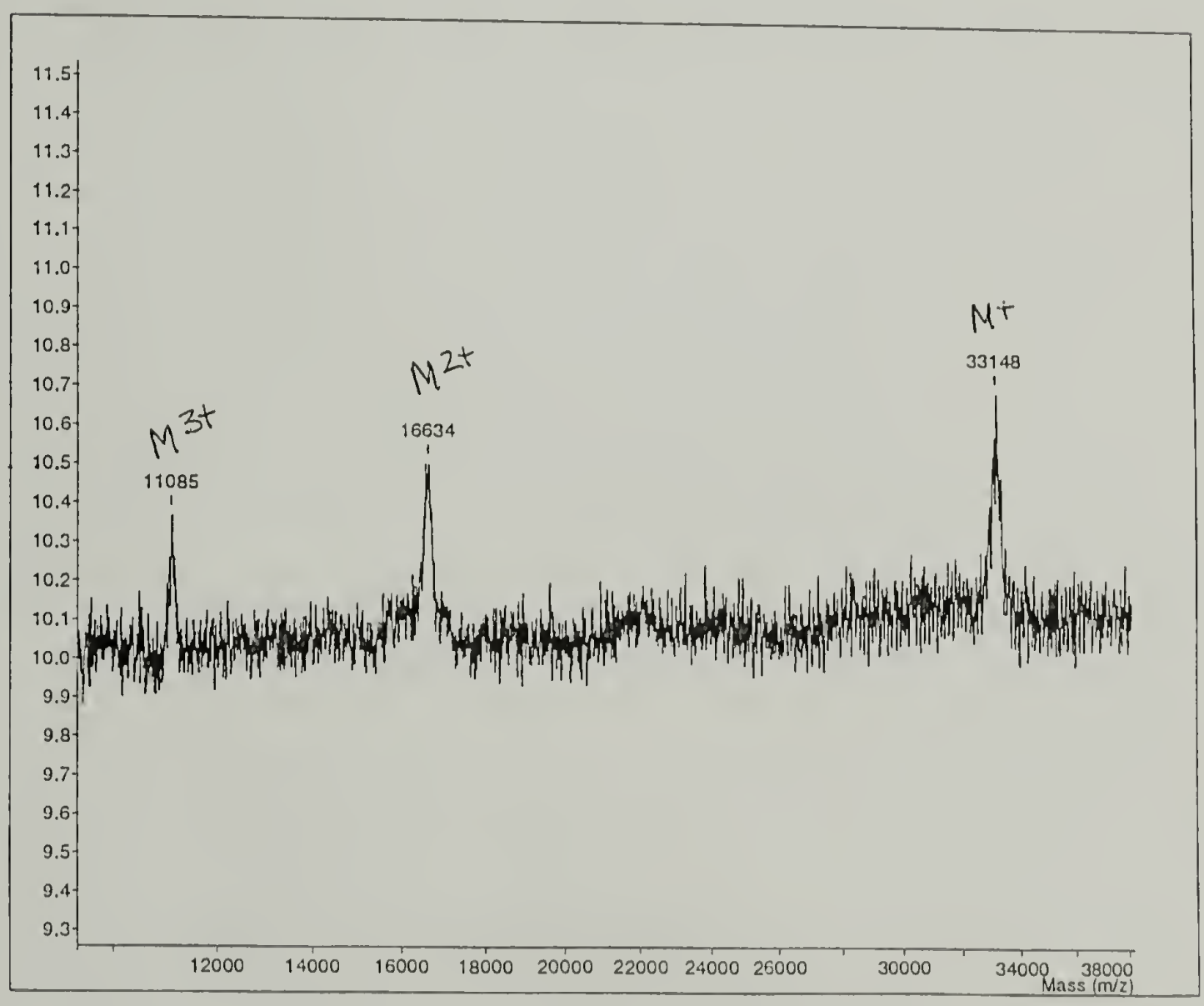


Figure 2.19 (a) Protein sequence **16** and (b) MALDI mass spectrum of **16** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 34450.

(a)

5' MRGSHHHHHHGSDDDDKWASGDLENEVAQLEREVRSLEDEAAELEQKVSR
LKNEIEDLKAEIGDHVAPRDTSYRDPMG[**(AG)$_3$PEG**]$_{28}$ARMPTSGDLKNKVAQL
KRKVRSLKDKAAELKQEVSRLENEIEDLKAKIGDHVAPRDTSMGGC 3'

(b)

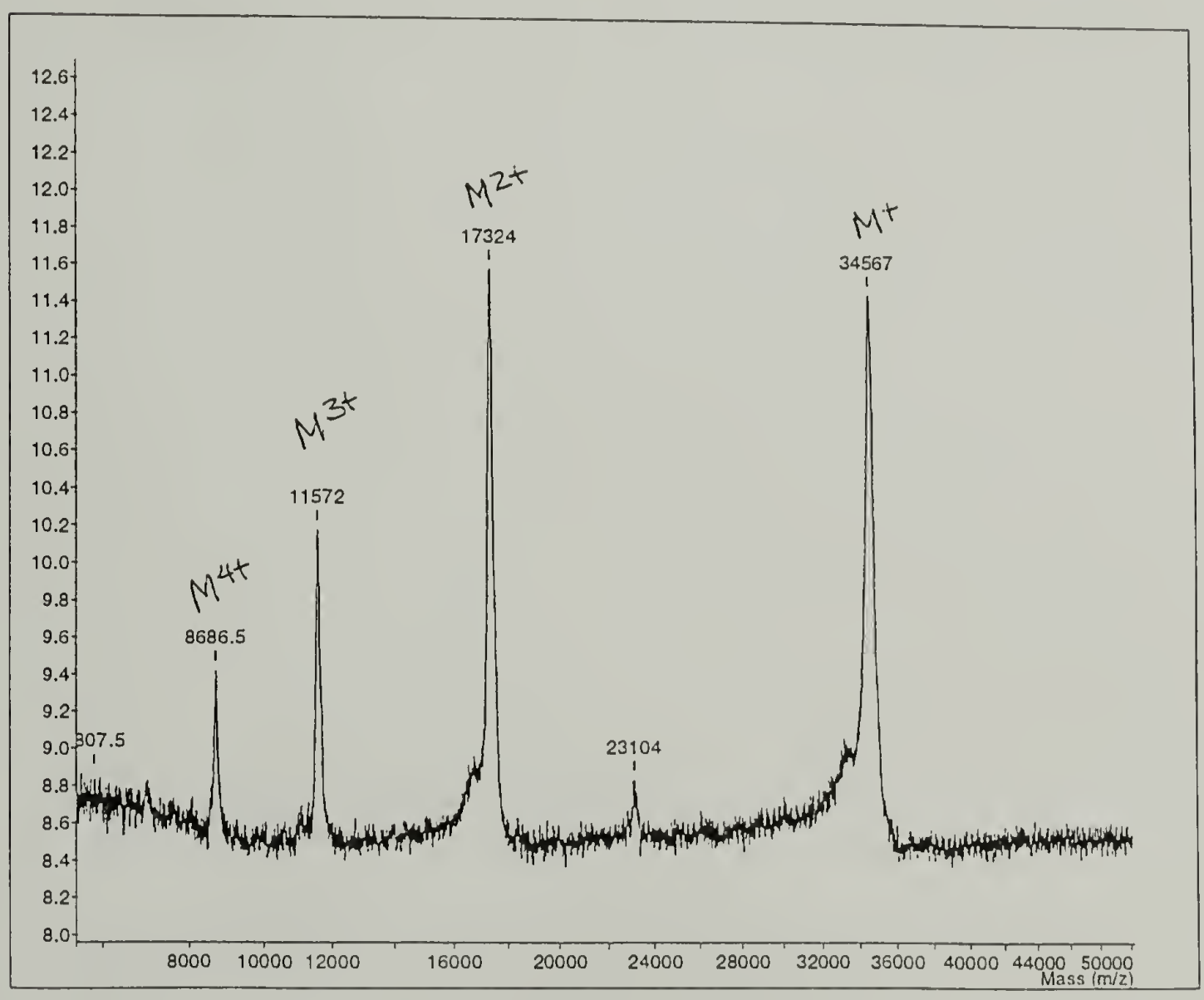


Figure 2.20 (a) Protein sequence **17** and (b) MALDI mass spectrum of **17** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 34445.

(a)

5’

MRGSHHHHHHGSDDDDKWASGDLKNKVAQLKRKVRSLKDKAAELKQEVSR
LENEIEDLKAKIGDHVAPRDTSYRDPMG**[(AG)$_3$PEG]$_{28}$**ARMPTSGDLKNKVAQL
KRKVRSLKDKAAELKQEVSRLENEIEDLKAKIGDHVAPRDTSMGGC 3’

(b)

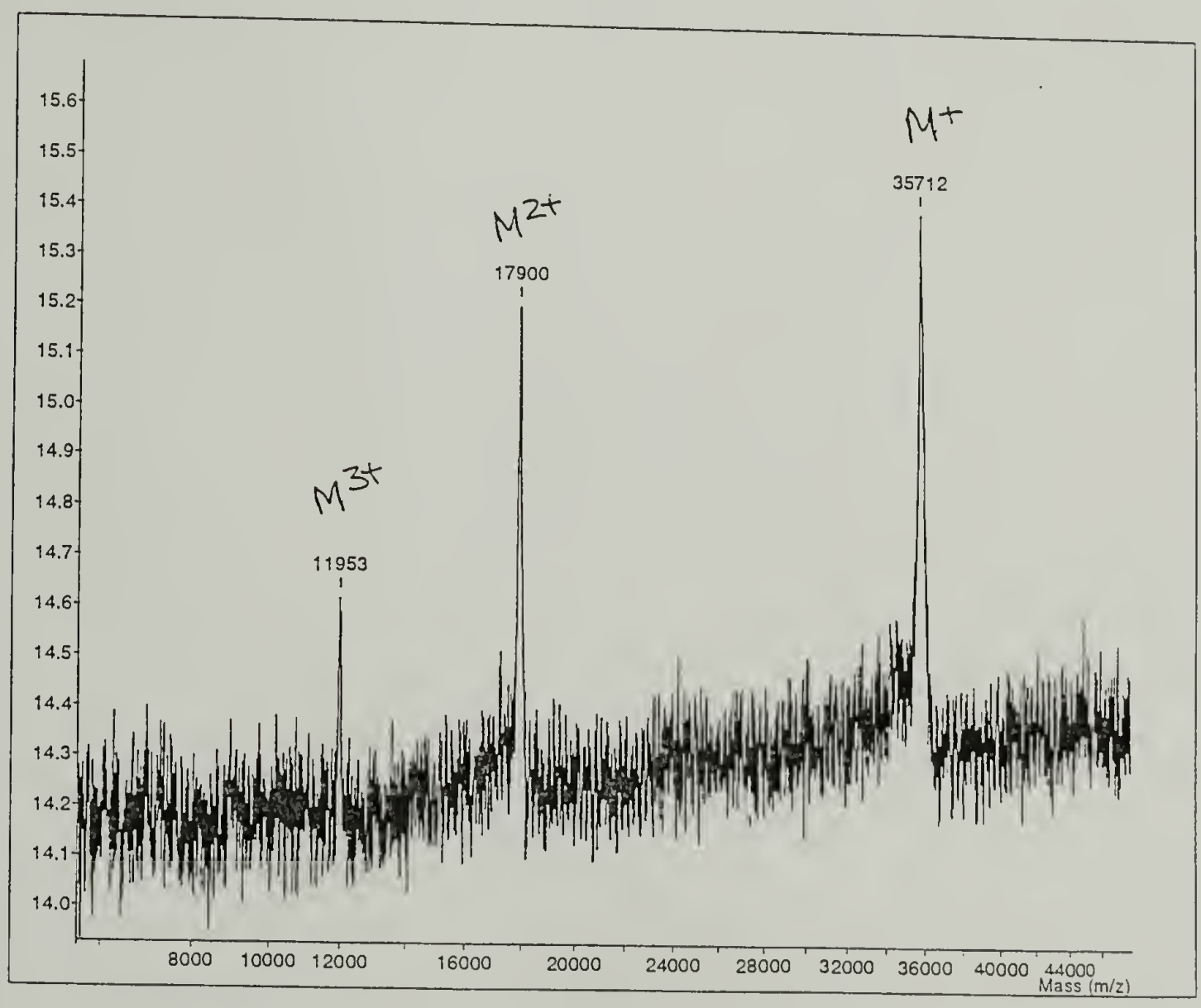


Figure 2.21 (a) Protein sequence **18** and (b) MALDI mass spectrum of **18** in 3,5-dimethoxy-4-hydroxycinnamic acid with calculated molecular weight of 34439.

Table 2.2 Summary of MALDI mass spectrometry data.

Protein	Domain Structure	$[M+H]^+$ (calculated)*	$[M+H]^+$ (observed)	relative error (%)
7	Acys	8550	8548	0.02
8	Bcys	8545	8559	0.16
9	Ctrp	10384	10433	0.47
10	AC_{10}Atrp	22092	22317	1.02
11	AC_{10}Btrp	22086	22133	0.21
12	BC_{10}Btrp	22080	22165	0.38
13	AC_{10}Acys	22440	22478	0.17
14	AC_{10}Bcys	22434	22506	0.32
15	BC_{10}Bcys	22429	22402	0.12
16	AC_{28}Acys	34451	33148	3.78
17	AC_{28}Bcys	34446	34567	0.35
18	BC_{28}Bcys	34440	35712	3.69

* Values are average molecular weights and refer to the protonated form of the polypeptide.

Table 2.3 Amino acid compositional analysis of proteins 7 and **8**.

amino acid	mol % (theor) 7	mol % (obs) 7	mol % (theor) **8**	mol % (obs) **8**
aspartic acid + asparagine	14.6	14.8	14.6	15.5
glutamic acid + glutamine	17.2	18.0	9.3	10.3
serine	8.0	8.3	8.0	8.3
glycine	8.0	8.6	8.0	8.7
histidine	9.3	9.2	9.3	8.9
arginine	6.6	6.8	6.6	6.8
threonine	1.4	1.7	1.4	1.6
alanine	8.0	7.9	8.0	8.1
proline	1.4	1.6	1.4	1.4
tyrosine	0.0	0.2	0.0	0.1
valine	5.3	4.8	5.3	4.5
methionine	2.7	2.3	2.7	2.3
cysteine	1.4	1.2	1.4	0.8
isoleucine	2.7	2.4	2.7	2.5
leucine	8.0	6.9	8.0	7.6
phenylalanine	0.0	0.1	0.0	0.1
lysine	5.3	5.4	13.2	12.7

Table 2.4 Amino acid compositional analysis of proteins **9** and **10**.

amino acid	mol % (theor) **9**	mol % (obs) **9**	mol % (theor) **10**	mol % (obs) **10**
aspartic acid + asparagine	4.1	4.1	8.5	8.4
glutamic acid + glutamine	8.2	8.0	16.0	14.7
serine	3.3	3.3	4.4	4.8
glycine	35.3	36.5	20.8	18.6
histidine	5.0	5.0	3.6	3.6
arginine	2.5	2.9	4.9	5.1
threonine	0.9	0.8	1.4	2.4
alanine	26.3	25.8	18.6	16.8
proline	9.9	10.1	6.2	6.0
tyrosine	0.9	0.9	0.5	1.1
valine	0.0	0.1	3.6	4.3
methionine	2.5	1.4	1.3	0.9
cysteine	0.0	0.1	0.0	0.1
isoleucine	0.0	0.0	1.8	2.5
leucine	0.0	0.1	5.3	6.2
phenylalanine	0.0	0.1	0.0	0.9
lysine	0.9	0.8	3.1	3.5

Table 2.5 Amino acid compositional analysis of proteins **11** and **12**.

amino acid	mol % (theor) **11**	mol % (obs) **11**	mol % (theor) **12**	mol % (obs) **12**
aspartic acid + asparagine	8.5	8.3	8.5	7.9
glutamic acid + glutamine	13.3	13.3	10.7	10.6
serine	4.4	4.4	4.4	4.6
glycine	20.8	21.0	20.8	21.4
histidine	3.6	3.5	3.6	3.8
arginine	4.9	5.0	4.9	5.0
threonine	1.4	1.3	1.4	1.4
alanine	18.6	18.5	18.6	18.6
proline	6.2	6.3	6.2	6.2
tyrosine	0.5	0.6	0.5	0.5
valine	3.6	3.5	3.6	3.4
methionine	1.3	1.3	1.3	0.8
cysteine	0.0	0.1	0.0	0.1
isoleucine	1.8	1.8	1.8	1.7
leucine	5.3	5.5	5.3	5.5
phenylalanine	0.0	0.1	0.0	0.1
lysine	5.8	5.6	8.4	8.4

Table 2.6 Amino acid compositional analysis of proteins **13** and **14**.

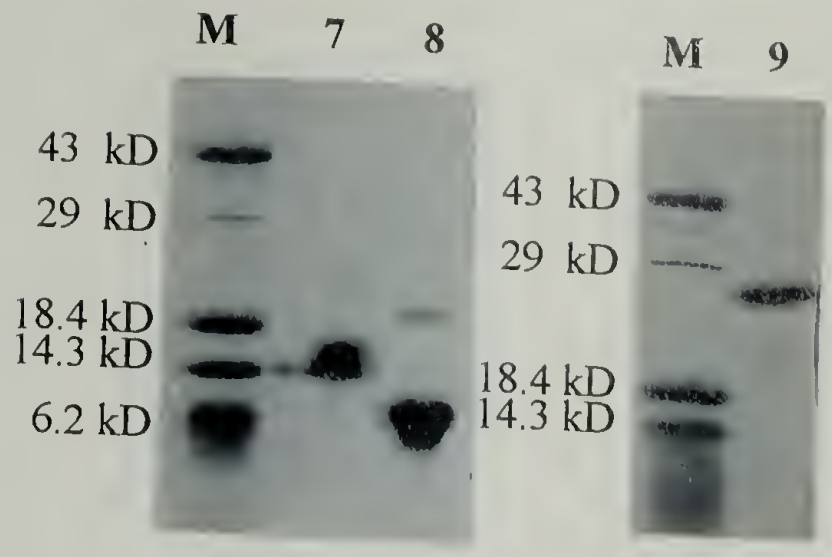
amino acid	mol % (theor) **13**	mol % (obs) **13**	mol % (theor) **14**	mol % (obs) **14**
aspartic acid + asparagine	8.3	7.5	8.3	7.7
glutamic acid + glutamine	15.7	15.6	13.1	12.9
serine	4.4	4.5	4.4	5.1
glycine	21.3	22.6	21.3	22.3
histidine	3.5	3.3	3.5	3.4
arginine	4.8	4.9	4.8	4.8
threonine	1.3	1.2	1.3	1.5
alanine	18.3	18.4	18.3	17.6
proline	6.1	6.3	6.1	6.7
tyrosine	0.5	0.5	0.5	0.6
valine	3.5	3.2	3.5	3.3
methionine	1.8	1.7	1.8	1.8
cysteine	0.5	0.3	0.5	0.3
isoleucine	1.8	1.6	1.8	1.7
leucine	5.2	5.3	5.2	5.1
phenylalanine	0.0	0.2	0.0	0.3
lysine	3.1	3.0	5.7	5.1

Table 2.7 Amino acid compositional analysis of proteins **15** and **16**.

amino acid	mol % (theor) **15**	mol % (obs) **15**	mol % (theor) **16**	mol % (obs) **16**
aspartic acid + asparagine	8.3	8.6	4.9	5.2
glutamic acid + glutamine	10.5	10.5	13.8	13.1
serine	4.4	5.8	2.6	2.9
glycine	21.3	19.7	30.9	30.3
histidine	3.5	3.9	2.1	2.5
arginine	4.8	4.9	2.8	3.3
threonine	1.3	1.4	0.8	0.9
alanine	18.3	16.5	24.5	23.7
proline	6.1	5.8	8.2	8.1
tyrosine	0.5	0.6	0.3	0.5
valine	3.5	3.7	2.1	2.2
methionine	1.8	1.4	1.0	0.8
cysteine	0.5	0.1	0.3	0.1
isoleucine	1.8	1.9	1.0	1.0
leucine	5.2	6.7	3.1	3.2
phenylalanine	0.0	0.1	0.0	0.1
lysine	8.3	8.4	1.8	1.9

Table 2.8 Amino acid compositional analysis of proteins **17** and **18**.

amino acid	mol % (theor) **17**	mol % (obs) **17**	mol % (theor) **18**	mol % (obs) **18**
aspartic acid + asparagine	4.9	4.8	4.9	5.0
glutamic acid + glutamine	12.3	12.0	10.7	10.3
serine	2.6	2.6	2.6	3.1
glycine	30.9	31.5	30.9	30.7
histidine	2.1	2.2	2.1	2.4
arginine	2.8	2.9	2.8	2.9
threonine	0.8	0.7	0.8	0.9
alanine	24.5	24.7	24.5	23.7
proline	8.2	8.3	8.2	8.3
tyrosine	0.3	0.4	0.3	0.4
valine	2.1	1.9	2.1	2.1
methionine	1.0	0.7	1.0	0.6
cysteine	0.3	0.2	0.3	0.1
isoleucine	1.0	0.9	1.0	1.1
leucine	3.1	3.1	3.1	3.8
phenylalanine	0.0	0.0	0.0	0.1
lysine	3.3	3.0	4.9	4.5



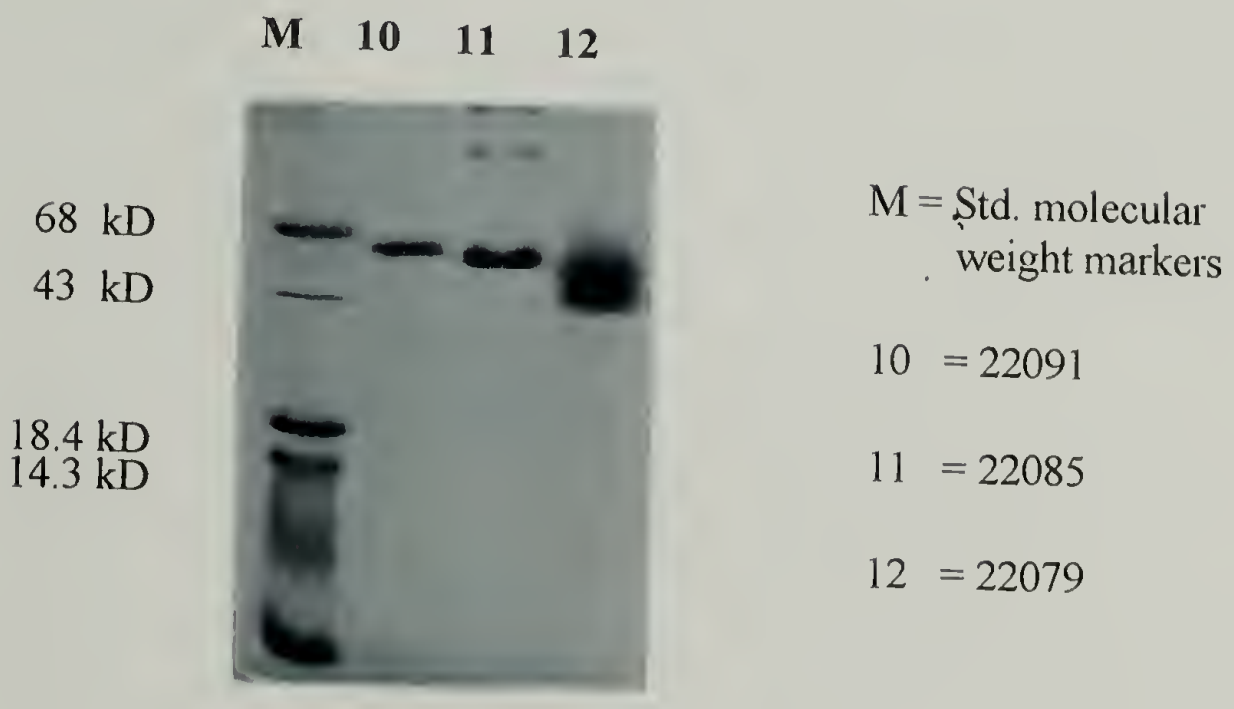
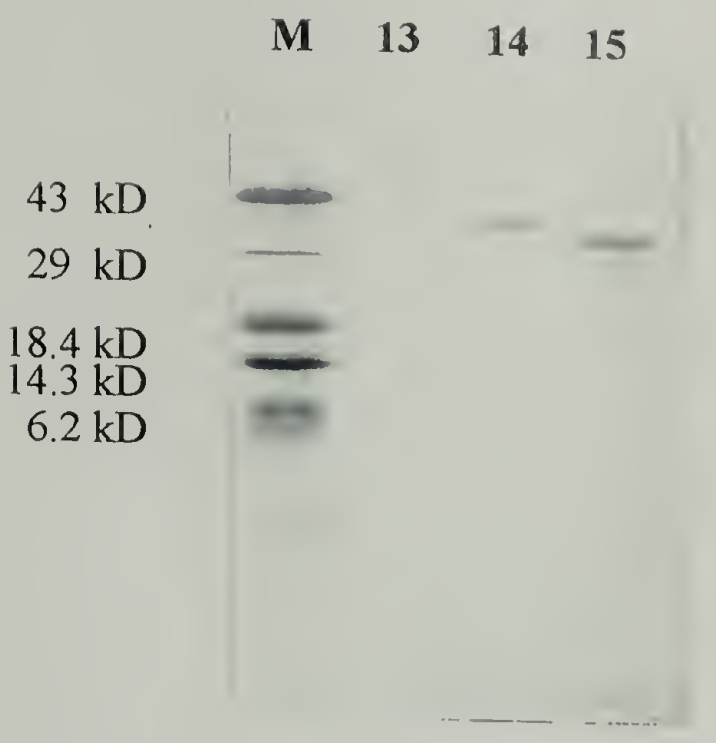


Figure 2.22 Purified proteins **7-12** visualized on a 14% sodium dodecyl sulfate (SDS)-polyacrylamide gel with Coomassie Brilliant Blue R-250. Calculated molecular weights are shown.

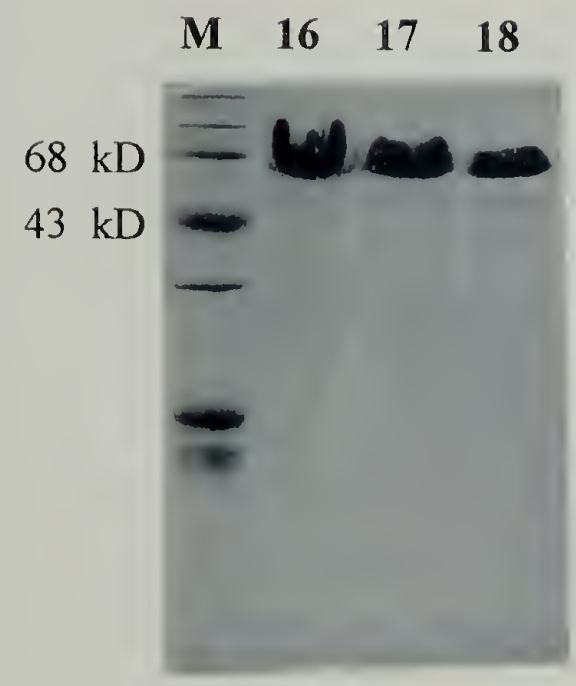


M = Std. molecular weight markers

13 = 22439

14 = 22433

15 = 22428



M = Std. molecular weight markers

16 = 34450

17 = 34445

18 = 34439

Figure 2.23 Purified proteins **13-18** visualized on a 14% sodium dodecyl sulfate (SDS)-polyacrylamide gel with Coomassie Brilliant Blue R-250. Calculated molecular weights are shown.

2.4 References

(1) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427.

(2) Zhang, G.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1992**, *25*, 3601.

(3) Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 536.

(4) Dong, W.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *PMSE Preprints* **1996**, *74*, 71.

(5) McGrath, K. P.; Butler, M. M.; DiGirolamo, C. M.; Petka, W. A.; Kaplan, D. L.; Laue, T. M. *submitted for publication.*

(6) Harbury, P. B; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401.

(7) McGrath, K. P.; Kaplan, D. L. *Mat. Res. Symp. Proc.* **1993**, *292*, 83.

(8) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

(9) Petka, W. A.; McGrath, K. P.; Kaplan, D. L.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polymer Preprints* **1994**, *35* (2), 452.

(10) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.

(11) Richardson, C.C. in *The Enzymes*; Academic Press: San Diego, 1981.

(12) Yanisch-Perron, C.; Viera J.; Messing, J. *Gene* **1985**, *33*, 103.

(13) Weiss, B.; Jacquemin-Sablon, A.; Live, T. R.; Fareed, G. C.; Richardson, C. C. *J. Biol. Chem.* **1968**, *243*, 4543.

(14) Sanger, F. *Annu. Rev. Biochem.* **1988**, *57*, 1.

(15) Tabor, S.; Richardson, C. C. *Proc. Nat. Acad. Sci. USA* **1987**, *84* (14), 4767.

(16) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.

(17) Cohen, S. A.; Meys, M.; Tarvin, T. L. in *The Pico-Tag Method: A Manual of Advanced Techniques for Amino Acid Analysis*; Millipore Corporation: Bedford, 1989.

(18) Heinrikson, R. L.; Meredith, S. C. *Anal. Biochem.* **1984**, *136*, 65.

(19) Bidlingmeyer, B. A.; Cohen, S. A.; Tarvin, T. L. *J. Chromatogr.* **1984**, *336*, 93.

(20) Laemmli, U. K. *Nature* **1970**, *227*, 680.

(21) Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1990**, *62*, 1836.

(22) Thannhauser, T. *Personal Communication.*

(23) Beavis, R. C.; Chait, B. T.; Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 7584.

(24) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213.

(25) Garfin, D. E. *Electrophoresis Methods* in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research*; Academic Press: San Diego, 1995.

CHAPTER 3

REVERSIBLE HYDROGELS FROM SELF-ASSEMBLING ARTIFICIAL PROTEINS

3.1 Objectives

Recombinant DNA technology was used to create a new class of protein-based materials that contain naturally occurring α-helical coiled coil protein ends, that act as physical crosslinks, separated by an artificial random coil poly(alanylglycine) protein, $[(AG)_3PEG]_{10}$, that acts as a flexible spacer. The physical aggregation of the terminal helices produces a three-dimensional network, with the flexible random coil retaining solvent and preventing precipitation of the chain. The result is physical gelation that is responsive to both pH and temperature changes. The reversibility of gelation results from the destabilizing effects of pH and temperature on the electrostatic and hydrophobic interactions of the helical ends. An investigation of the acidic helical protein (**7**), random coil poly(alanylglycine) protein (**9**), and acidic triblock copolymer protein (**13**) shown in Figure 3.1 will be presented. Both pH and temperature dependence of the three proteins in solution will be discussed in addition to the pH, temperature, and concentration dependence of gelation of protein **13**.

3.2 Introduction

Protein engineering of new functional macromolecular materials promises to be a valuable tool in creating novel polymers for biomedical applications. For instance, the combination of natural protein domains with artificial domains provides a pathway for making materials that have both functional and structural properties. These properties determine how a material responds to a combination of environmental stimuli such as pH, ionic strength, temperature, pressure, and various constituent molecules. Hydrogels are a special class of functional materials that respond to different kinds of stimuli in aqueous environments. They can be designed to swell or shrink under certain physiological conditions (1) and they can be used to encapsulated cells, drugs, and other molecules for site specific use (2-6).

The combination of both ionic and hydrophobic blocks within a synthetic polymeric chain has been shown to yield pH and temperature sensitive gelled networks. Chen *et al.* (7) have designed graft copolymers of poly(N-isopropyl acrylamide) (PNIPAAm) and poly(acrylic acid) (PAAc) to investigate the response of the chain to changes in temperature and pH with changes in copolymer composition. The results show that graft copolymers having between 50 and 80 wt% AAc have lower critical solution temperatures (LCST) that are pH dependent. The ionization of the -COOH affects the hydrogen bonding interactions between the grafted NIPAAm and the backbone AAc groups. The authors suggest that the temperature-induced pH-dependent behavior of these graft copolymers may be fabricated into hydrogels. A similar study involved preparing hydrogels of graft copolymers of polyethyleneglycol (PEG) on a

poly(methacrylic acid) (PMMAc) backbone (8). Hydrogen bonded complexes were obtained between the -O- and -COOH groups at low pH. Temperature sensitivity was achieved at pH 4 but not at pH 7.4 where the gels were highly swollen.

In addition to the gels described above, chemically cross-linked gels can also exhibit pH and temperature dependent behavior. Annaka *et al.* (9) have prepared copolymer gels of acrylic acid and methacryl-amido-propyl-trimethyl ammonium chloride by free radical polymerization. Both electrostatic (attractive or repulsive) and hydrogen bonding interactions result in the existence of several phases with both pH and temperature changes. These phases were determined by measuring the equilibrium swelling degree of a cylindrically shaped gel sample. Reversible swelling with pH and denaturant was found to exist in chemically cross-linked protein gels by the same group (10).

Our interests are to synthesize protein-based materials that behave as pH and thermally reversible hydrogels. By using genetic engineering principles, variables such as sequence, stereochemistry, and length are precisely controlled. In turn, it is possible to create non-covalent protein networks where the strength of the physical crosslinks, the pH and temperature at which the materials gel, and the pore size are also controlled. Reported herein is one of the block copolymers synthesized through genetic engineering techniques that has both pH and temperature sensitive domains within the polymer backbone.

Natural protein domains that are being used to control the gelation process are of a well-known class of α-helical proteins, coiled-coils. Transcriptional regulatory

proteins of this class include yeast GNC4 (11, 12) and mammalian C/EBP (13) and nuclear transforming proteins of this class include Jun, Fos, and Myc (11-17). The essential features of coiled-coil proteins are found in the sequence of seven amino acids that repeat a minimum of four times (18) to form a 7_2 α-helical structure (11). In the general heptad repeat denoted $(a\ b\ c\ d\ e\ f\ g)_n$, *a* and *d* positions are hydrophobic residues that facilitate and stabilize interchain association of α-helices by forming complementary hydrophobic helical faces that interact through hydrophobic and van der Waals interactions (19-22). In addition to the hydrophobic core, which is formed by the side-to-side packing of the *a* and *d* residues, the formation of coiled-coils is further modulated by interactions between regularly spaced charged groups at the *e* and *g* positions. The stabilization of the super coiled helices is enhanced or diminished by positioning attractive or repulsive charged groups at these positions (22, 23) and the propensity to form parallel or antiparallel coiled-coils is dependent on the charge-charge interactions between helices (24, 25).

The properties of both coiled-coils and a disordered water soluble protein, $[(AG)_3PEG]_{10}$, are being combined to form a triblock protein that self-assembles into a three dimensional network where the α-helical regions are designed to be the physical crosslinks and the alanylglycine portion as the solvent-swollen domain. The amino acid sequences of the proteins being investigated are shown in Figure 3.1. The *Helix*, which contains a 42 amino acid sequence (six internal heptad repeats, *abcdefg*), constitutes the coiled-coil region. The choice of residues at positions *a* and *d* was based on the *a* / *d* residue patterns see in the *Jun* oncogene product (26), which forms a weakly stabilized homodimer. An algorithm developed by Lupas *et al.* (27), which identifies the most

probable amino acids at positions *a-g*, was used to select residues occupying *b*, *c*, and *f* positions in coiled-coils. Residues that would increase the solubility of the coiled-coil complex in water were chosen first for these positions. Acidic or basic groups, glutamic acid and lysine, respectively, were placed at the *e* and *g* positions to provide intermolecular attraction or repulsion between chains. The coil domain used in the design is the polyelectrolyte, $[(AG)_3PEG]_{10}$, which was previously shown (28) to adopt disordered coil structures in solution and in the solid state.

3.3 Experimental

3.3.1 Biosynthesis of Macromolecules

Oligonucleotides were synthesized on a *Biosearch*™ Model 8700 DNA synthesizer by phosphoramidite chemistry (29) and ligated into the polylinker region of the pUC18 cloning vector (30). A 213 bp fragment encoding α-helical protein **7** was ligated into the EcoRI and HindIII restriction sites of pUC18 to yield pWAP-L2A. Likewise, a 351 bp fragment encoding random coil protein **9** was ligated into the EcoRI and HindIII restriction sites of pUC18 to yield pWAP-L1C. Recombinant DNA was cloned in *E.coli* strain DH5αF' before coding and non-coding sequences were verified by DNA sequence analysis. DNA fragments that encode **7** and **9** independently were used to form recombinant DNA that encodes for **13**. This combination of DNA was achieved through *Nhe* I and *Spe* I sites. All DNA fragments were isolated by BamHI digestion, and directionally ligated into the *E. coli* expression vector pQE9 (Qiagen, Chatsworth,

CA) to form a NH_2-terminal His fusion product. The pQE9 plasmid contains an *E. coli* phage T5 promoter and two *lac* operator sequences. Recombinant DNA plasmids encoding 7, **9**, and **13** were designated pQE9-L2A, pQE9-L1C_{10}, and pQE9-L2AC_{10}A respectively. Each of these DNA sequences contain a enterokinase cleavage site directly after the His tag. The host used for protein expression was *E. coli* strain SG13009 containing pREP4 repressor plasmid (Qiagen, Chatsworth, CA). Cultures containing these DNA sequences were grown at 37 °C in 2 L of TB medium (16 g Bacto-Tryptone, 10 g Yeast Extract, 5 g NaCl per liter) with ampicillin (100 μg/ml) and kanamycin (50 μg/ml) until the optical densities (600 nm, OD_{600}) were in excess of 2. Isopropyl-β-thiogalactoside (IPTG) (1 mM) was added and protein synthesis was induced for a period of 4 hours at 37°C. The cells were then centrifuged (22100*g* for 30 minutes) and 50 ml 6 M guanidine-HCl buffer containing Na_2HPO_4 (0.1 M, pH 8) was added. The cells were lysed by storing in a -80 °C freezer and the supernatant was collected for protein purification. Purification of the target proteins was achieved by metal affinity chromatography using a nickelnitrilotriacetic acid (Ni^{2+}-NTA) resin supplied by *Qiagen*™. Purified protein yields of 7, **9**, and **13** are 122 mg, 26 mg, and 56 mg per liter of growth medium respectively.

3.3.2 Characterization of Artificial Proteins

3.3.2.1 Additional Purification and Verification of Protein Synthesis

Reversed phase high performance liquid chromatograms were acquired on a Waters™ 717plus Autosampler, 486 Tunable Absorbance Detector, and 600 Controller that was equipped with a Vydac™ C18 column (Supelco, Bellefonte, PA). Proteins were dissolved in distilled, deionized water and filtered (0.22 μm) prior to injection. Elution profiles were monitored at 215 nm with a linear 2% gradient of water (0.1% trifluoroacetic acid) to acetonitrile that was run over a period of 60 minutes at a flow rate of 1 mL/min (Figure 3.2). Solvents were purged with helium. Amino acid compositional analysis as well as matrix assisted laser desorption mass spectrometry (MALDI-MS) were performed in the Analytical Chemistry and Peptide/DNA Synthesis Facility at Cornell University. Amino acid compositional analysis was made on a Pico-Tag Amino Acid Analysis System (*Millipore Corp.*, Bedford, MA). Amino acids were derivatized with phenyl isothiocyanate (PITC) and analyzed by reversed phase-HPLC with a 4.6 x 300 mm Nova Pack C18 column (*Waters Corp.*, Bedford, MA). All amino acid analyses were within 5% of the theoretical values. For MALDI experiments, protein samples (1-10 μM) were mixed with a molar excess of matrix solution (3,5-dimethoxy-4-hydroxycinnamic acid (0.05 M) for masses greater than 20,000 and α-cyano-4-hydroxycinnamic acid (0.05 M) for masses less than 20,000 in 30% acetonitrile/water, 0.1% trifluoroacetic acid). The probe surface was irradiated with

nitrogen pulsed laser at 337 nm. Detection of the protein ions was recorded in the positive ion mode with individual spectra being obtained with single laser shots of about 10^6 W/cm^2 irradiance. An accelerating voltage of 20 keV was used. The calculated mass of the **7** is 8549 and a peak at 8548 was found. Likewise, the calculated mass of **9** is 10383 and the peak at 10433 was reported and the calculated mass of **13** is 22439 whereas the mass analysis gave a peak at 22478.

3.3.2.2 Circular Dichroism (CD) Measurements

CD spectra were recorded on an Aviv 62DS spectropolarimeter (Lakewood, NJ) in 10 mM NaH_2PO_4, 150 mM NaCl, pH 7.4 (adjusted with 1 N NaOH). Final protein solution concentrations (after preparing 1 mg/ml filtered with 0.22 μm cellulose acetate membranes) were determined by amino acid analysis. Wavelengths from 250 nm to195 nm were scanned at 1.0 nm bandwidth with points taken every 0.5 nm. Experiments were performed in a 1 mm path length rectangular cell (Helma) at 25 °C. The instrument was calibrated with an aqueous solution of (1S)-(+)-(10)-camphorsulfonic acid; constant N_2 flushing was employed. Mean residue molar ellipticity is reported at wavelengths 195-250 nm ([θ], deg cm^2 dmol^{-1}) and is calculated from the following equation: $[\theta] = [\theta]_{obs} \times MRW/(c \times l)$, where $[\theta]_{obs}$ is ellipticity measured in millidegrees, MRW is the mean residue molecular weight or the molecular weight of the peptide divided by the number of amino acid residues, c is the peptide

concentration in g/L determined by quantitative amino acid analysis, and l is the optical path length in mm.

The calculation of protein secondary structure in solution was performed by three different selection procedures: ridge regression, variable selection, and neural networks. *AVIV* CD data files were converted to the proper format prior to running each program. The three different selection methods are described below. All programs were obtained from Dr. N. J. Greenfield (31).

Ridge Regression Analysis (CONTIN). The version of the CONTIN program, by Provenchur and Glöckner (32), was provided by Dr. S.Y. Venyaminov (Dept. Biochemistry and Molecular Biology, Mayo Foundation, Rochester, MN) and uses a variation of the method of least squares (33). The CD spectrum of an unknown protein can be fit by a linear combination of spectra of known conformation. The basis set used in this analysis has 16 original reference proteins plus poly-L-glutamate. Four basis structures were evaluated: α-helix, β-sheet, β-turn, and remainder. The errors associated with each calculation are reported as the variance.

Self-Consistent Method (SELCON). The SELCON program was provided by R.W. Woody with the original FORTRAN source code from Sreerama and Woody (34, 35) modified to run on a personal computer. Four basis structures were evaluated: α-helix, beta, turn, and other. Thirty-three protein reference spectra (36) were used. Basis set proteins are arranged in order of increasing root-mean-square (RMS) differences from the CD spectra of the protein to be analyzed. Spectra which are least like the spectrum of interest are systematically deleted. A series of iterations are made before self-consistency is obtained. Errors are reported in standard deviation.

Neural Networks (K2D). The K2D neural net analysis program was provided by Dr. M. Andrade. This algorithm (37) evaluates data between 200-240 nm to form 40 input neurons which are designated as the real neuron CD spectrum. In the "training" phase, connections are made between input of the real CD spectrum with those of protein standards. The weights of the connections are adjusted until the error between the calculated and the actual secondary structure is minimized. This is reported as the square of the euclidean distance between the real and the winning neuron CD spectra. This distance gives an estimation of the mean error in the prediction of values for α-helix, beta, and random secondary structures. If the distance larger than 0.08, the prediction will not be reliable.

Thermal melting curves were determined by monitoring the CD signal at 222 nm as a function of temperature. The signal was collected every 1 °C with an equilibration time of 1 minute. Spectra were collected in 10 mM NaH_2PO_4, 150 mM NaCl (pH adjusted with 1 N HCl or 1 N NaOH; equilibrated overnight). A thermostatically controlled cuvette holder (HP Model 89101A) was used to regulate the temperature to within ±0.2 °C. A two state mechanism was used to describe the equilibrium between folded and unfolded protein states. From this two state model, the thermal unfolding temperature or the temperature where the fraction unfolded is equal to fraction folded ($\Delta G = 0$) was determined by taking the first derivative of the CD signal, $[\theta]_{222\,nm}$, with respect to temperature^{-1} (K) and finding the minimum of this function (38, 26). All thermal melts were repeated in order to check reversibility; T_m values were within ±2 °C of each other.

3.3.2.3 Diffusing Wave Spectroscopy

The dynamic intensity auto correlation function of multiply-scattered light from a dilute suspension (0.5% solids diluted from stock) of tracer particles was measured with a *Coherent*™ Laser source (512 nm, 100 mW) in a transmission geometry. A single photon correlator with correlation software (*ALV*, Germany) was used. The tracer particles, 0.3 μm sulfonated polystyrene beads (SPS) (obtained stock as 8.8% solids, surface charge density of 1.5 $\mu C/cm^2$, area per charge group of 1060 $Å^2$/group) were obtained from *Interfacial Dynamics Corporation*. Samples were placed into a 10mm x 5mm x 45mm optical glass cuvette (*NSG Precision Cell*) and data was collected for a duration of six hours. Preparation of the latex suspensions was accomplished as follows. Purified protein (50 mg) was placed into the glass cuvette where 943 μL of tris(hydroxymethyl)aminomethane (Tris) (buffering range of 6.9-7.1) was added. The sample was mixed thoroughly before 1 μL volumes of KOH (8M) were added incrementally to change the pH. The gel containing the SPS beads was left for approximately three days to equilibrate at which time additional KOH was added if the desired pH was not found. The final volume of the gel containing SPS beads was adjusted to 1 mL once the pH of the gel stabilized. The final concentration of the gel was 5% (w/v). Concentration studies were made with samples prepared by diluting this volume.

3.3.2.4 UV-Vis Measurements

The reaction of 5,5'-dithiobis(2-nitrobenzoic acid) (Ellman's reagent) with protein **13** in buffered solutions (10 mM Na_2HPO_4, 150 mM NaCl, pH 7.9 or 10 mM Tris, pH 7.1) was monitored at 412 nm with a *Hitachi* Model U-2000 Double-Beam UV/Vis spectrophotometer. The amount of free thiol in solution was calculated by measuring the amount of colored anion produced over time with the following equation,

$$C_o = A/\varepsilon \times l \qquad \{1\}$$

where C_o is the molar concentration of free thiol in solution, A is the absorbance at 412 nm, ε is the extinction coefficient of the colored anion (13,600 M/cm), and l is path length (1 cm). The reaction of β-mercaptoethanol with Ellman's reagent was used as a control. To a 1 mL sample (300 μL of protein solution, 500 μL water, 200 μL of buffer) 7.6 μL of Ellman's reagent (0.01 M) was added. Absorbance was monitored up to 600 seconds following addition of the reagent.

3.4 Results and Discussion

The leucine zipper helix in the recombinant proteins contains six heptad repeats, and so is 42 amino acids in length. The charge pattern at the *e* and *g* positions in the heptad repeat, shown as underlined amino acids in Figure 3.1, is largely derived from glutamic acid. In addition to the enrichment of acid functionality in the helical portion, the disordered coil domain contains a glutamic acid in one out of every nine amino acids. The triblock design is a combination of both helix and coil such that each chain will have

regions of high charge density (the helical domains) separated by regions of low charge density (the coil domain). More specifically, the high charge density region results from the localization of charges on the 7_2 helix where the *a* and *d* positions align on one face of the helix and the *e* and *g* positions are on either side of them. Adjacent helices will form favorable hydrophobic interactions but will have either charge-charge attraction or repulsion at the *e* and *g* positions, depending on pH. In the case of the *Helix* in **7** and **13**, it would be expected that the interactions of the ionized glutamic acid sidechains would destabilize the α-helix and reduce the amount of interchain helical-helical association. The opposite would be true for the lysine-glutamic acid interactions at the *e* and *g* positions. These latter interactions are believed to be minor contributors to the helical-helical stabilization because there are almost twice as many acid-acid interactions as there are acid-base interactions. Isoelectric point measurements indicate that a negative charge excess would exist at pH >5.7 for **7** and pH > 5.5 for **13**.

The secondary structures of the target proteins (5 μM) in aqueous salt solutions (pH 7.4, 10 mM Na_2HPO_4, 150 mM NaCl) were determined by circular dichroism spectroscopy (CD) over the wavelength range 195-250 nm (Fig. 3.3). The results show that **9** is predominantly random coil, as evidenced by the minimum in mean residue ellipticity at 201 nm. The CD scans of **7** and **13** have the characteristic α-helical amide transitions at 222, 208, and 190 nm. Methods for analyzing protein secondary structure in solution with circular dichroism spectroscopy have been reviewed by Greenfield (34). The accuracy in predicting specific protein conformations from test data depends upon the mathematical method used as well as the protein basis spectra chosen for each method. The CD spectrum of each protein in Figure 3.3 was analyzed by three different

methods: ridge regression, neural networks, and self-consistent method (described in the Experimental Section). The results of these methods are reported in Table 3.1 in addition to the expected α-helical conformational values. The expected α-helical content is calculated from the number of amino acids in the *Helix* region divided by the total number of amino acids in the protein. For protein **7** the "measured" (fit) α-helical content is within 20% of the predicted value while for **9** and **13** the values are within 10%. Both the ridge regression and neural network analyses predict higher α-helical contents for **7** and **13** than the self-consistent method. These data suggest the helix may propagate further into the chain than would be expected. In any case, it appears that the helicity of the coiled-coil domains of the triblock is not reduced by introducing the disordered coil regions between helical ends.

Table 3.1 Percent of α, β, and random conformations present in 5 μM solutions of **7**, **9**, and **13**.

Protein	Algorithm	α	β	random	error*	predicted α**
7	ridge regression	0.64	0.19***	0.17	0.008	0.55
7	neural network	0.78	0.01	0.21	0.227	0.55
7	self-consistent	0.60	0.16***	0.23	0.001	0.55
9	ridge regression	0.06	0.75***	0.20	0.013	0.00
9	neural network	0.06	0.44	0.50	0.182	0.00
9	self-consistent	0.08	0.64***	0.28	0.004	0.00
13	ridge regression	0.41	0.35***	0.24	0.022	0.37
13	neural network	0.47	0.14	0.39	0.182	0.37
13	self-consistent	0.35	0.35***	0.31	0.002	0.37

* Errors are averaged over α, β, and random conformations and are reported as variance for ridge regression, standard deviation for self-consistent, and mean error for neural network.

** Values calculated from number of amino acids in the helical region divided by the total number of amino acids in the chain.

*** Value includes calculations for both β-sheet and β-turn.

Gelation through physical crosslinking requires association of the coiled-coil elements of the triblock. In order to determine whether the unfolding behavior of the helical end blocks is affected by attachment of the coil domain, the pH dependence of the transition temperature, T_m, was measured for both **7** (5 μM) and **13** (5 μM) in solution. The similarity of the unfolding processes is shown in Figure 3.4 and in Table 3.2. At pH values greater than 10, the unfolding transitions of both **7** and **13** occur between 30 and 40°C. The low T_m is expected in basic solution in which the glutamic acid residues are deprotonated and charge-charge repulsion of the helices will occur. In addition, the helical content at 0 °C is reduced from 22055 (pH 7.6) to 16835 (pH 11.5) deg cm^2 $dmol^{-1}$ for **7** and from 13722 (pH 7.6) to 10294 (10.8) deg cm^2 $dmol^{-1}$ for **13**. This would imply that the individual helices are partially unfolded at highly basic conditions and low temperatures. When the pH is lowered to values between 7 and 10, the T_m values increase slightly but are within 10 °C of each other. In the range of pH 7-10 and lower, the lysine (pKa =10) is charged and can form interhelical salt bridges with the glutamic acid. Upon further decrease of pH to 6, the T_m for **7** and **13** occurs at 88°C and 81°C respectively. At pH 5 and lower, no transitions are observed for either protein at temperatures below 100 °C and the helical content decreases only slightly throughout the 0-100 °C temperature range. Acidic conditions effectively stabilize the helix so that no thermal denaturation of **1** or **3** is observed.

Table 3.2 T_m values as a function of pH for proteins 7 and **13**.

Protein	pH	T_m(°C)	$-[\theta]_{222}$ 0 °C
7	6.1	88	21385
7	6.9	64	22143
7	7.6	51	22055
7	9.4	50	21979
7	10.7	40	20674
7	11.5	31	16835
13	6.1	81	14218
13	6.9	55	14125
13	7.6	54	13722
13	9.5	49	11888
13	10.8	31	10294

The thermal denaturation experiments described above were measured in solutions containing 150 mM NaCl. There have been many observations and discussions about the effect that NaCl has on the stability of coiled-coil interactions in solution (39-43). The most important observation is that while the addition of NaCl (> 100mM) has no significant effect on coiled-coil stability at neutral pH, the stability of the coiled-coil increases drastically in acidic environments (43). Similar behavior was observed for the helical domains in this study. Experiments show that a T_m is observed for protein 7 at 85 °C in 10 mM Na_2PO_4, pH 2.1 (without NaCl). No thermal transitions at 222 nm are observed from 0 to 100 °C for either 7 or **13** at pH 2.1. The lack of a thermal transition at before 100 °C suggests that the helices are stabilized with the addition of NaCl in acidic environments. It is our intent to investigate the solution behavior of 7 and **13** under physiological conditions and therefore, the addition of NaCl should not significantly affect the thermal stability of the helices in the neutral pH range.

The concept of reversible gelation as a function of pH in this work is based on the separation of charge along the backbone of the triblock protein. For instance, if the

protein chains are in a low pH solution, the sidechain glutamic acids would be protonated, resulting in an increased stability of the coiled-coil regions. Precipitation would occur because highly associated oligomer states of the helical domains can be formed. The coil domain would also have protonated acids which would cause it to collapse to a hydrophobically packed globule. As the pH of the solution increases, deprotonation of the glutamic acids in the coil domain would allow swelling of the flexible spacer while some deprotonation of the acid groups in the helical domain would not disrupt the helical-helical packing. The protein would behave as a noncovalent gelled network where the reversibility with temperature in the gelled state results from thermal dissociation of the leucine zippers regions. Finally, if the gelled solution is increased to a high enough pH, all of the glutamic acids become deprotonated in the helical domains. This decreases the state of association of the helices so that the proteins flow as the chains are free to slip past one another. This proposed gelation process is depicted in a general schematic in Figure 3.5.

The solution data serve as guidelines for obtaining the conditions for gelation. In these experiments, T_m values over 37°C are of interest because the applicability of using this gel as a delivery vehicle in the body or for application on the body will require that the material is not affected at body temperature. In order to investigate the pH and temperature dependence of gelation, dynamic light scattering experiments were done. Diffusing wave spectroscopy (DWS) (44-48) was used to measure the viscoelastic behavior of **13**. The measurement obtained is the dynamic intensity auto correlation function of light that is multiply scattered from a suspension of tracer particles (0.3 μm sulfonated polystyrene latexes). The calculation of the autocorrelation function in DWS

requires two approximations to be made. The first approximation is that as light propagates through a highly scattering medium, each photon is scattered a large number of times. The propagation of light is described by the diffusion approximation where all interference effects within the medium are neglected and the phases of all photons are completely randomized (47). The second approximation is that the conservation of scattering momentum at every point along the full path can be neglected and therefore, the individual scattering events are approximated by the contribution of an average scattering event (48).

From the autocorrelation function, the time average mean-square displacement of the tracer particles as a function of time is determined. In addition, the shear creep compliance can be obtained with the following relationship (49),

$$J(t) = (\pi a / kT) \langle r^2(t) \rangle \qquad \{2\}$$

where a is the radius of the bead and $\langle r^2(t) \rangle$ is the average mean square displacement of the beads (μm^2).

For a viscous liquid, the correlation function is a decreasing, convex function of time while for an elastic medium, there is a plateau in the correlation function (49). A viscoelastic medium will show some aspects of each limiting behavior. The correlation function was measured for protein **13** (5% (w/v), 10 mM Tris buffer, 25°C) at pH 8.0, 8.8, and 9.5 (not shown). The corresponding mean square displacement curves are shown in Figure 3.6. These data suggest that the gel point lies between pH 8.0 and 9.5. At pH 8.0, **13** is an elastic gel, at pH 8.8, a viscoelastic gel, and at pH 9.5, a viscous

liquid. The elastic behavior in these curves is shown in the plateau region over time scales of 1 x 10^{-5} to 1 sec. The plateau region reflects that the mean square displacement of the tracer particles is relatively small during this time scale. In an elastic medium, particles will not move very far from their starting points. On the other hand, the tracer particles are not hindered in a viscous fluid and the mean square displacement will grow with time.

The thermal dependence of gelation for **13** was recorded by cycling a 5% gel through a series of temperatures ranging from 23 °C to 55 °C at pH 7.8. The average mean square displacement versus time at various temperatures is plotted in Figure 3.7. These data were acquired in the following temperature order: 23, 39.6, 47.7, 55.2, 42.7, 24.9 and 23 °C. These data show that as the temperature is increased to 55°C, the gel becomes a fluid. An interesting observation in the thermal analyses is the hysteresis that occurs after the material is heated to 55 °C and returned to 23 °C (Figure 3.8). Most likely, there is some time-dependence for the helices to re-fold into the original associated state. The two to three hour equilibration times allotted between six hour experiments might not have been long enough for the gel to return to its native state.

The dependence of gel formation as a function of concentration was also investigated for **13** at pH 7.9. The DWS results show that **13** becomes a fluid at 3% (w/v) whereas at 5%, **13** is an elastic medium (Figure 3.9). The 4% sample responds as a viscous liquid at early times but at intermediate times, the emergence of a plateau is apparent. Therefore, the critical concentration for forming an elastic gel that spans the entire volume of interest is between 4-5%. When concentrations are between 3-4%, the

protein is most likely forming soluble aggregates that viscosify the solution by increasing the apparent molecular weight.

The pH, temperature, and concentration behavior of **13,** as determined by DWS experiments, indicates that these materials are not simple gels. At very short times ($t \leq 10^{-6}$ sec), the correlation function, thus average mean square displacement, for any of the above experiments does not give any information about the gel behavior. The tracer particles are probing very short length scales and local fluctuations at these times. On the other hand, viscous-like behavior is observed for all of the experiments in the very long time limit. The existence of a plateau in the correlation function provides evidence for gel behavior at some range of times. With further experiments, this time range will be evaluated.

The contribution of the disulfide linkage to gelation was investigated for protein **13** where one C-terminus cysteine exists on each protein chain. The fraction of free thiol in a solution of **13** was determined by the Ellman reaction (50). A linear increase in absorbance with increase in concentration was observed for the control, β-mercaptoethanol. With **13**, no change was observed in absorbance as the concentration of protein was increased. This suggests that **13** is fully oxidized in solution under physiological conditions. It is then expected that the thiol groups remain oxidized in the conditions used to prepare gels of **13**. Introduction of 10 mM β-mercaptoethanol into a 5% (w/v) gel does not cause dissolution of the gel; therefore, it can be concluded that the disulfide linkage of the protein chains are not responsible for gelation.

From the previous discussion, the driving force for gelation is believed to be the non-covalent association of α-helical coiled-coil end blocks that contain both hydrophobic and electrostatic attractive forces. The formation of the gel of protein **3** can be made at concentrations greater than 4%(w/v). In the gelation process, there are two separate reversibility issues to consider; one with pH and one with temperature. This reversible process results from the pH dependent electrostatic interactions of the α-helices as well as extension and contraction of the disordered coil domain. The thermoreversibility of gelation depends on the energy needed to break the non-covalent hydrophobic and electrostatic interactions.

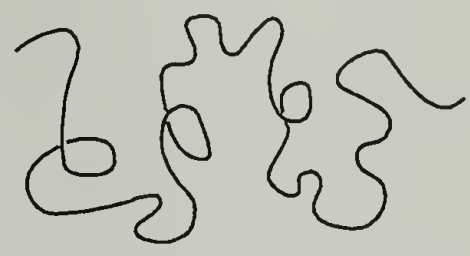
In most synthetic polymer systems, such as acrylamide, methacrylate, or cellulose derivatives, the gel behavior and solubility of the system is an additive function of the side groups (51). The solubility behavior of the protein described in this work is most likely dominated by secondary structure. This statement is based on the fact that the helix-coil transition depends on temperature as well as pH and this is what drives the solubility-induced transitions seen in the gel of **13**.

3.5 Conclusions

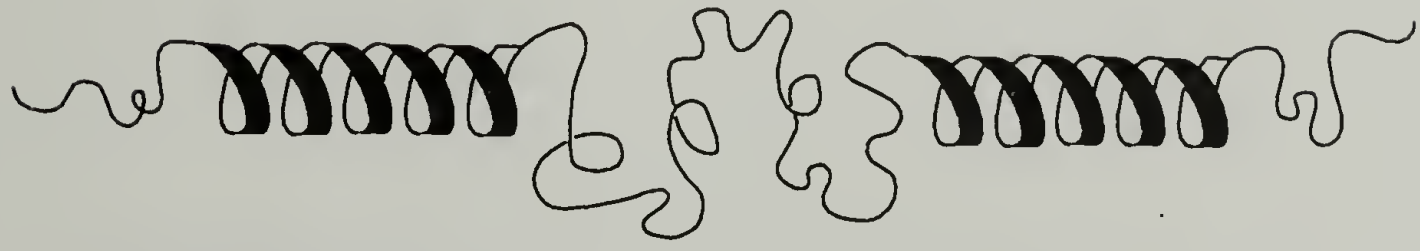
It has been shown that genetic engineering principles can be applied to the synthesis of a protein-based materials that have defined structure and function. The triblock protein, **13**, forms an elastic physical gel at physiological pH and temperature conditions whereas the control protein blocks do not. The advantages of the triblock design are that different coiled-coil domains of various stabilities can be introduced as

the physical crosslinks and the flexible spacer can be shortened or lengthened to have precise control over the pore size of the gel. The charge density of the flexible spacer can also be changed in order to compare solubility as well as swelling behavior. These are future directions that are currently being pursued in our group.

MRGSHHHHHHGSDDDDKWA- *Helix* -IGDHVAPRDTSMGGC (7)



MRGSHHHHHHGSDDDDKASYRDPMG- $[(AG)_3PEG]_{10}$ -ARMPTSW (9)



MRGSHHHHHHGSDDDDKWA- *Helix* -IGKHVAPRDTSYRDPMG- $[(AG)_3PEG]_{10}$

ARMPTSGD- *Helix* -IGDHVAPRDTSMGGC (**13**)

Helix = SGDLENEVAQLEREVRSLEDEAAELEQKVSRLKNEIEDLKAE

Figure 3.1 Amino acid sequences of the three proteins studied. **(7)** consists of 76 amino acids of which 42 are of the leucine zipper, *Helix*. **(9)** consists of 122 amino acids where 90 amino acids are of the alanylglycine repeat, $[(AG)_3PEG]_{10}$. **(13)** consists of 230 amino acids of which 84 are of the *Helix* repeat and 90 are of the alanylglycine repeat. Charge patterns in the *e* and *g* positions of the *Helix* heptad repeat, $(a\,b\,c\,d\,e\,f\,g)_6$ are underlined. Abbreviations for the amino acids are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val and W, Trp.

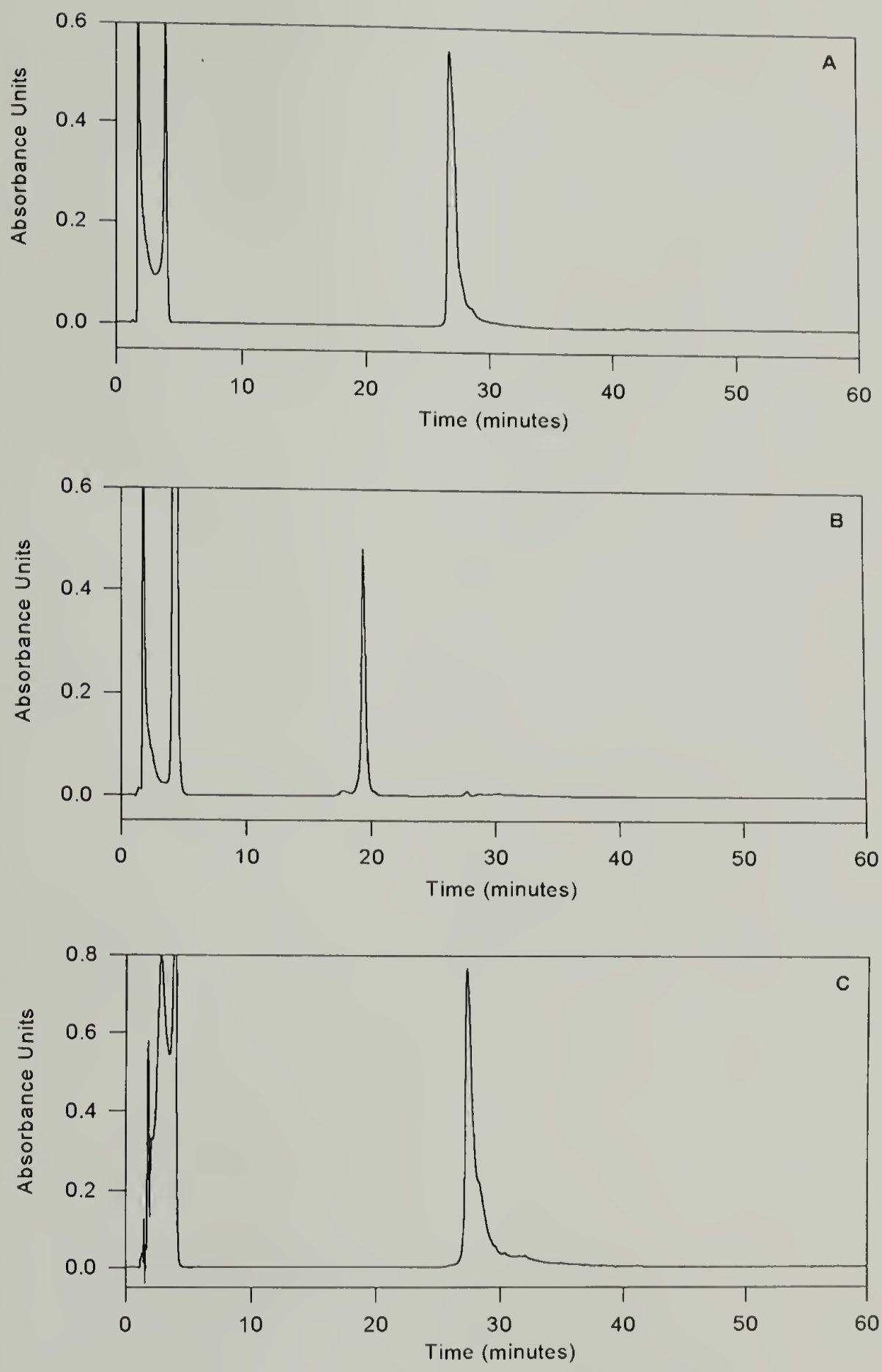


Figure 3.2 Reversed phase high performance liquid chromatograms of genetically engineered proteins. **(A)** 7 (76 μg) eluted at 26.76 minutes. **(B)** **9** (54 μg) eluted at 19.36 minutes. **(C)** **13** (54 μg) eluted at 27.33 minutes.

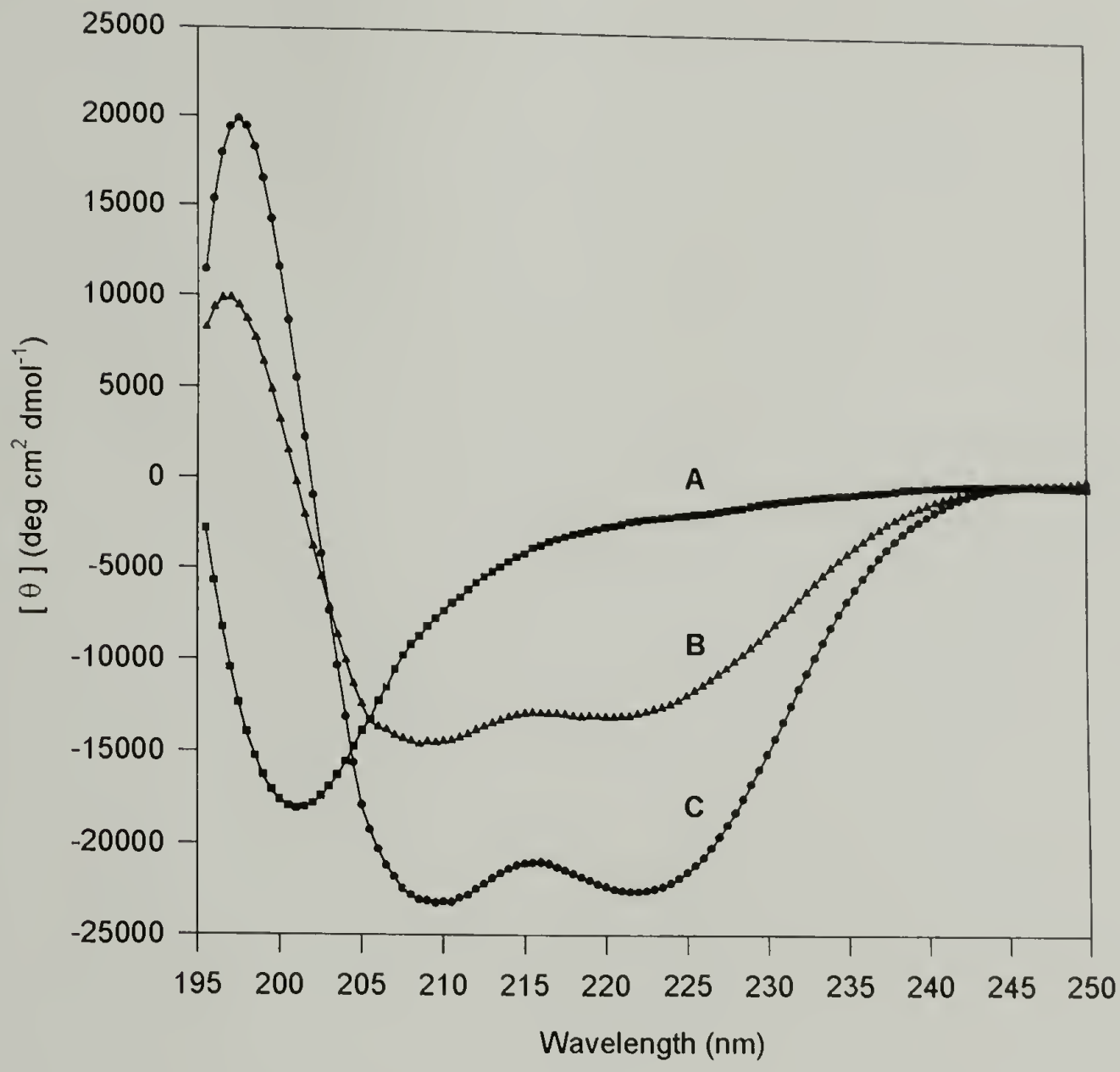


Figure 3.3 Secondary structural analyses using circular dichroism spectroscopy. **(A) 9** (5 μM). **(B) 13** (5 μM). **(C)** 7 (5 μM). Spectra were recorded at 25°C in 10 mM NaH_2PO_4, 150 mM NaCl, pH 7.4 (upon addition of 1 N NaOH).

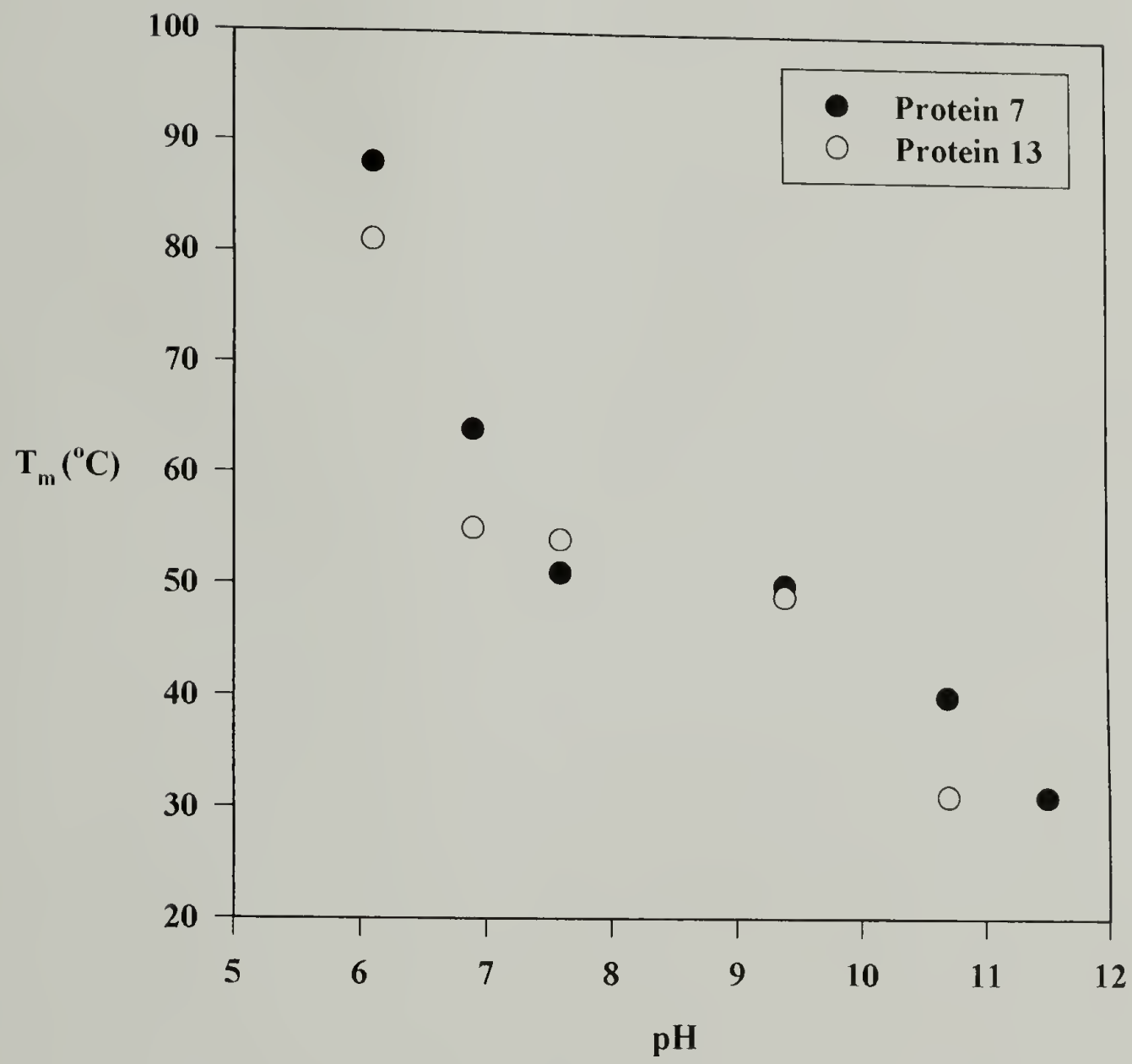


Figure 3.4 Thermal melting transitions of 7 (5 μM) and **13** (5 μM) monitored at 222 nm as a function of pH. Measurements were made in 10 mM NaH_2PO_4, 150 mM NaCl (pH 7.6). T_m was determined by taking the first derivative of the CD signal $[\theta]_{222}$ with respect to temperature^{-1} (K) and finding the minimum of this function.

Figure 3.5 General schematic of the physical gelation process of monodisperse triblock copolymer **13**.

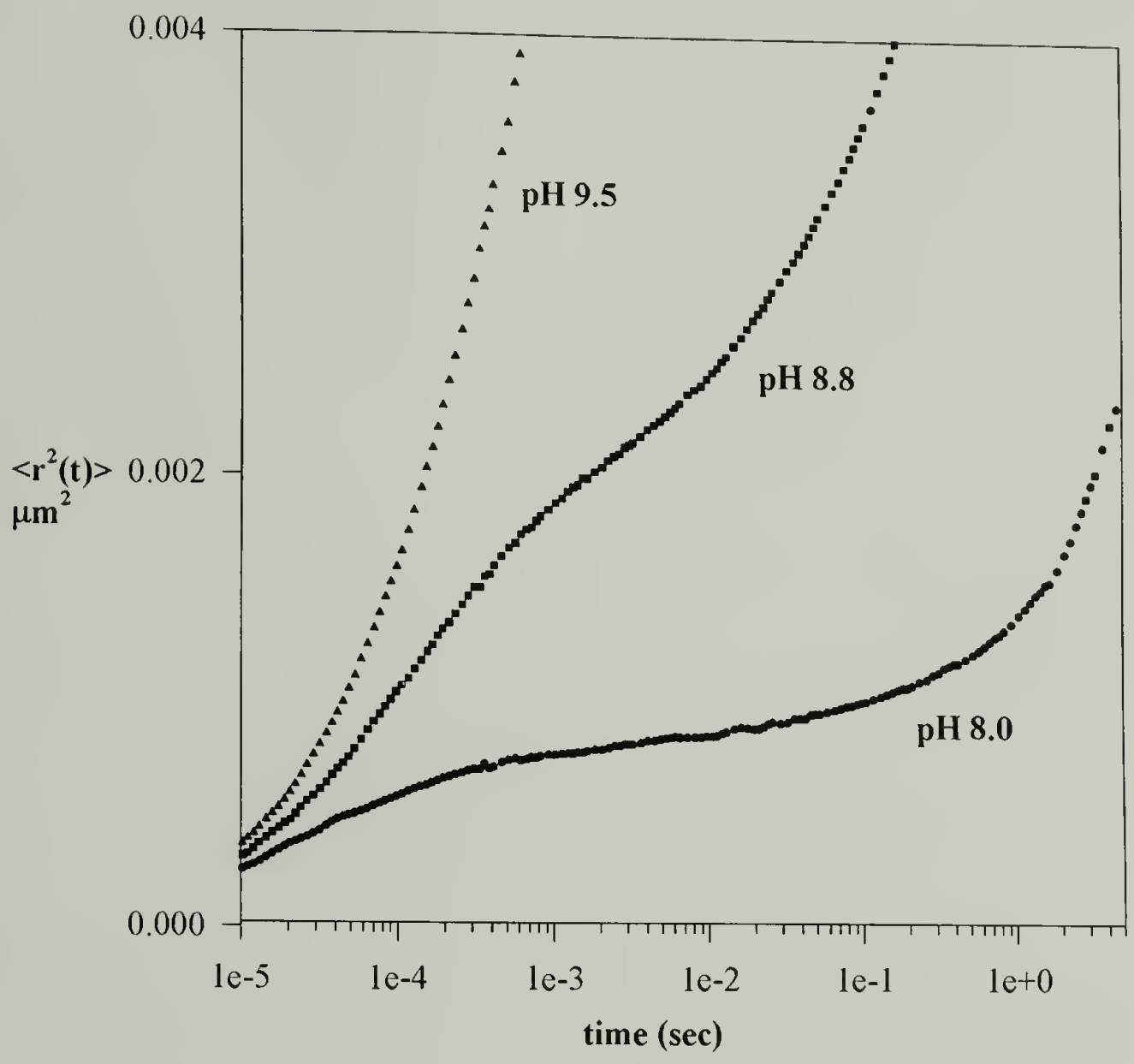


Figure 3.6 Mean square displacement as a function of time in a linear-log representation for 5% (w/v) of protein **13** in 10 mM Tris buffer at pH 8.0, 8.8, and 9.5.

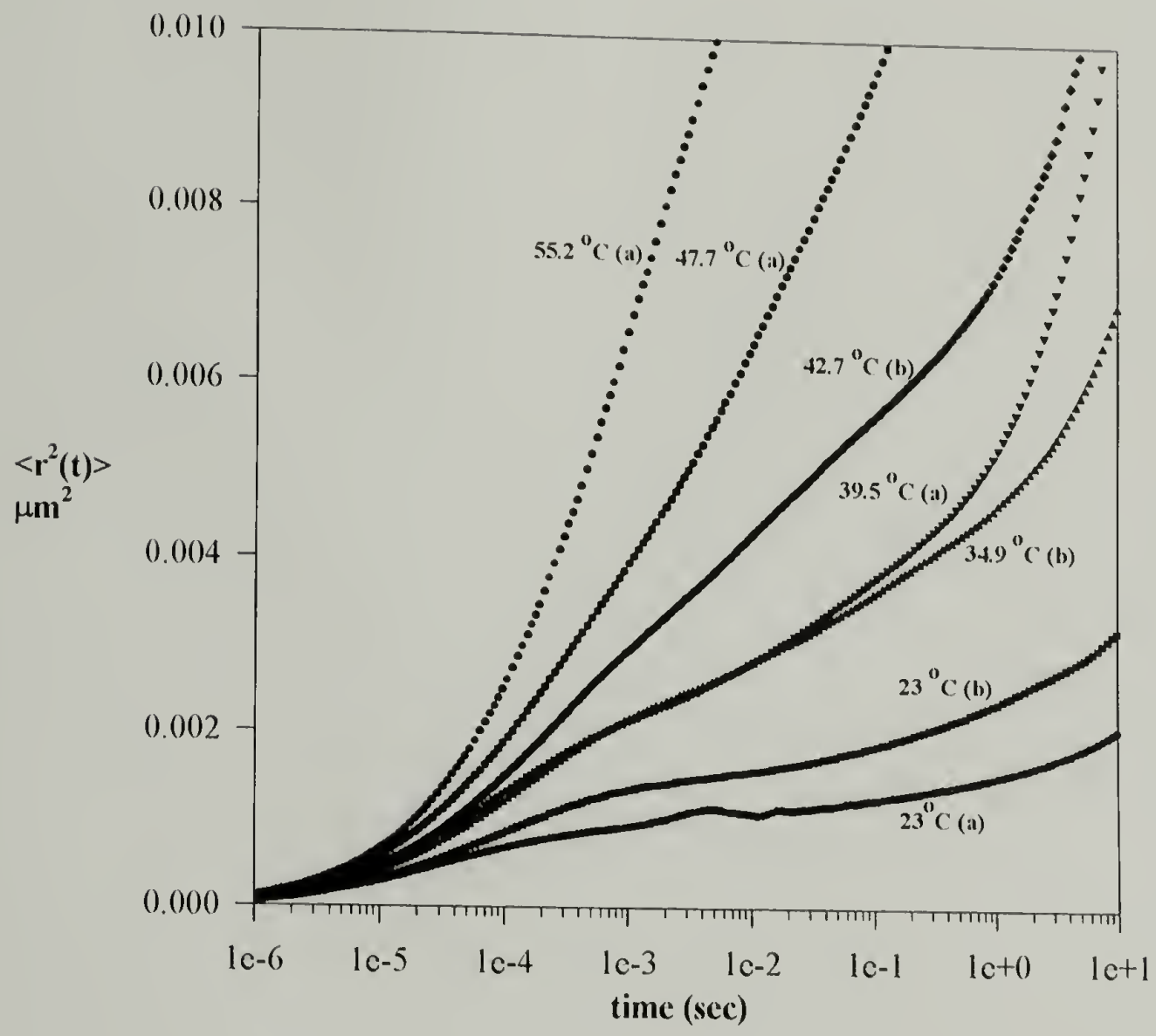


Figure 3.7 Mean square displacement as a function of time at temperatures 23 °C to 55 °C for protein **13** (5% (w/v), 10 mM Tris buffer, pH 7.8). **(a)** signifies data collected upon heating the sample from 23 °C to 55.2 °C and **(b)** signifies the data collected on cooling the sample from 55.2 °C to 23 °C.

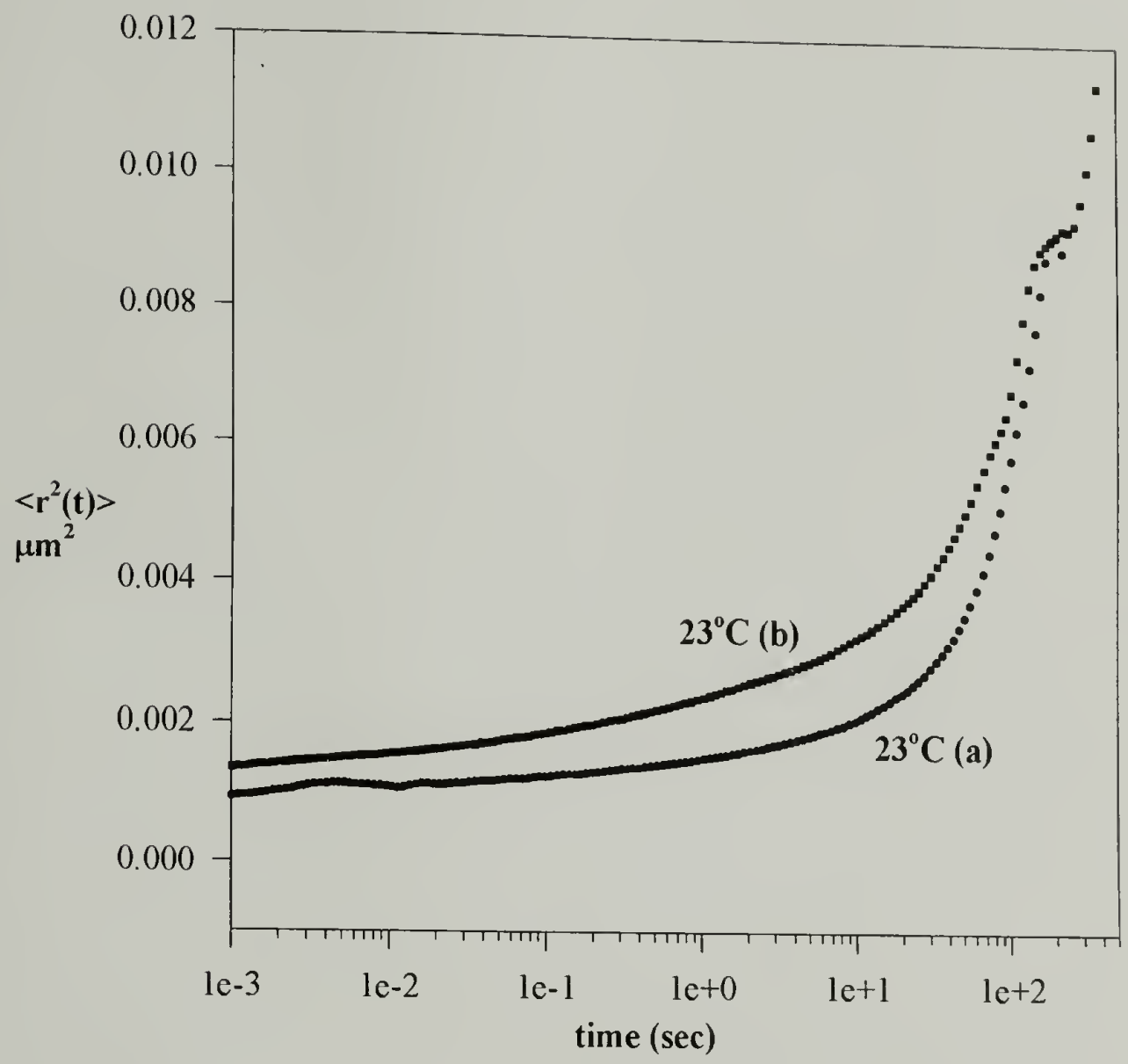


Figure 3.8 Mean square displacement as a function of time for **13** at temperatures **(a)** 23 °C before thermal cycling and **(b)** 23 °C after thermal cycling.

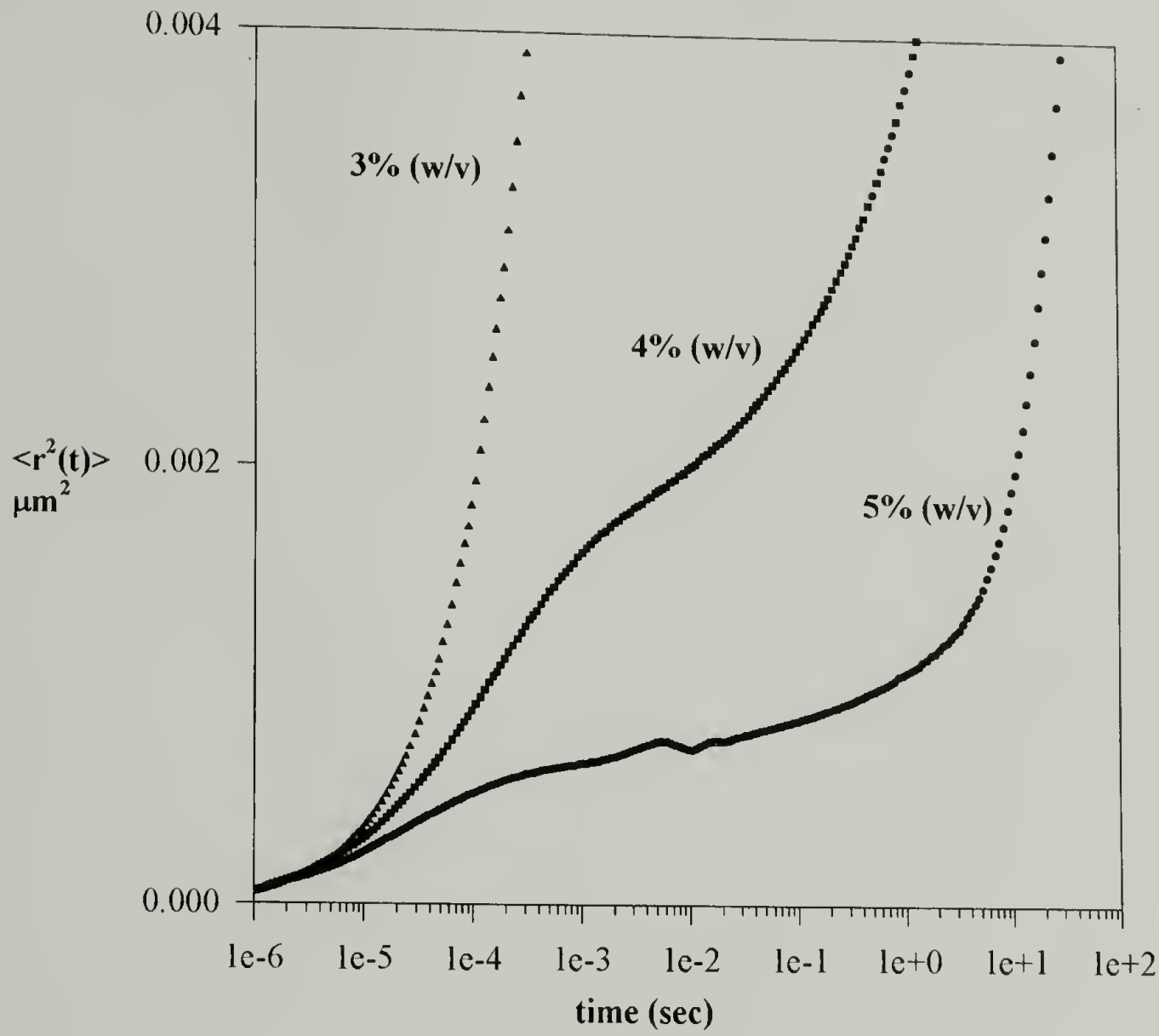


Figure 3.9 Concentration dependence of the mean square displacement as a function of time for **13** at pH 7.9 and 23 °C.

3.6 References

(1) Yoshida, R.; Uchida, K.; Kaneko, Y.; Sakai, K.; Kikuchi, A.; Sakurai, Y.; Okano, T. *Nature* **1995**, *374*, 240-242.

(2) Lim, F.; Sun, A. M. *Science* **1980**, *210*, 908.

(3) Hoffman, A. S. *Chemtech* **1986**, 426.

(4) Cohen, S.; Bañó, M. C.; Vissher, K. B.; Chow, M.; Allcock, H. R.; Langer, R. *J. Am. Chem. Soc.* **1990**, *112*, 7833.

(5) Pathak, C. P.; Sawhney, A. S.; Hubbell, J. A. *J. Am. Chem. Soc.* **1992**, *114*, 8311.

(6) Peppas, N. A.; Langer, R. *Science* **1994**, *263*, 1715.

(7) Chen, G.; Hoffman, A. S. *Nature* **1995**, *373*, 49.

(8) Klier, J.; Scranton, A. B.; Peppas, N. A. *Macromolecules* **1990**, *23*, 4944.

(9) Annaka, M.; Tanaka, T.; *Nature* **1992**, *355*, 430.

(10) Annaka, M.; Tanaka, T. *Physica A* **1994**, *204*, 40.

(11) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

(12) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538.

(13) Mitchell, P. J.; Tjian, R. *Science* **1989**, *245*, 371.

(14) Turner, R.; Tjian, R. *Science* **1989**, *243*, 1689.

(15) Bohmann, D.; Bos, T. J.; Admon, A.; Nishimura, T.; Vogt, P. K.; Tjian, R. *Science* **1987**, *238*, 1386.

(16) Gentz, R.; Rauscher III, F. J.; Abate, C.; Curran, T. *Science* **1989**, *243*, 1695.

(17) Schuermann, M.; Hunter, J. B.; Hennig, G.; Müller, R. *Nucleic Acids Research* **1991**, *19*, 739.

(18) Lau, S. Y. M.; Taneja, A. K.; Hodges, R. S. *J. Biol. Chem.* **1984**, *259*, 13253.

(19) Hodges, R. S.; Saund, A. K.; Chong, P. C. S.; St.-Pierre, S. A.; Reid, R. E. *J. Biol. Chem.* **1981**, *256*, 1214.

(20) Hodges, R. S.; Zhou, N. E.; Kay, C. M.; Semchuk, P. D. *Peptide Research* **1990**, *3*, 137.

(21) Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419.

(22) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1992**, *267*, 2664.

(23) Graddis, T. J.; Myszka, D. G.; Chaiken, I. M. *Biochemistry* **1993**, *32*, 12664.

(24) Monera, O. D.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1993**, *268*, 19218.

(25) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol.* **1994**, *237*, 500.

(26) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Cell* **1992**, *68*, 699.

(27) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.

(28) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.

(29) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett* **1983**, *24*, 245.

(30) Yanisch-Perron, C.; Vieira, J.; Messing, J. *Gene* **1985**, *33*, 103.

(31) Greenfield, N. J. *Anal. Biochem.* **1996**, *235*, 1.

(32) Provencher, S. W.; Glöckner, J. *Biochemistry* **1981**, *20*, 33.

(33) Venyaminov, S. Y.; Baikalov, I. A.; Shen, Z. M.; Wu, C.-S. C.; Yang, J. T. *Anal. Biochem.* **1993**, *214*, 17.

(34) Sreerama, N.; Woody, R. W. *Analytical Biochemistry* **1993**, *209*, 32.

(35) Sreerama, N.; Woody, R. W. *J. Mol. Biol.* **1994**, *242*, 497.

(36) Johnson, W. C., Jr. *Proteins: Struc. Funct. Genet.* **1990**, *7*, 205.

(37) Andrade, M. A.; Chacon, P.; Merolo, J. J.; Moran, F. *Protein Eng.* **1993**, *6*, 383.

(38) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*; W. H. Freeman and Company: San Francisco, 1980.

(39) McLachlan, A. D.; Stewart, M. *J. Mol. Biol.* **1975**, *98*, 293.

(40) Lumb, K. J.; Kim, P. S. *Science* **1995**, *268*, 436.

(41) Krylov, D.; Mikhailenko, I.; Vinson, C. *EMBO J.* **1994**, *13*, 2849.

(42) Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8898.

(43) Yu, Y.; Monera, O. D.; Hodges, R. S.; Privalov, P. L. *J. Mol. Biol.* **1996**, *255*, 367.

(44) Maret, G.; Wolf, P. E. *Z. Phys. B* **1987**, *65*, 409.

(45) Fraden, S.; Maret, G. *Phys. Rev. Lett* **1990**, *65*, 512.

(46) Pine, D. J.; Weitz, D. A.; Zhu, J. X.; Herbolzheimer, E. *J. Phys. France* **1990**, *51*, 2101.

(47) Weitz, D. A.; Zhu, J. X.; Durian, D. J.; Gang, H.; Pine, D. J. *Physica Scripta* **1993**, *T49*, 610.

(48) Weitz, D. A.; Pine, D. J. in *Dynamic Light Scattering, The Method and Some Applications*, Brown, W. Ed.; Clarendon Press: Oxford, 1993.

(49) Mason, T. G.; Gang, H.; Weitz, D. A. *J. Mol. Struct.* **1996**, *383*, 81.

(50) Ellman, G. L. *Archives of Biochemistry and Biophysics* **1959**, *82*, 70.

(51) Gehrke, S. *Personal Communication* .

CHAPTER 4

INCORPORATION OF 5',5',5'-TRIFLUORO-L-LEUCINE AT THE HYDROPHOBIC INTERFACE OF LEUCINE ZIPPERS

4.1 Objectives

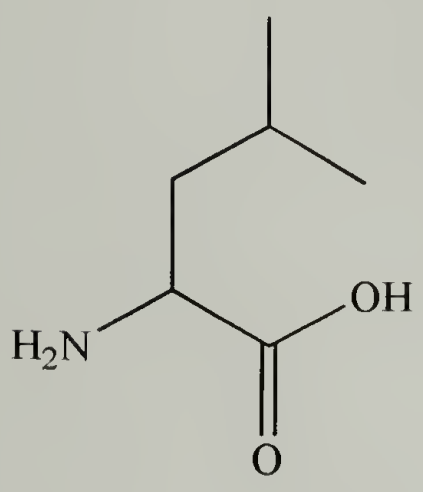
The stabilization of coiled coil structures is believed to be modulated by interactions between regularly spaced hydrophobic and charged groups that are on adjacent associated helices. In particular, it is believed that the hydrophobic residues at the *a* and *d* position in the general heptad repeat, $(a\ b\ c\ d\ e\ f\ g)_n$, determines the global folding of the helices while the polar groups at positions *e* and *g* stabilize the helices through salt bridges. To evaluate the effect of substitution of the hydrophobic residues at position *d*, an unnatural amino acid analogue, 5',5',5'-trifluoro-L-leucine (Tfl), was incorporated in place of leucine (Leu) in the recombinant leucine zipper coiled coil proteins, Recognins A1 (A1) and B1 (B1) (Figure 4.1). An investigation of the incorporation of Tfl into the recombinant proteins with bacterial hosts auxotrophic for leucine will be presented. Also to be discussed is the isolation of the Tfl containing chains and the effect of Tfl incorporation on the thermal stabilities of the interacting α-helices.

4.2 Introduction

The incorporation of non-natural amino acids into protein-based materials is one way to introduce new chemical functionality and physical properties into natural polymer systems. The syntheses of these kinds of materials provide an opportunity to explore biological changes or activities that result from the substitution of non-natural amino acids for natural ones (1-3) and to study unique structural and functional properties that are associated with these materials. Amino acid analogs can be introduced into cellular proteins either by direct incorporation during protein synthesis or by post-translational modification. Overall, there are three pathways to incorporate non-natural amino acids into proteins: *in vivo* biological synthesis (4), *in vitro* chemical and biological synthesis (5-13), and chemical synthesis by solution or solid-phase methods (14, 15). There has been an ongoing research effort in our group to utilize the *in vivo* biological synthetic route to incorporate non-natural amino acids into polymer chains (16-21) .

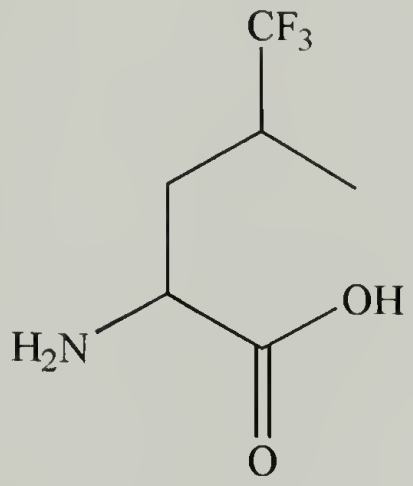
One non-natural amino acid that is incorporated into *E. coli* proteins in the absence of its natural analog (**1**) is 5',5',5'-trifluoroleucine (**2**) (22-24) shown below. Fluorine mimics the geometry of hydrogen (van der Waals radii of fluorine 1.35Å to hydrogen 1.2Å) while differing in chemical character (electronegativity of fluorine 3.98 versus 2.2 for hydrogen). The lipophilic nature of the trifluoromethyl group makes it an ideal candidate for the modification of membrane spanning or hydrophobically associating proteins. Tfl imparts unique chemical and physiological stability to the non-natural amino acid containing proteins (25, 26) while the fluorine atom can be used in ^{19}F

NMR to monitor structural changes in proteins (27) or to study the motion of fluorinated proteins within lipid bilayers (28).



Leucine (Leu)

1



5',5',5'-Trifluoroleucine (Tfl)

2

One class of proteins known to associate through hydrophobic interactions are the coiled-coil proteins (29). These proteins contain a seven amino acid repeat, $(a\ b\ c\ d\ e\ f\ g)_n$, where the *a* and *d* positions are non-polar amino acids, with *d* being mostly leucine. The result of this 3-4 hydrophobic repeat is a characteristic 7_2 α-helix where the *a* and *d* positions align on one face of the helix so that it can hydrophobically associate with adjacent helices. Coiled-coil proteins are good candidates for structural modification since they are found in over 200 proteins in the body and are involved in many recognition processes (30). Their roles range from gene regulation such as the yeast transcription factor, GCN4 (31), to membrane fusion and infection processes of human immunodeficiency virus type I (32). The substitution of leucine (Leu) with 5',5',5'-trifluoroleucine (Tfl) in coiled-coils would be one way of modifying residues that

are believed to be the driving force for the association of α-helices through hydrophobic interactions (33, 34).

Coiled-coil proteins being used for the incorporation of Tfl are shown in Figure 4.1. These proteins were prepared through genetic engineering techniques and were designed to form stable water soluble coiled-coils with leucines at every *d* position as found in the *Jun* oncogene protein (35). An algorithm developed by Lupas *et al.* (30) was used to identify the most probable amino acids positioned at *a*, *b*, *c*, and *f* found in naturally occurring coiled-coils. The sequence of amino acids at these positions were also chosen based on their ability to increase water solubility of the chain and to make the corresponding DNA manipulations easy. Acidic or basic groups, glutamic acid and lysine, were placed at the *e* and *g* positions to provide intermolecular attraction or repulsion of chains at varying pH.

The incorporation of Tfl into the helical chain at the *d* position of the heptad repeat, which normally encodes leucine during protein synthesis, was accomplished by using a leucine auxotrophic strain of *E. coli* and media shifting techniques developed by Kothakota *et al.* (19). We rationalized that by modifying the α-helical hydrophobic face with trifluoromethyl groups we would be able to increase the stability of the interacting helices in solution. This work reports the syntheses of Tfl chains A1 and B1 containing a population of eight possible Leu position being substituted. The level of Tfl incorporation and initial thermal stability measurements of the α-helices will be discussed.

4.3 Experimental Section

4.3.1 Materials

5',5',5'-Trifluoroleucine (mixture of stereoisomers) was purchased from MTM Research Chemicals (Windham, NH). Leucine auxotrophs of strain mc1000 (F^- lacΔx74anaD139/(Ana Abioc-leu)Δ7679 gal U gal K rspL) containing pREP4 plasmid (*Qiagen*™, Chatsworth, CA) were obtained from Professor Fournier's Lab, Department of Biochemistry and Molecular Biology (University of Massachusetts, Amherst). The stable auxotroph is designated JM-1. Expression vectors containing A1 and B1 DNA sequences (pQE9-A1 and pQE9-B1) were obtained from the U.S. Army Natick RD&E Center (Natick, MA).

4.3.2 General Methods

Electrospray mass spectra were collected on a *Fisons* VG Platform II™ electrospray mass spectrometer. A1 and B1 protein spectra were also recorded as standards. All samples were run in 50:50 acetonitrile:water with 0.2% formic acid at 12 pmol/μL concentrations. Data was recorded with a positive ion mode (full scan 600-1500 m/z), an infusion rate of 5μL/min, and cone voltages of either 35 or 50 V. A *Waters*™ high performance liquid chromatography (HPLC) system (717 plus Autosampler, 486 Tunable Absorbance Detector, and 600 Controller) was used to inject 50 μL samples at flow rates of 1 mL/min onto a *Vydac*™ C18 reversed phase column.

Millenium™ software was used in the analysis of the collected spectra. Detection at 210 nm was accomplished over a 60 minute gradient of water/0.1% trifluoroacetic acid, TFA to acetonitrile/0.1% TFA. This gradient consisted of 90:10 (10 min), 50:50 (20 min), and 100:0 (30 min). Amino acid analysis was performed by the Analytical Chemistry and Peptide/DNA Synthesis Facility at Cornell University, Ithaca, NY. Amino acid analysis was carried out on a Pico-Tag Amino Acid Analysis System (36). Duplicate samples were hydrolyzed in constant boiling HCl at 150 °C for 110 min. Appropriate blanks, controls and standards were hydrolyzed in the same vessel as a batch hydrolysis. The amino acids were derivatized with PITC and analyzed by reversed phase HPLC (37, 38). The resulting phenylthiocarbamoyl amino acid derivatives were separated on a 4.6 x 300 mm Nova Pack C18 column employing a modified Pico-Tag buffer system. Circular dichroism spectra were recorded on an Aviv 62DS spectropolarimeter (Lakewood, NJ). Mean residue molar ellipticity reported at 222 nm ($[\theta]_{222}$, deg cm^2 $dmol^{-1}$) was calculated from the following equation: $[\theta] = [\theta]_{obs} \times MRW/(c \times l)$, where $[\theta]_{obs}$ is ellipticity measured in millidegrees, MRW is the mean residue molecular weight or the molecular weight of the peptide divided by the number of amino acid residues, c is the peptide concentration in g/L determined by quantitative amino acid analysis, and l is the optical path length in mm. Thermal melting curves were determined by monitoring the CD signal of the protein solutions (3 μM) at 222 nm as a function of temperature. Data was collected from 0 to 90 °C in 1 °C increments with an equilibration time of 1 minute. Spectra were collected in 50 mM sodium phosphate buffer, 100 mM NaCl, pH 7.4. Two Peltier units on both sides of the 1 cm path length rectangular cell (Helma) were used to regulate the temperature to within ±0.2 °C. A two

state mechanism was used to describe the equilibrium between folded and unfolded protein states. From this two state model, thermal melting temperature, T_m, was calculated as the temperature at which 50% of the helix is unfolded. To get this value, ellipticity readings were normalized to the fraction of the protein folded using the standard equation:

$$f_n = ([\theta]-[\theta]_u)/([\theta]_n-[\theta]_u) \qquad \{1\}$$

where $[\theta]_n$ and $[\theta]_u$ represent the ellipticity values for the fully folded and fully unfolded species, respectively, at 222 nm. All thermal melting transitions were determined to be reversible (± 2 °C) with 96% (A1), 88% (B1), 99% (Tfl A1), and 97% (Tfl B1) of the initial helix contents recovered.

4.3.3 Protein Expression and Purification

pQE9-A1 and pQE9-B1 were used to transform strain mc1000 (F^- lacΔx74anaD139/(Ana Abioc-leu)Δ7679 gal U gal K rspL) containing pREP4 to yield JM-1/pQE9-A1 and JM-1/pQE9-B1. Single colonies of cells were grown to saturation in 5 ml of 2xYT medium (16 g Bacto-Tryptone, 10 g Yeast Extract, 5 g NaCl per liter) containing ampicillin (200 μg/ml) and kanamycin (25 μg/ml) overnight at 37°C. Each of these cultures was transferred to 1 L of sterilized YT medium containing the same concentrations of antibiotics. The cells were incubated until optical densities (600 nm) of 1 were reached and then centrifuged at 5000 rpm for 10 minutes. The supernatant was then decanted and the cells were washed with 1 L of 1% M9 salt solution (60 g Na_2HPO_4, 30 g KH_2PO_4, 5 g NaCl, 10 g NH_4Cl per liter). Again the cells were pelleted

and resuspended in 2L of M9 medium (10% v/v M9 salt solution, 10% 19 amino acid solution (0.4 g of each amino acid, except leucine, per liter), 0.1% $MgSO_4 \cdot 7H_2O$ (1M), 0.1% $CaCl_2$ (0.01M), 0.1% vitamin B_1, 4% glucose (20% w/v solution), 0.33% ampicillin (200 μg/ml), 0.16% kanamycin (25 μg/ml) and 75.3% distilled deionized water). Leucine (40 mg/mL ; final concentration of 40 μg/mL) or 5',5',5'-trifluoro-D,L-leucine (40 mg/mL ; final concentration of 40 μg/mL) was added in addition to the M9 medium. The cells were grown for 15 minutes before isopropyl-β-thiogalactoside (IPTG, 1mM) was added to induce protein synthesis. Cell cultures were removed and centrifuged at 5000 rpm for 10 minutes at various times after induction with IPTG. All pelleted cells were resuspended in 25 ml of lysis buffer (6 M guanidine·HCl, 0.1 M Na_2HPO_4, pH 8) and placed in -80 °C freezer overnight. The supernatants from the cell lysates were subjected to nickel(II)-NTA metal affinity chromatography. The collected protein fractions were frozen at -80°C and lyophilized to give total dry weights for all proteins of 75-100 mg per liter of growth medium.

4.4 Results and Discussion

DNA encoding the leucine zipper proteins was first introduced into bacterial hosts auxotrophic for leucine. Expression of A1 and B1 was carried out by shifting the growth medium to minimal medium containing 19 or 20 amino acids prior to induction with isopropyl-β-thiogalactoside. Figures 4.2 and 4.3 show the results of the protein expressions in media lacking Leu and Tfl, in the presence of Leu but not Tfl, and lacking Leu but enriched with Tfl. Prior to induction there is no protein accumulation but one to

three hours after induction, cells in medium containing either Leu or Tfl showed protein expression. The calculated masses for A1 and B1 are 8305 and 8299 respectively. Relatively broad bands are found on the polyacrylamide gels for both proteins; A1 migrates between molecular weights of 6200 and 14000 of the standard protein marker and B1 migrates at approximately 6200.

Substitution of Tfl for leucine in A1 and B1 was confirmed by electrospray mass spectrometry (39, 40) and the results shown in Figures 4.4-4.13. Increases in the population of Tfl substitution was monitored as a function of time after protein induction. Proteins were isolated and purified with Ni^{2+} metal affinity chromatography at 1, 2, 3, and 4 hours after induction. Figures 4.4 and 4.5 are the electrospray results for A1 and B1, respectively, grown in Leu (40 μg/ml) supplemented medium for 4 hrs, and Figures 4.6-4.13 are the results for A1 and B1 grown in Tfl (40 μg/ml) medium over 1 to 4 hrs.

The mass spectra shown in Figures 4.4 and 4.5 contain $[M+H]^+$ peaks at 8306.25 and 8300.45 that correspond to calculated masses of 8305 and 8299 for A1 and B1 proteins respectively. In addition, a peak appears at a $[M+H]^+$ that is +14 mass units higher than the calculated average mass for either protein. This corresponds to a single methylation reaction and it is believed that this reaction is a post-translational event where the N-terminal methionine is being methylated (41-43). N-monomethyl-methionine has been found in several proteins and the levels of methylation reported in literature are in the range of 30-50% (44). Specifically, a protein sequence that can trigger methylation in proteins expressed in *E. coli* is an N-terminal methionine followed by a basic residue (Lys or Arg) in the second position. Both A1 and B1 proteins have

MetArg as the first two amino acids that are translated and therefore seem likely to be methylated based on prior literature. A mass peak at 8343 found in both A1 and B1 protein spectra corresponds to a chain that has a formylated methionine at the N terminus and the mass peak at 8278 in Figure 4.4 is unassigned.

In Figures 4.6-4.13, a population of peaks, approximately 54 mass units apart, appear in the spectra of proteins expressed in Tfl enriched medium. The substitution of one Tfl for one Leu corresponds to a mass increase of 54. These peaks confirm that eight Leu positions in the protein chain are being substituted with Tfl over time. The chains free of Tfl are believed to result from the fact that the cells first use trace amounts of leucine left in the growth medium for protein translation.

Reversed phase high performance liquid chromatography (RPLC) was used to separate the Tfl-substituted chains from the unsubstituted chains (Figures 4.14 and 4.15) and was used to monitor the relative amounts of substitution during protein syntheses when the concentration of Tfl was varied from 4 μg/ml to 32 μg/ml (Table 4.1). On reversed phase columns, C8 groups are known to disrupt interchain association of coiled-coil proteins without denaturing the α-helices (45). The exposed non-polar face of the helix readily partitions into the hydrophobic stationary phase at a retention time dependent on hydrophobicity at that face. It has been observed that the substitution of valine and isoleucine for leucine at the *d* position results in retention times that are shorter than those observed for the leucine containing chains (46).

Substitution of Tfl for Leu at various *d* positions along the A1 and B1 chains results in broad peaks that elute at longer retention times than the relatively narrow peaks of the unsubstituted chains. Figures 4.14 and 4.15 show elution profiles for A1

(HPLC purified), A1 (Ni^{2+} purified), Tfl A1 (Ni^{2+}) and B1 (HPLC purified), B1 (Ni^{2+} purified), and Tfl B1 (Ni^{2+}) respectively. The broad peaks eluted at retention times of 35.22-36.50 min for Tfl A1 and 31.48-32.79 min for Tfl B1 are assigned as the Tfl populations. It was this population that was isolated with RPLC for the amino acid analyses and thermal denaturation experiments described below.

In order to assess the dependence of protein synthesis on the concentration of Tfl, varying levels (4 to 32 μg/ml) of Tfl were added to the cell growth medium. Table 4.1 describes the extent of Tfl incorporation with varying Tfl concentrations. An exponential skim was used to resolve the narrow peak from the shoulder of the broad peak. The relative area of the narrow peak to the total area was used as a qualitative measure of the extent of Tfl incorporation. It is shown that by increasing the total Tfl concentration from 4 μg/ml to 32 μg/ml, the amount of Tfl incorporated is increased as evidenced by the increase in the area of the broad peak.

Table 4.1 Extent of Tfl incorporation with Tfl concentration in growth medium.

Sample	Concentration of Tfl (μg/ml)[a]	Fraction of Leu Chains (%)[b]
Tfl A1	4	100
Tfl A1	8	14
Tfl A1	16	12
Tfl A1	32	7
Tfl B1	4	100
Tfl B1	8	N/A
Tfl B1	16	12
Tfl B1	32	10

[a] The amount of Tfl is the final concentration in the cell growth medium.
[b] Relative area of narrow peak, corresponding to fully leucinated chains, to total area.

Amino acid analyses were performed on Tfl A1 and Tfl B1 populations that were isolated and purified by reversed phase HPLC. Comparison of the percent leucine content suggested 66% and 38% levels of Tfl substitution for A1 and B1 respectively (Tables 4.2 and 4.3). Initial measurements of thermal stabilities of α-helical Leu and Tfl containing A1 and B1 were made by circular dichroism spectroscopy (Table 4.4 and Figure 4.16). Thermal stabilities of the Tfl-substituted helices recorded at 222 nm were higher than those of the corresponding Leu-substituted helices by 8°C and 13°C respectively. These data suggest that the increase in hydrophobicity caused by the substitution of trifluoromethyl for the methyl group at the 5-position of leucine causes an increased stability of the α-helix for both A1 and B1.

Table 4.2 Amino acid compositional analysis of Recognins A1 and Tfl A1.

amino acid	mol % (theor)	mol % (obs) A1	mol % (obs) Tfl A1
aspartic acid + asparagine	12.2	12.5	12.4
glutamic acid + glutamine	17.6	18.4	18.9
serine	8.1	8.3	8.7
glycine	6.8	7.1	7.7
histidine	8.1	8.1	7.3
arginine	8.1	7.9	7.6
threonine	1.4	1.5	1.5
alanine	9.5	9.4	10.2
proline	1.4	1.7	1.5
tyrosine	0.0	0.1	0.2
valine	4.1	3.7	4.1
methionine	2.7	2.3	2.7
cysteine	0.0	0.1	0.1
isoleucine	4.1	3.5	3.7
leucine	10.8	9.7	3.1
phenylalanine	0.0	0.1	0.2
lysine	5.4	5.6	10.2

Table 4.3 Amino acid compositional analysis of Recognins B1 and Tfl B1.

amino acid	mol % (theor)	mol % (obs) B1	mol % (obs) Tfl B1
aspartic acid + asparagine	12.2	12.0	12.8
glutamic acid + glutamine	9.5	9.9	10.5
serine	8.1	8.3	8.9
glycine	6.7	7.2	7.9
histidine	8.1	7.7	8.4
arginine	8.1	8.0	8.2
threonine	1.4	1.4	1.6
alanine	9.5	9.7	9.5
proline	1.4	1.5	1.4
tyrosine	0.0	0.1	0.1
valine	4.1	3.9	3.3
methionine	2.7	2.2	2.2
cysteine	0.0	0.1	0.3
isoleucine	4.1	3.8	3.6
leucine	10.8	10.9	7.1
phenylalanine	0.0	0.0	0.1
lysine	13.5	13.2	14.0

Table 4.4 Thermal melting transitions of A1, Tfl A1, B1, and Tfl B1 monitored at 222 nm.

Dimer Complex	T_m in PBS (°C)[a]
A1-A1	58
Tfl A1-Tfl A1	66
B1-B1	75
Tfl B1-Tfl B1	88

[a] T_m = temperature at which 50% of the helix is unfolded in 50 mM sodium phosphate, 100 mM NaCl, pH 7.4.

These results raise an important question about the relative stabilities of the A1 and B1 proteins. Why does 66% replacement of leucine with trifluoroleucine in A1 only

result in an 8 °C increase in T_m whereas a 38% replacement in B1 results in a 13 °C increase in T_m? One possible explanation is that an increased level of Tfl will destablize helical-helical interactions due to the increased size of fluorine over that of hydrogen and to the difference in chemical character between fluorine and hydrogen. The introduction of Tfl at the *d* position in the heptad repeat could disrupt the overall packing of coiled-coils. At lower levels of Tfl substitution, the effect of increased size and electronegativity on the packing density might be less significant than at higher levels. It is speculated that at near complete Tfl substitution at the *d* position the helices would be very stable due to favorable fluorine-fluorine interaction. The interaction between fluorine and hydrogen is not favorable and thus at intermediate levels of Tfl substitution, the helices would be destabilized. Eventually, we would like to optimize this system so that a relationship between Tfl substitution and thermal stability can be developed.

4.5 Conclusions

An *in vivo* synthetic strategy is established for the incorporation of 5',5',5'-trifluoroleucine at the Leu positions of recombinant coiled-coil proteins, A1 and B1. The results demonstrate that coiled-coils become more thermally stable with relatively low levels of Tfl substitution. The use of leucine auxotrophs to prepare unique coiled-coil proteins containing Tfl groups can be applied to a variety of proteins where Leu participates in folding, stability, activity, or function. With optimization of this strategy, the enhancement of certain physical properties in natural polymer systems can be achieved for biological or materials applications.

N' MRGSHHHHHHGSMA - *Helix* - IGDLNNTSGIRRPAAKLN C'

Helix = SGDLENEVAQLEREVRSLEDEAAELEQKVSRLKNEIEDLKAE = A1

Helix = SGDLKNKVAQLKRKVRSLKDKAAELKQEVSRLENEIEDLKAK = B1

Figure 4.1. Amino acid sequences for recombinant leucine zipper proteins, A1 and B1. Eight leucine positions in each chain (bolded) can be substituted with Tfl. Charge patterns in the *e* and *g* positions of the *Helix* heptad repeat, $(a\,b\,c\,d\,e\,f\,g)_6$, are underlined. Abbreviations for the amino acids are : A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and W, Trp.

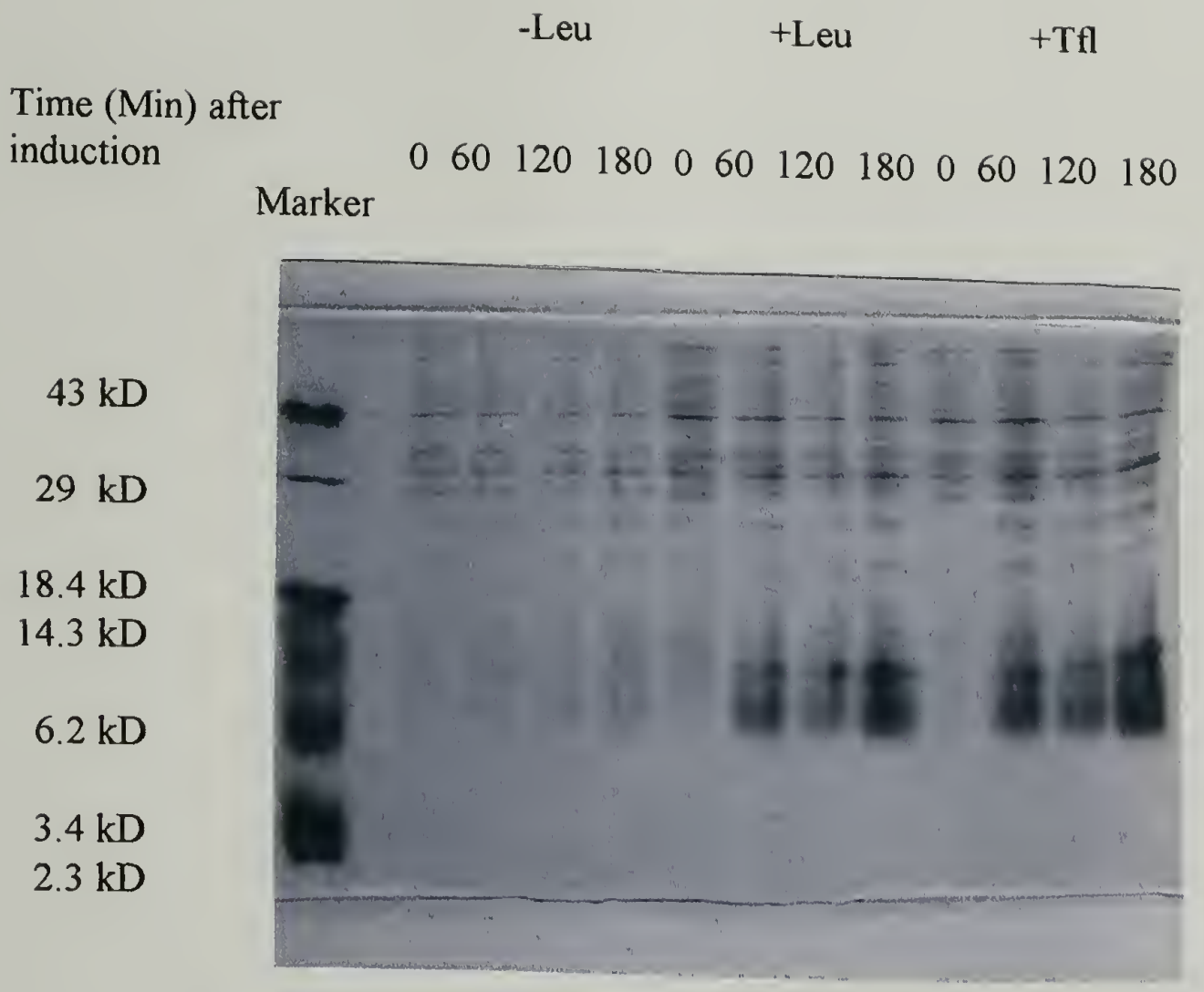


Figure 4.2 *In vivo* synthesis of Recognin A1 after culture was shifted to minimal medium supplemented with amino acids. The lanes are indicated as follows: Lane 1, low molecular weight marker; lanes 2-5, medium lacking both leucine and Tfl; lanes 6-9, medium containing leucine (40 μg/ml) but lacking Tfl; and lanes 10-13, medium lacking leucine but containing Tfl (40 μg/ml). Protein accumulation was monitored for three hours after induction with IPTG (1 mM) and was visualized on a 15% SDS-polyacrylamide gel with Coomassie Brillant Blue R-250.

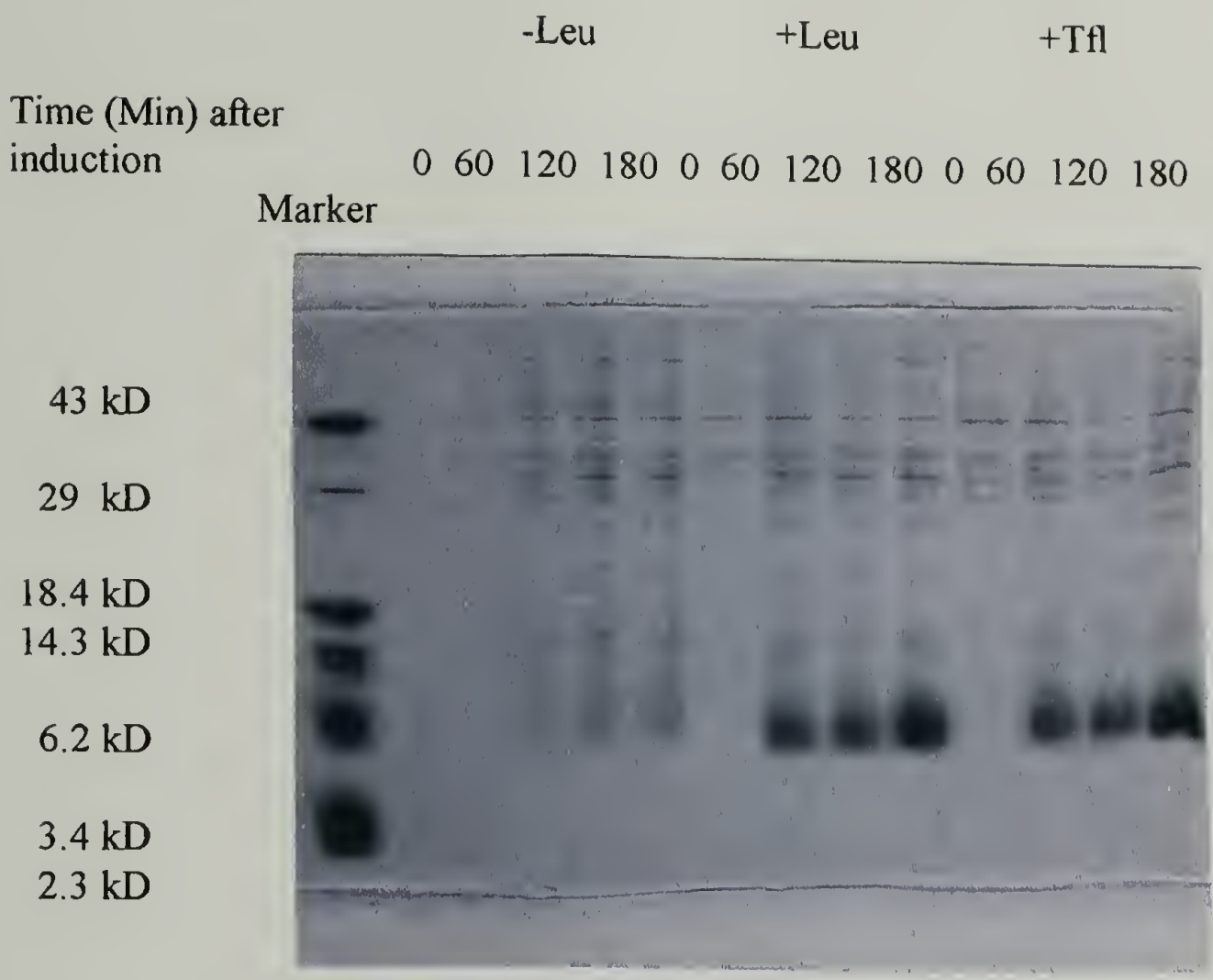


Figure 4.3 *In vivo* synthesis of Recognin B1 after culture was shifted to minimal medium supplemented with amino acids. The lanes are indicated as follows: Lane 1, low molecular weight marker; lanes 2-5, medium lacking both leucine and Tfl; lanes 6-9, medium containing leucine (40 μg/ml) but lacking Tfl; and lanes 10-13, medium lacking leucine but containing Tfl (40 μg/ml). Protein accumulation was monitored for three hours after induction with IPTG (1 mM) and was visualized on a 15% SDS-polyacrylamide gel with Coomassie Brillant Blue R-250.

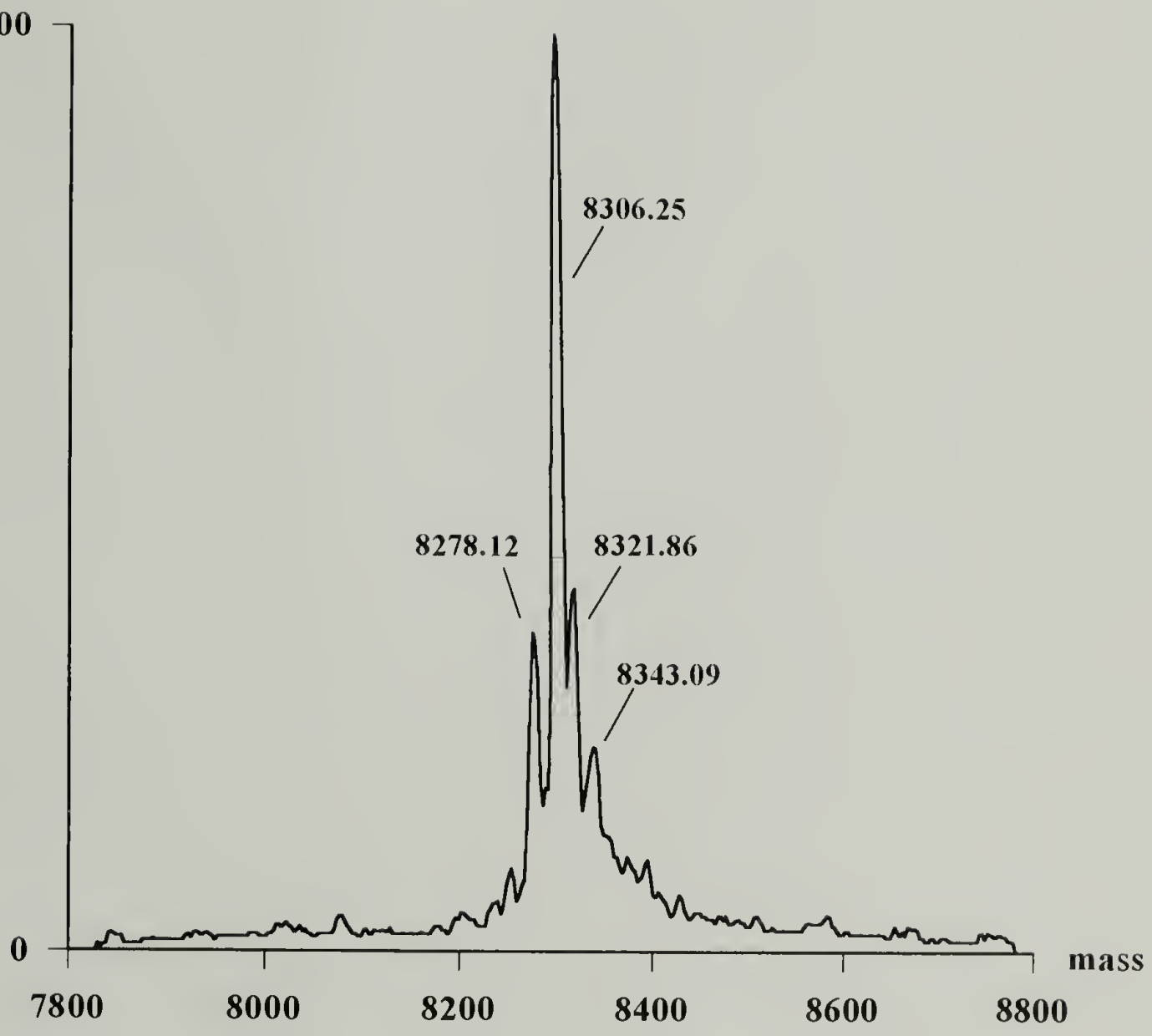


Figure 4.4 Electrospray mass spectrometry results for Recognin A1 (12 pmol/μl) in 50:50 acetonitrile:water with 0.2% formic acid at an infusion rate of 5 μl/min (35 cone voltage, positive ion scan).

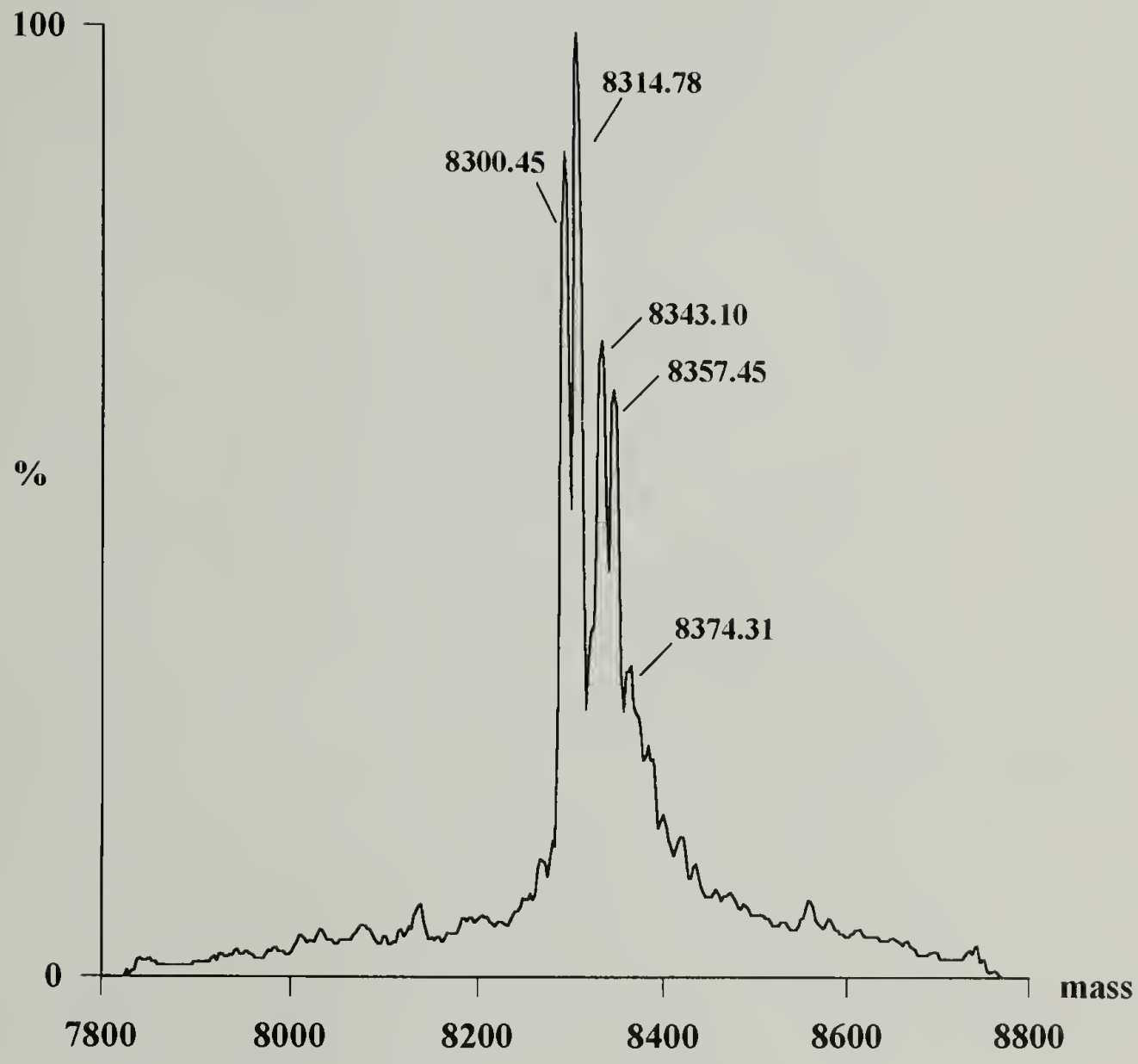


Figure 4.5 Electrospray mass spectrometry results for Recognin B1 (12 pmol/μl) in 50:50 acetonitrile:water with 0.2% formic acid at an infusion rate of 5 μl/min (50 cone voltage, positive ion scan).

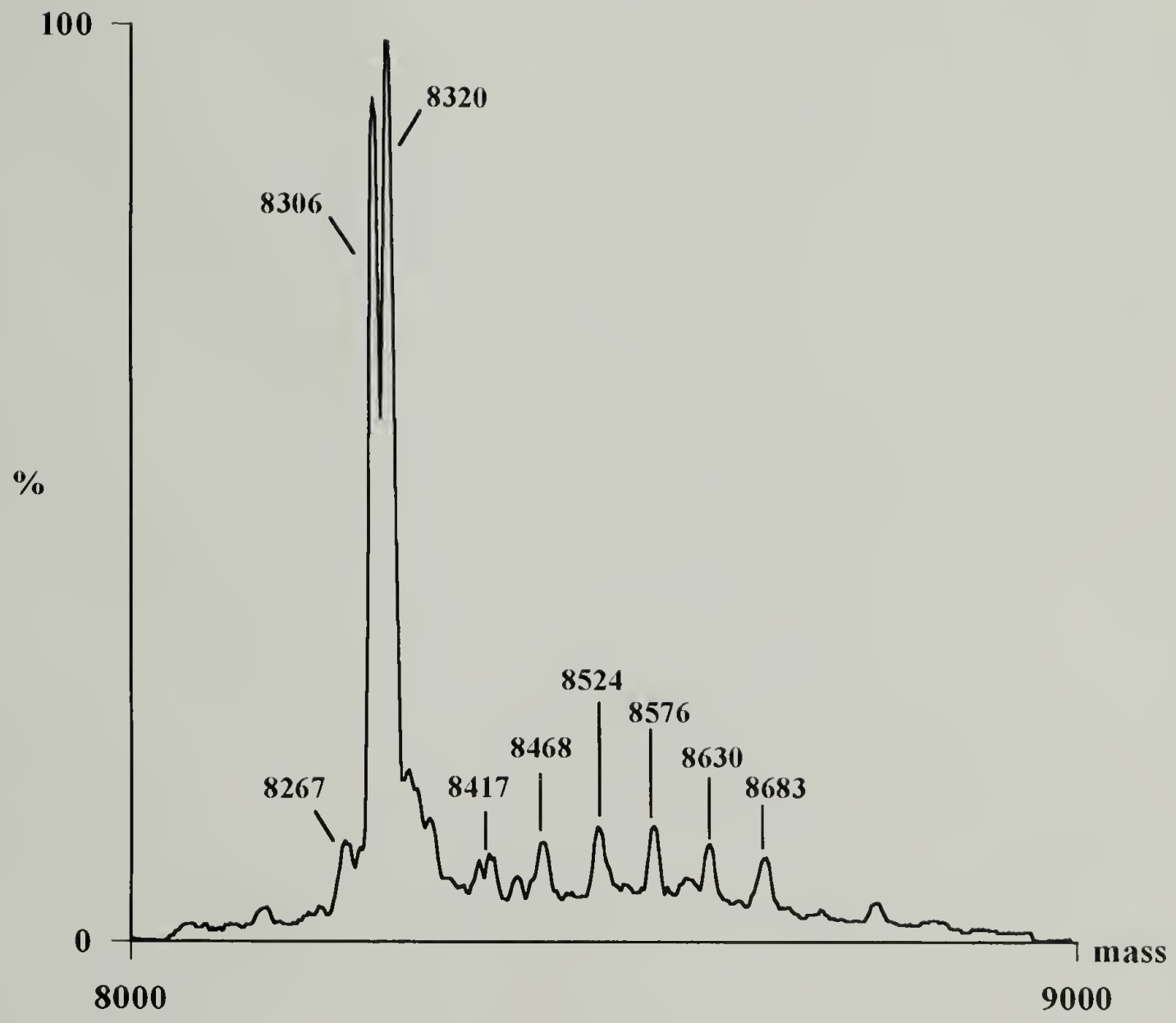


Figure 4.6 Electrospray mass spectrometry results for Tfl A1 1 hour after protein initiation. 12 pmol/µl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 µl/min (35 cone voltage, positive ion mode).

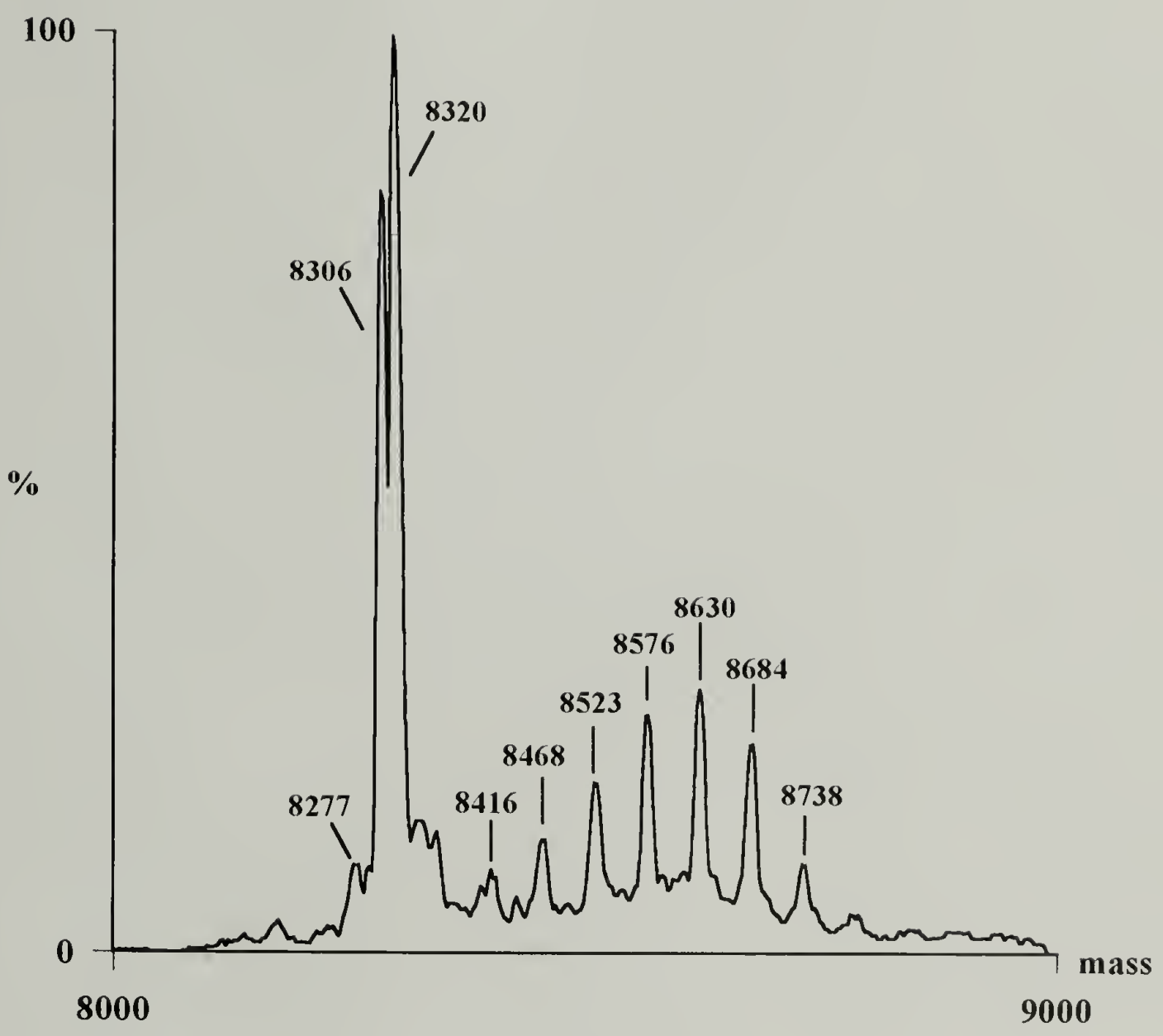


Figure 4.7 Electrospray mass spectrometry results for Tfl A1 2 hours after protein initiation. 12 pmol/μl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 μl/min (35 cone voltage, positive ion mode).

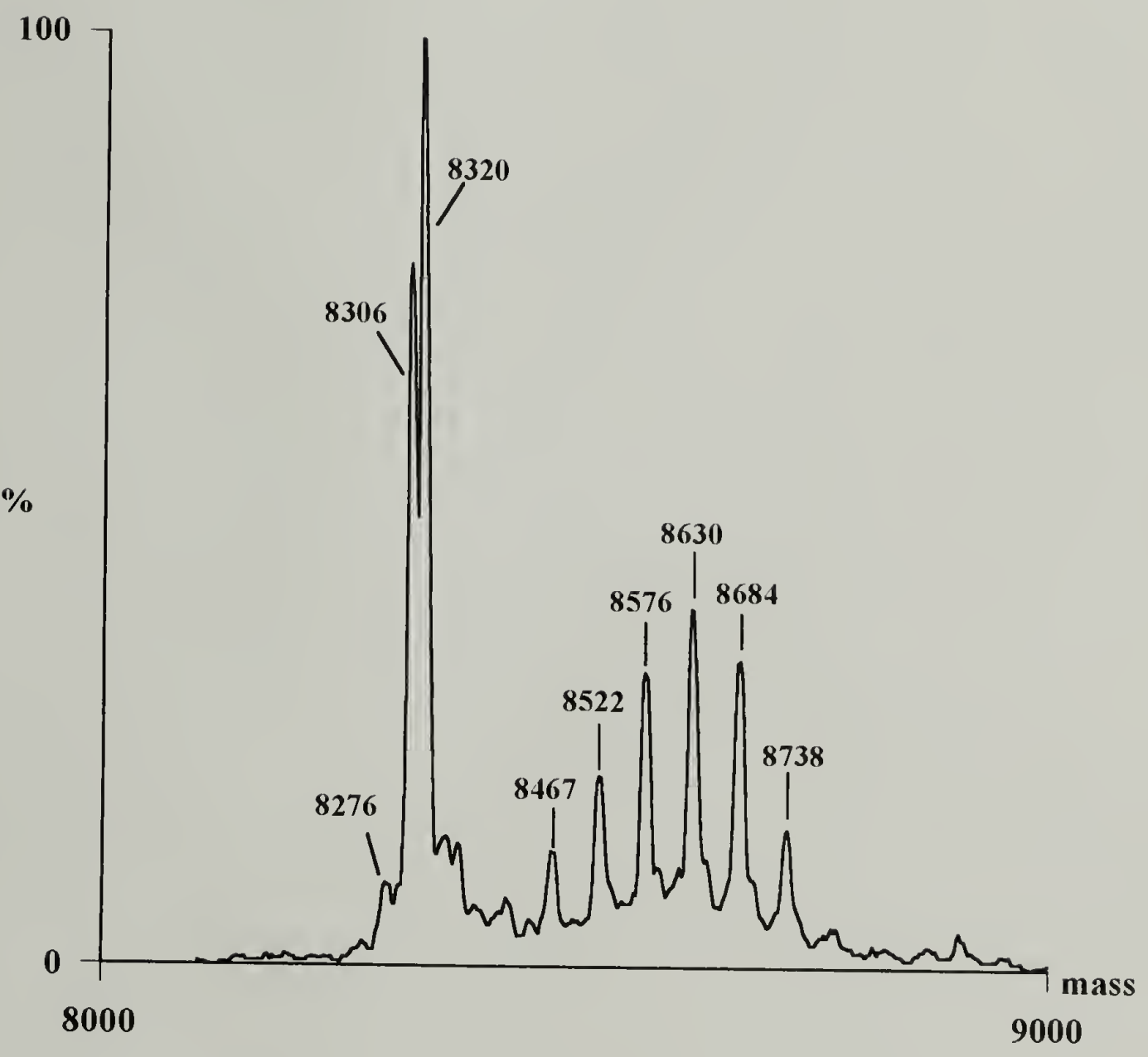


Figure 4.8 Electrospray mass spectrometry results for Tfl A1 3 hours after protein initiation. 12 pmol/μl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 μl/min (35 cone voltage, positive ion mode).

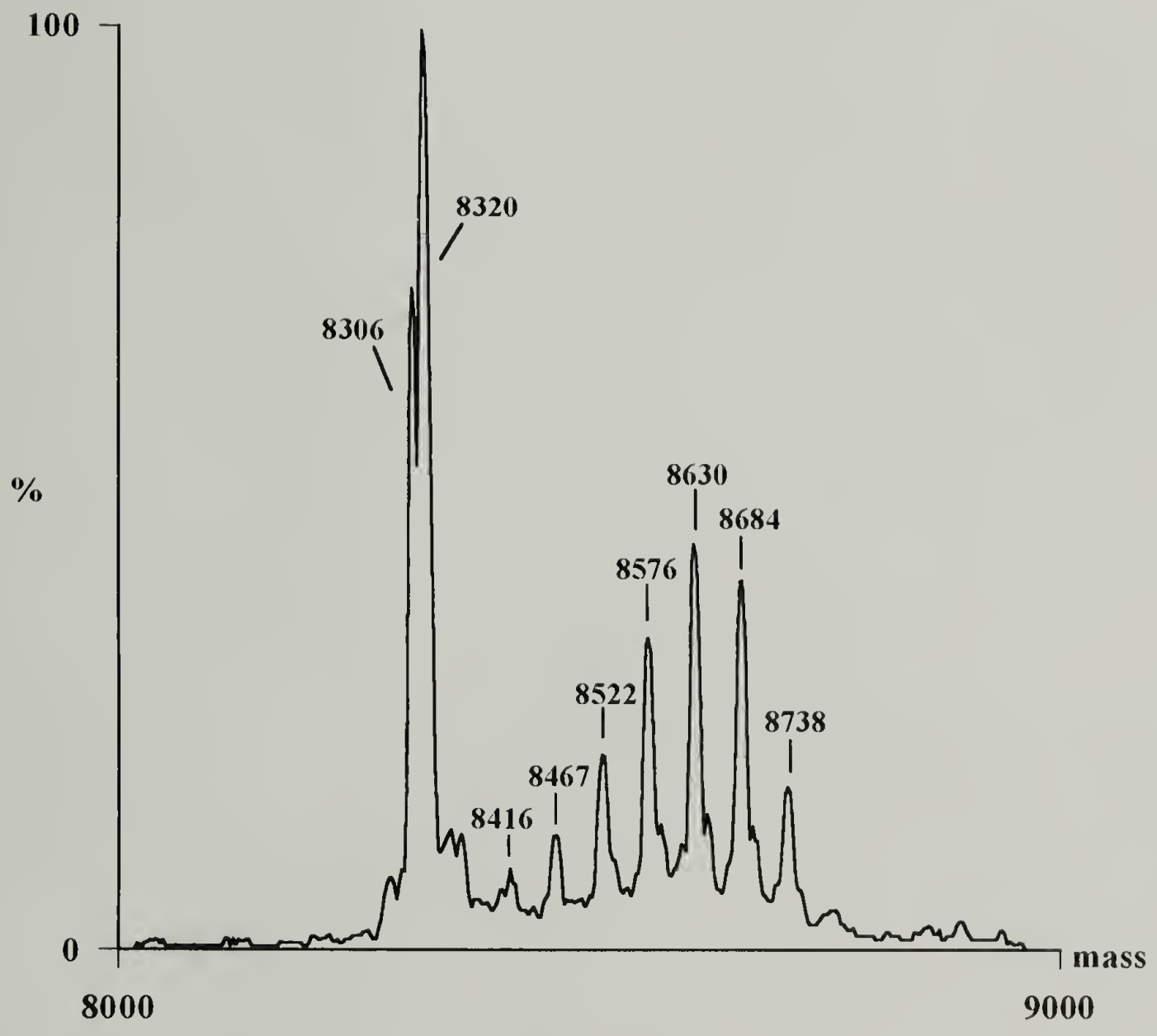


Figure 4.9 Electrospray mass spectrometry results for Tfl A1 4 hours after protein initiation. 12 pmol/μl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 μl/min (35 cone voltage, positive ion mode).

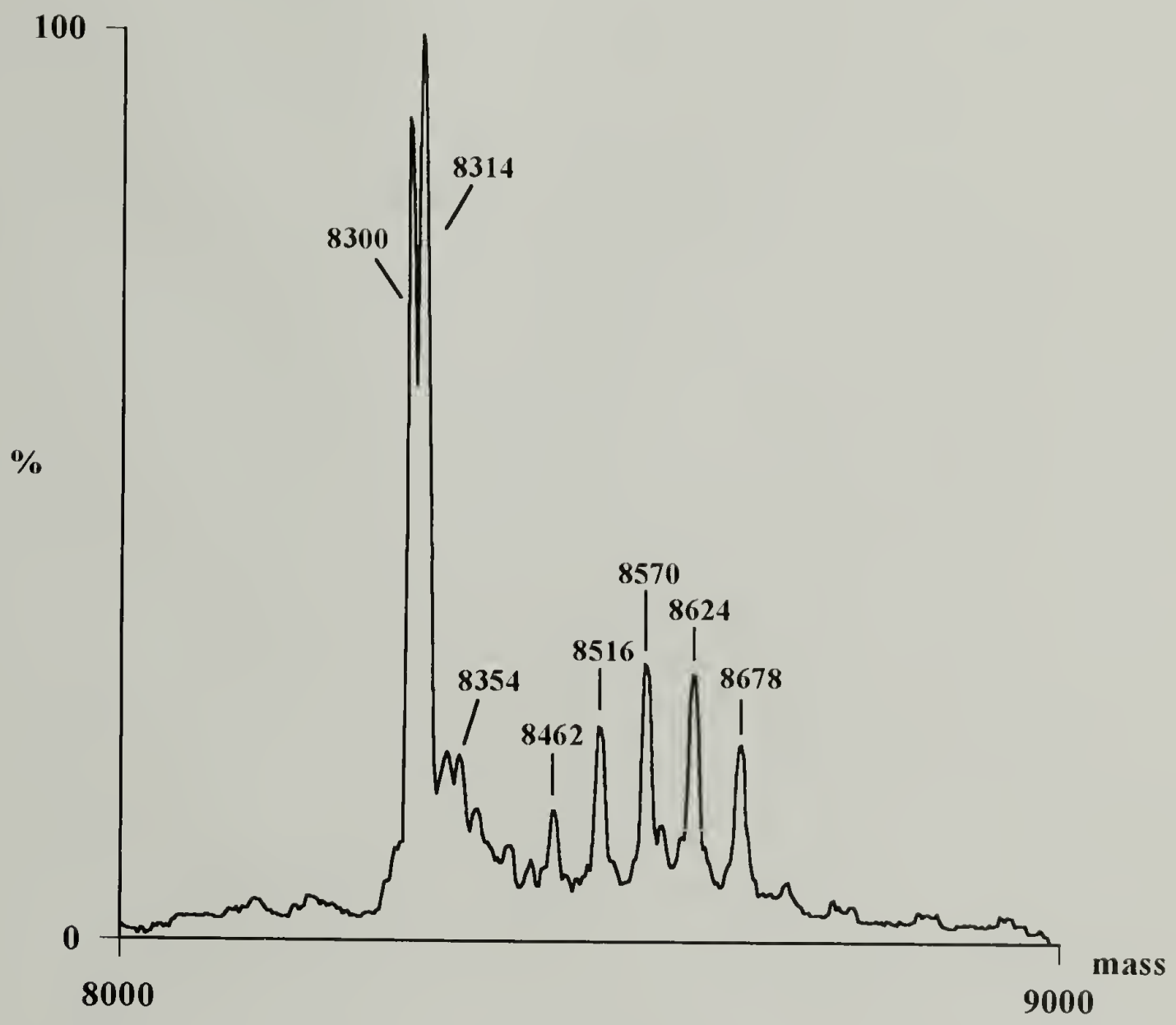


Figure 4.10 Electrospray mass spectrometry results for Tfl B1 1 hour after protein initiation. 12 pmol/µl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 µl/min (50 cone voltage, positive ion mode).

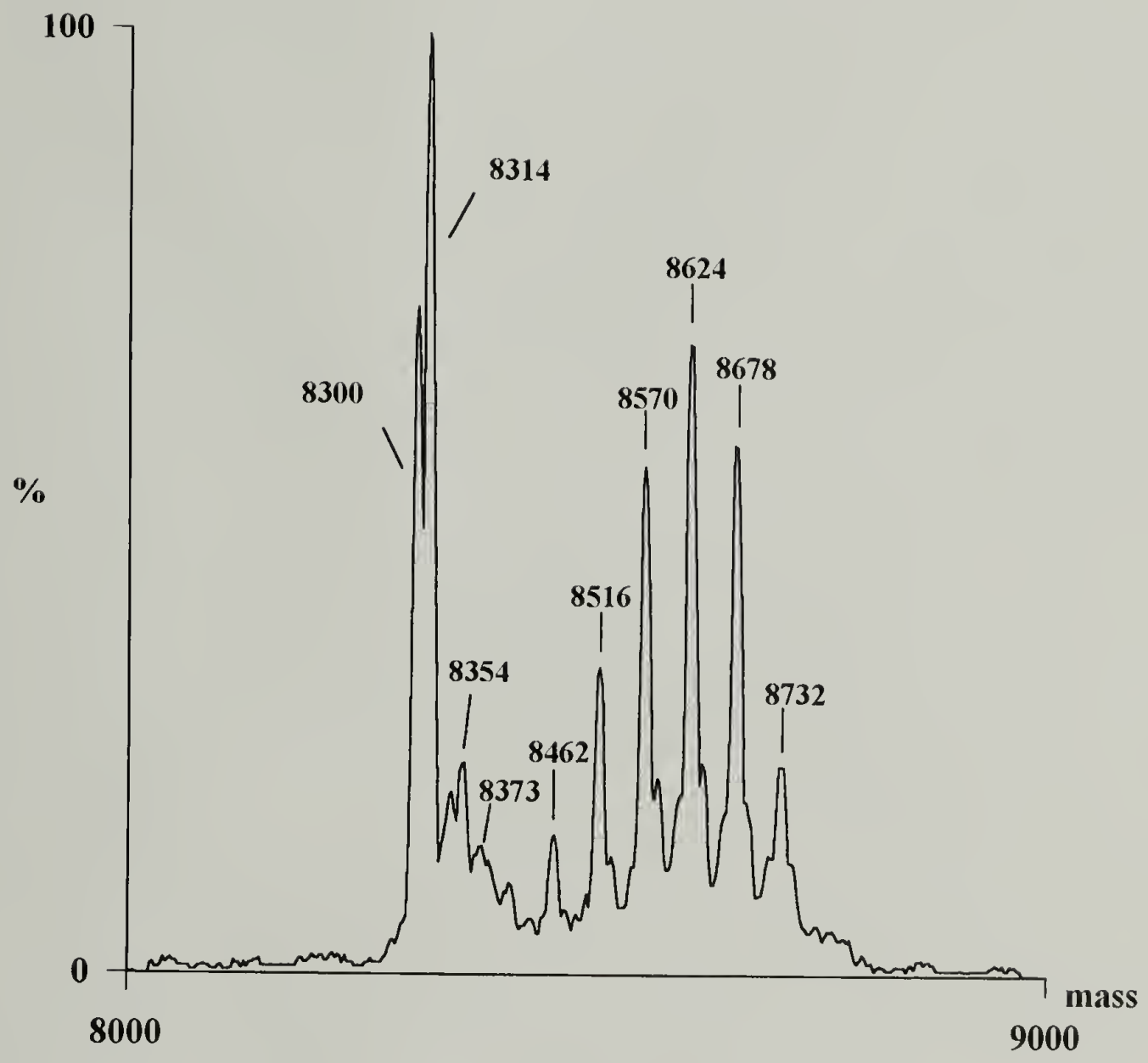


Figure 4.11 Electrospray mass spectrometry results for Tfl B1 2 hours after protein initiation. 12 pmol/µl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 µl/min (50 cone voltage, positive ion mode).

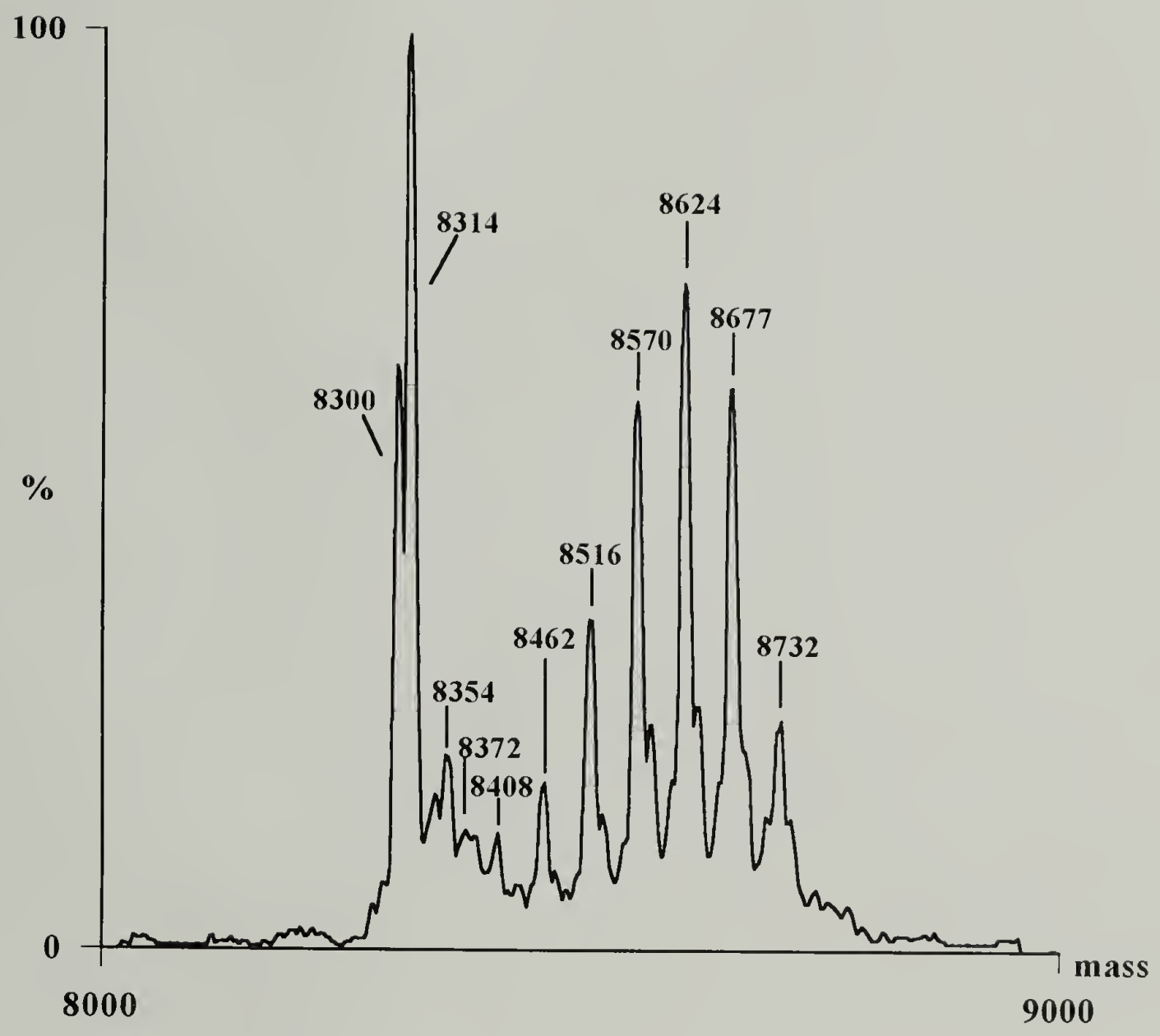


Figure 4.12 Electrospray mass spectrometry results for Tfl B1 3 hours after protein initiation. 12 pmol/µl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 µl/min (50 cone voltage, positive ion mode).

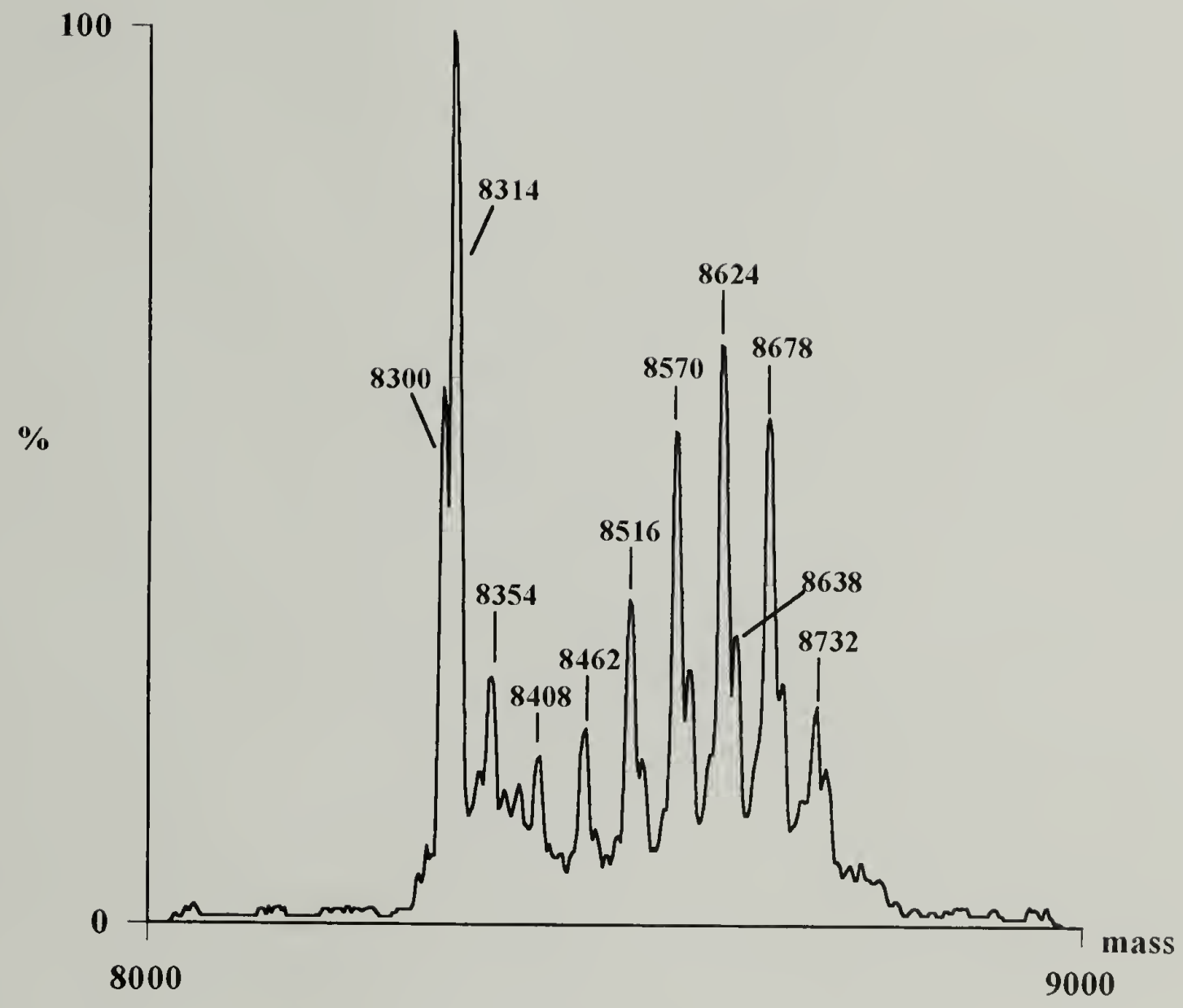


Figure 4.13 Electrospray mass spectrometry results for Tfl B1 4 hours after protein initiation. 12 pmol/µl in 50:50 acetonitrile:water with 0.2% formic acid was injected at a rate of 5 µl/min (50 cone voltage, positive ion mode).

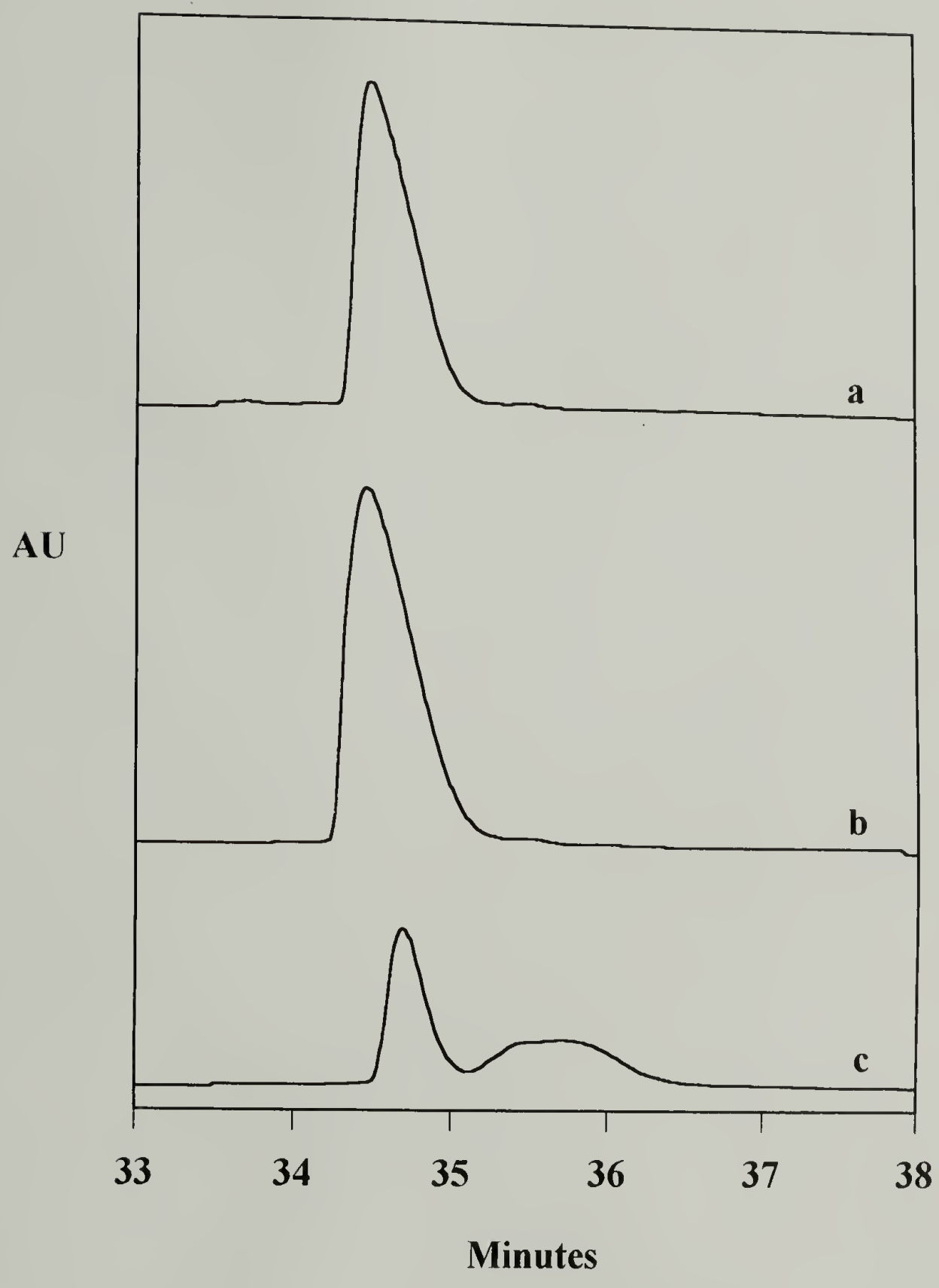


Figure 4.14 Reversed phase high performance liquid chromatograms of (**a**) HPLC purified Recognin A1 (**b**) Ni^{2+} metal affinity purified Recognin A1 and (**c**) Ni^{2+} metal affinity purified Tfl A1 detected at 210 nm on a C18-reverse phase column with H_2O (0.1% TFA): CH_3CN (0.1% TFA) gradient of 100:0 to 50:50 over 60 minutes.

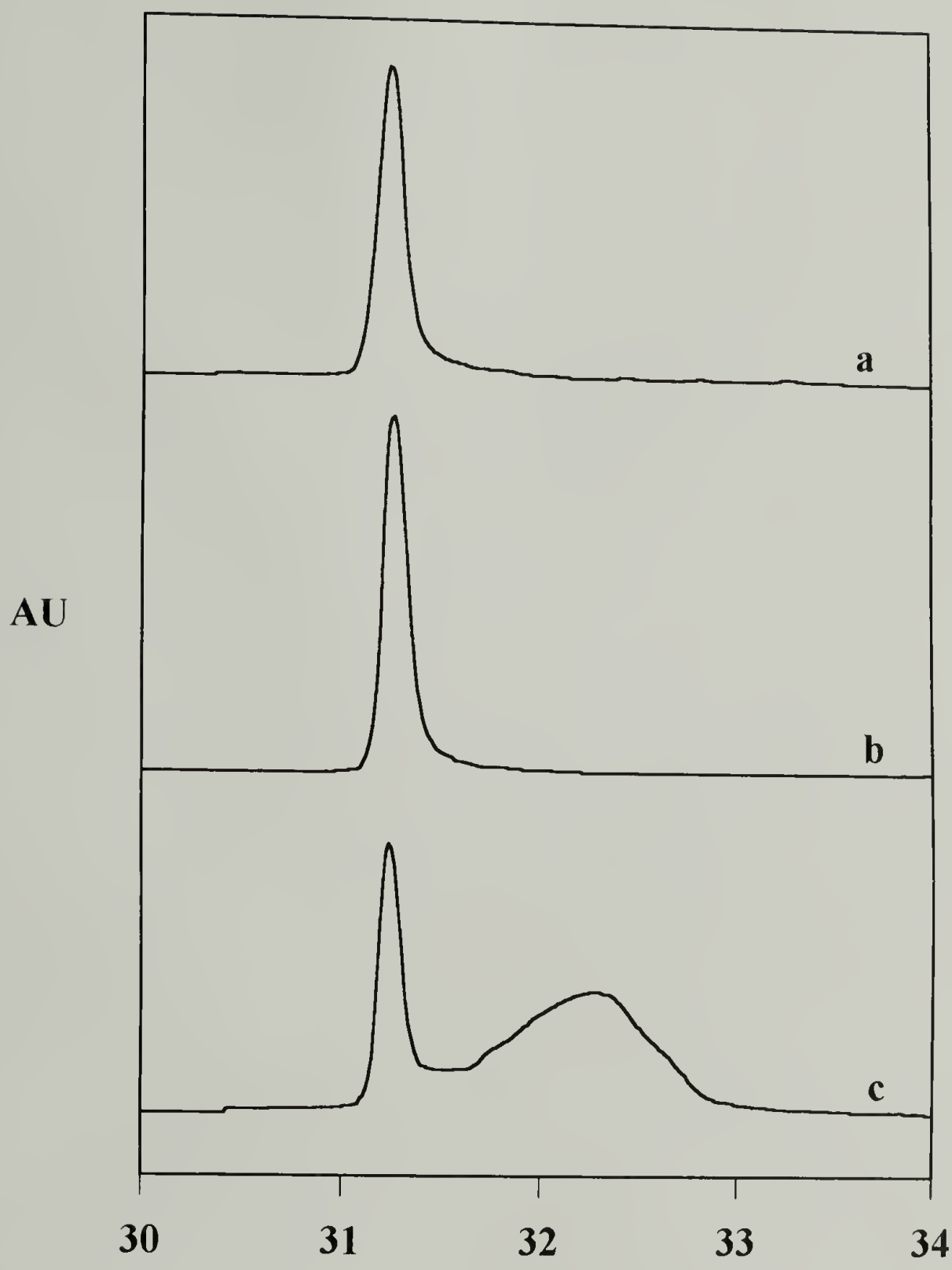
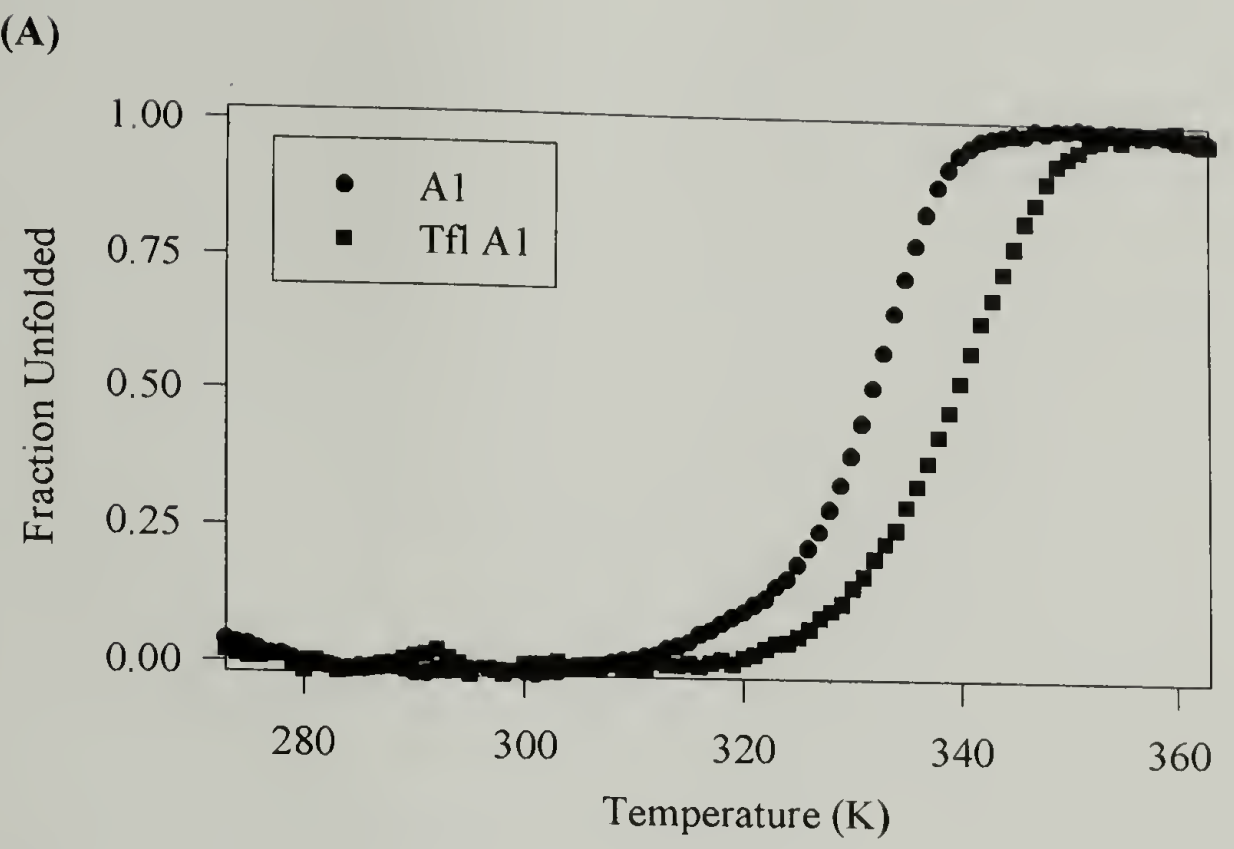


Figure 4.15 Reversed phase high performance liquid chromatograms of (**a**) HPLC purified Recognin B1 (**b**) Ni^{2+} metal affinity purified Recognin B1 and (**c**) Ni^{2+} metal affinity purified Tfl B1 detected at 210 nm on a C18-reverse phase column with H_2O (0.1% TFA): CH_3CN (0.1% TFA) gradient of 100:0 to 50:50 over 60 minutes.



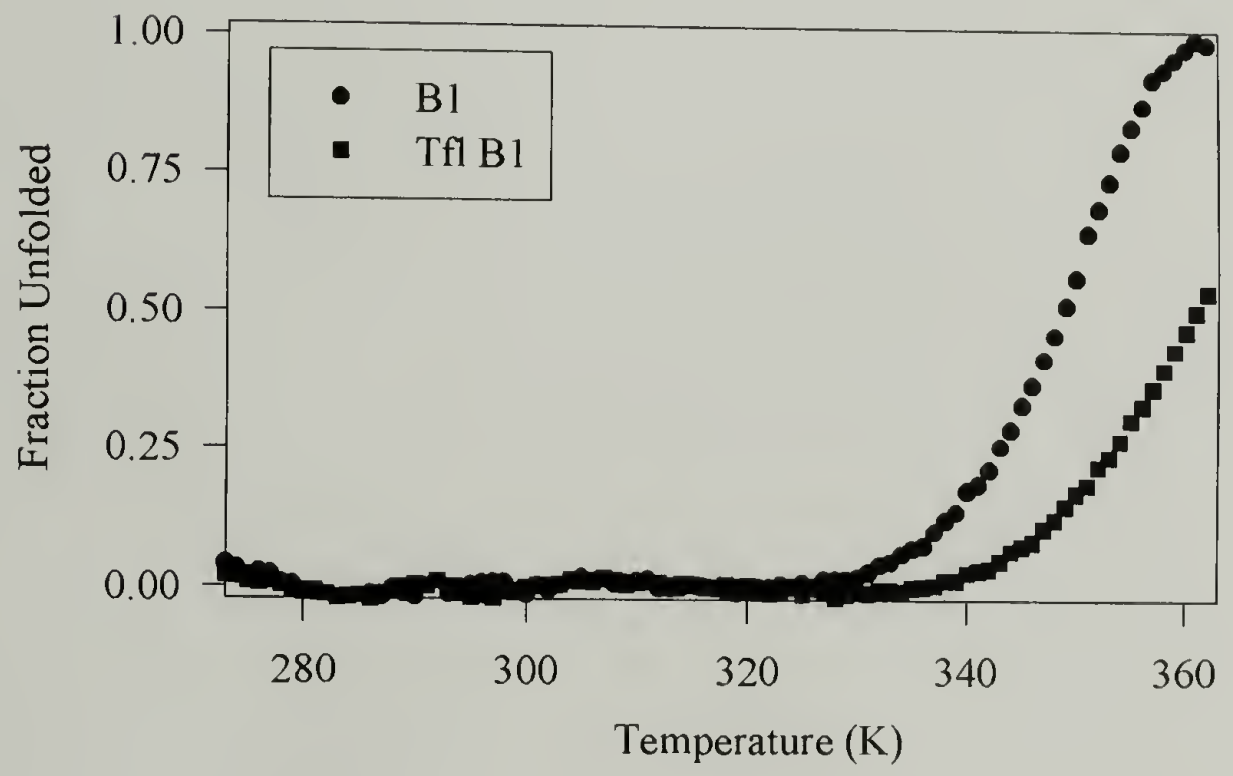


Figure 4.16 Thermal melting profiles of (**A**) A1 and Tfl A1 and (**B**) B1 and Tfl B1 monitored with circular dichroism at 222 nm. Spectra were acquired in 50 mM Na_2HPO_4, 100 mM NaCl, pH 7.4 at protein concentrations of 3 μM. T_m was determined as the temperature where 50% of the helix is unfolded.

4.6 References

(1) Attias, J.; Schlesinger, M. J.; Schlesinger, S. *J. Biol. Chem.* **1969**, *244*, 3810.

(2) Gueguen, P.; Padron, M.; Perbal, B.; Herve, G. *Biochim. Biophys. Acta* **1980**, *615*, 59.

(3) Pines, M.; Rosenthal, G. A.; Applebaum, S. W. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5480.

(4) Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205.

(5) Heckler, T. G.; Chang, L-H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1984**, *23*, 1468.

(6) Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P-I.; Hecht, S. M. *Biochemistry* **1988**, *27*, 7254.

(7) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182.

(8) Bain, J. D.; Diala, E. S.; Glabe, C. G.; Wacker, D. A.; Lyttle, M. H.; Dix, T. A.; Chamberlin, A. R. *Biochemistry* **1991**, *30*, 5411.

(9) Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197.

(10) Hecht, S. M. *Acc. Chem. Res.* **1992**, *25*, 545.

(11) Mendel, D.; Ellman, J.; Schultz, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 4359.

(12) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537.

(13) Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. *J. Am. Chem. Soc.* **1989**, *111*, 8013.

(14) Bodanszky, M.; Bodanszky, A. *The Practice of Peptide Synthesis*; Springer Verlag: Berlin, 1984.

(15) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149.

(16) Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779.

(17) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471.

(18) Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 536.

(19) Kothakota, S.; Yoshikawa, E.; Murphy, O. J.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Polym. Sci. Part A Polym Chem.* **1995**, *33*, 1267.

(20) Kothakota, S.; Dougherty, M. J.; Fournier, M. J.; Mason, T. L. *Macromol. Symp.* **1995**, *98*, 573.

(21) Deming, T. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1996**, *29*, 1442.

(22) Rennert, O. M.; Anker, H. S. *Biochemistry* **1963**, *2*, 471.

(23) Fenster, E. D.; Anker, H. S. *Biochemistry* **1969**, *8*, 269.

(24) Fenster, E. D. *Biochim. Biophys. Acta* **1971**, *228*, 701.

(25) Walborsky, H. M.; Lang, J. H. *J. Am. Chem. Soc.* **1956**, *78*, 4314.

(26) Yale, H. L. *J. Med. Pharm. Chem.* **1959**, *1*, 121.

(27) Chambers, S. E.; Lau, E. Y.; Gerig, J. T. *J. Am. Chem. Soc.* **1994**, *116*, 3063.

(28) Hagen, D. S.; Weiner, J. H.; Sykes, B. D. *Biochemistry* **1978**, *17*, 3860.

(29) Hodges, R. S.; Saund, A. K.; Chong, P. C. S.; St.-Pierre, S. A.; Reid, R. E. *J. Biol. Chem.* **1981**, *256 No. 3*, 1214.

(30) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.

(31) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

(32) Rabenstein, M.; Shin, Y-K. *Biochemistry* **1995**, *34*, 13390.

(33) Zhu, B-Y.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1993**, *2*, 383.

(34) Su, J. Y.; Hodges, R. S.; Kay, C. M. *Biochemistry* **1994**, *33*, 15501.

(35) McGrath, K. P.; Kaplan, D. L. *Mat. Res. Soc. Symp. Proc.* **1993**, *292*, 83.

(36) Cohen, S. A.; Meys, M.; Tarvin, T. L. *The Pico-Tag Method: A Manual of Advanced Techniques for Amino Acid Analysis*, Millipore Corporation: Bedford, 1989.

(37) Heinrikson, R. L.; Meredith, S. C. *Anal. Biochem.* **1984**, *136*, 65.

(38) Bidlingmeyer, B. A.; Cohen, S. A.; Tarvin, T. L. *J. Chromatogr.* **1984**, *336*, 93.

(39) Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451.

(40) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

(41) Chen, R.; Brosius, J.; Wittmann-Liebod, B. *J. Mol. Biol.* **1977**, *111*, 173.

(42) Stock, A.; Clarke, S.; Clarke, C.; Stock, J. *FEBS Lett.* **1987**, *220*, 8.

(43) Stock, A. in *Advances in Post-Translational Modifications of Proteins and Aging*; Zappia, V.; Galletti, P.; Porta, R.; Wold, F., Eds.; Plenum Press: New York, 1988.

(44) Apostol, I.; Aitken, J.; Levine, J.; Lippincott, J.; Davidson, J. S.; Abbott-Brown, D. *Protein Sci.* **1995**, *4*, 2616.

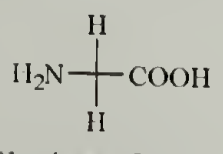
(45) Zhou, N. E.; Mant, C. T.; Hodges, R. S. *Pept. Res.* **1990**, *3*, 8.

(46) Hodges, R. S.; Zhu, B. Y.; Xhou, N. E.; Mant, C. T. *J. Chromatogr.* **1994**, *676*, 3.

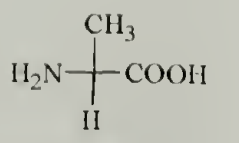
APPENDIX

LIST OF AMINO ACIDS

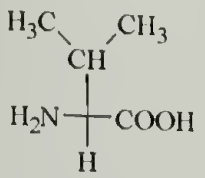
ALIPHATIC



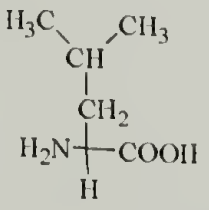
Glycine (Gly) G



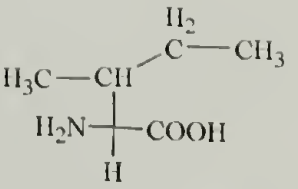
Alanine (Ala) A



Valine (Val) V

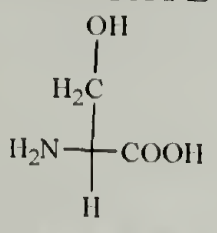


Leucine (Leu) L

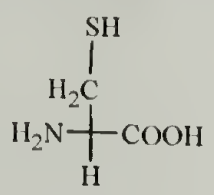


Isoleucine (Ile) I

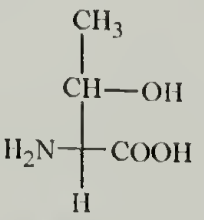
HYDROXYL OR SULFUR CONTAINING



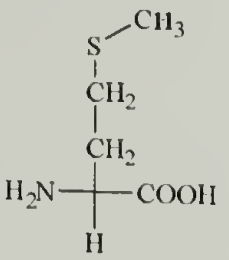
Serine (Ser) S



Cysteine (Cys) C

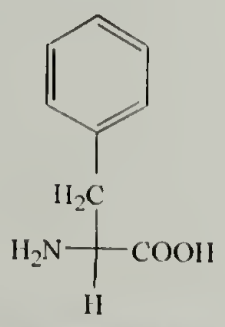


Threonine (Thr) T

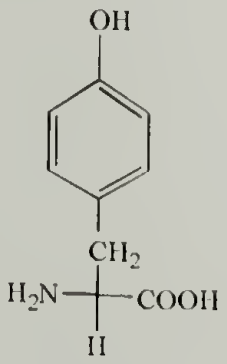


Methionine (Met) M

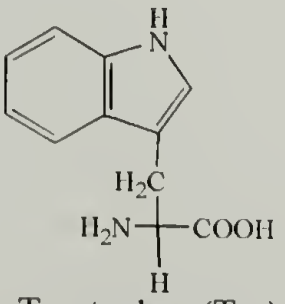
AROMATIC



Phenylalanine (Phe) F

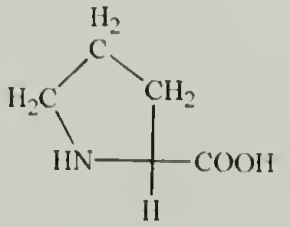


Tyrosine (Tyr) Y



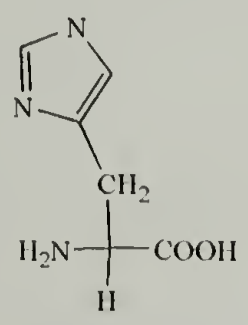
Tryptophan (Trp) W

CYCLIC

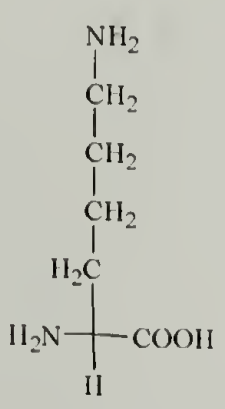


Proline (Pro) P

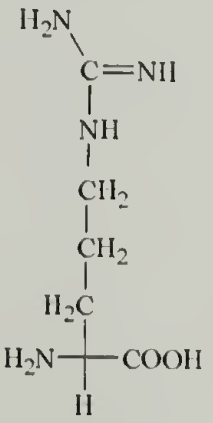
BASIC



Histidine (His) H

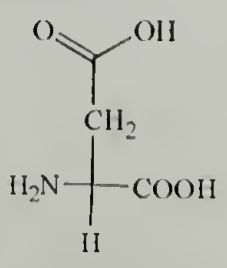


Lysine (Lys) K

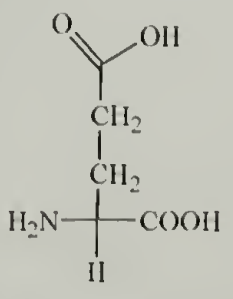


Arginine (Arg) R

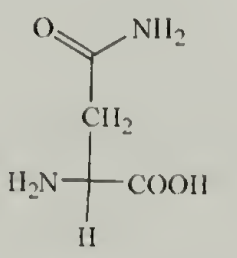
ACIDIC



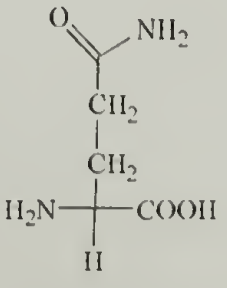
Aspartic Acid (Asp) D



Glutamic Acid (Glu) E



Asparagine (Asn) N



Glutamine (Gln) Q

BIBLIOGRAPHY

Andrade, M. A.; Chacon, P.; Merolo, J. J.; Moran, F. *Protein Eng.* **1993**, *6*, 383.

Annaka, M.; Tanaka, T. *Nature* **1992**, *355*, 430.

Annaka, M.; Tanaka, T. *Physica A* **1994**, *204*, 40.

Apostol, I.; Aitken, J.; Levine, J.; Lippincott, J.; Davidson, J. S.; Abbott-Brown, D. *Protein Sci.* **1995**, *4*, 2616.

Asquith, R. S. *Chemistry of Natural Protein Fibers*; Plenum Press: New York, 1977.

Astbury, W. T.; Reed, R.; Spark, L. C. *Biochem. J.* **1948**, *43*, 282.

Astbury, W. T.; Haggith, J. W. *Biochim. Bio-phys. Acta* **1953**, *10*, 483.

Attias, J.; Schlesinger, M. J.; Schlesinger, S. *J. Biol. Chem.* **1969**, *244*, 3810.

Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. *J. Am. Chem. Soc.* **1989**, *111*, 8013.

Bain, J. D.; Diala, E. S.; Glabe, C. G.; Wacker, D. A.; Lyttle, M. H.; Dix, T. A.; Chamberlin, A. R. *Biochemistry* **1991**, *30*, 5411.

Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537.

Bamford, C. H.; Brown, L.; Elliott, A.; Hanby, W. E.; Trotter, I. F. *Nature* **1952**, *169*, 357.

Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, *196*, 657.

Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1990**, *62*, 1836.

Beavis, R. C.; Chait, B. T.; Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 7584.

Bidlingmeyer, B. A.; Cohen, S. A.; Tarvin, T. L. *J. Chromatogr.* **1984**, *336*, 93.

Bodanszky, M.; Bodanszky, A. *The Practice of Peptide Synthesis*; Springer Verlag: Berlin, 1984.

Bohmann, D.; Bos, T. J.; Admon, A.; Nishimura, T.; Vogt, P. K.; Tjian, R. *Science* **1987**, *238*, 1386.

Candau, S.; Bastide, J.; Delsanti, M. *Advances in Polymer Science* **1982**, *44*, 27.

Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*; W. H. Freeman and Company: San Francisco, 1980.

Cappello, J.; Crissman, J.; Dorman, M.; Mikolajczak, M.; Textor, G.; Marquet, M.; Ferrari, F. *Biotechnol. Prog.* **1990**, *6*, 198.

Chambers, S. E.; Lau, E. Y.; Gerig, J. T. *J. Am. Chem. Soc.* **1994**, *116*, 3063.

Chambon, F.; Winter, H. H. *Polymer Bulletin* **1985**, *13*, 499.

Chen, Y.-H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350.

Chen, R.; Brosius, J.; Wittmann-Liebod, B. *J. Mol. Biol.* **1977**, *111*, 173.

Chen, G.; Hoffman, A. S. *Nature* **1995**, *373*, 49.

Chum, H. L., Ed. *Polymers from Biobased Materials*; Noyles Data Corporation: New Jersey, 1991.

Cochran, W.; Crick, F. H. C.; Vand, V. *Acta Cryst.* **1952**, *5*, 581.

Cohen, C.; Holmes, K. *J. Mol. Biol.* **1963**, *6*, 423.

Cohen, S. A.; Meys, M.; Tarvin, T. L. *The Pico-Tag Method: A Manual of Advanced Techniques for Amino Acid Analysis*, Millipore Corporation: Bedford, 1989.

Cohen, S.; Bañó, M. C.; Vissher, K. B.; Chow, M.; Allcock, H. R.; Langer, R. *J. Am. Chem. Soc.* **1990**, *112*, 7833.

Conway; Parry *Int. J. Biol. Macromol.* **1991**, *13*, 14.

Creel, H. S. *Ph.D. Dissertation* **1994,** University of Massachusetts, Amherst.

Creel, H.. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213.

Creighton, T. E., Ed. *Protein Folding*; W.H. Freeman and Co.: New York, 1992.

Crick, F. H. C. *Nature* **1952**, *170*, 882.

Crick, F. H. C. *Acta Cryst.* **1953**, *6*, 685.

Crick, F. H. C. *Acta Cryst.* **1953**, *6*, 689.

DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622.

Deming, T. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1996**, *29*, 1442.

Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779.

Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.

Ellman, G. L. *Archives of Biochemistry and Biophysics* **1959**, *82*, 70.

Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197.

Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

Fenster, E. D.; Anker, H. S. *Biochemistry* **1969**, *8*, 269.

Fenster, E. D. *Biochim. Biophys. Acta* **1971**, *228*, 701.

Flory, P. J.; Weaver, E. S. *J. Am. Chem. Soc.* **1960**, *82*, 4518.

Fraden, S.; Maret, G. *Phys. Rev. Lett* **1990**, *65*, 512.

Garfin, D. E. in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research*; Glasel, J. A.; Deutscher, M. P.; Eds.; Academic Press: San Diego, 1995.

Gehrke, S. *Personal Communication* .

Gentz, R.; Rauscher III, F. J.; Abate, C.; Curran, T. *Science* **1989**, *243*, 1695.

Glover, J. N. M.; Harrison, S. C. *Nature* **1995**, *373*, 257.

Godard, P.; Biebuyck, J. J.; Daumerie, M.; Naveau, H.; Mercier, J. P. *J. Polym. Sci.: Polymer Phys. Ed.* **1978**, *16*, 1817.

Graddis, T. J.; Myszka, D. G.; Chaiken, I. M. *Biochemistry* **1993**, *32*, 12664.

Green, S. M.; Meeker, A. K.; Shortle, D. *Biochemistry* **1992**, *31*, 5717.

Greenfield, N. J.; Hitchcock-DeGregori, S. E. *Biochemistry* **1995**, *34*, 16797.

Greenfield, N. J. *Anal. Biochem.* **1996**, *235*, 1.

Grohmann, K.; Himmel, M. E. *Advances in Protein-Derived Materials*; Noyes Data Corporation: Park Ridge, 1991.

Gueguen, P.; Padron, M.; Perbal, B.; Herve, G. *Biochim. Biophys. Acta* **1980**, *615*, 59.

Guenet, J.-M. *Thermoreversible Gelation of Polymers and Biopolymers*; Academic Press Inc.: San Deigo, 1992.

Hagen, D. S.; Weiner, J. H.; Sykes, B. D. *Biochemistry* **1978**, *17*, 3860.

Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *371*, 80.

Hecht, S. M. *Acc. Chem. Res.* **1992**, *25*, 545.

Heckler, T. G.; Chang, L.-H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1984**, *23*, 1468.

Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P.-I.; Hecht, S. M. *Biochemistry* **1988**, *27*, 7254.

Heinrikson, R. L.; Meredith, S. C. *Anal. Biochem.* **1984**, *136*, 65.

Hilger, C.; Drager, M.; Stadler, R. *Macromolecles* **1992**, *25*, 2498.

Hodges, R. S.; Sodek, J.; Smillie, L. B.; Jurasek, L. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *37*, 299.

Hodges, R. S.; Saund, A. K.; Chong, P. C. S.; St.-Pierre, S. A.; Reid, R. E. *J. Biol. Chem.* **1981**, *256 No. 3*, 1214.

Hodges, R. S.; Zhou, N. E.; Kay, C. M.; Semchuk, P. D. *Pept. Res.* **1990**, *3*, 123.

Hodges, R. S.; Zhu, B. Y.; Xhou, N. E.; Mant, C. T. *J. Chromatogr.* **1994**, *676*, 3.

Hodges, R. S. *Biochem. Cell Biol.* **1995**, *74*, 133.

Hoffman, A. S. *Chemtech* **1986**, 426.

Hudson, S. M.; Cuculo, J. A. *J. Polym. Sci.: Polymer Chemistry Edition* **1982**, *20*, 499.

Jirgensons, B. *Optical Activity of Proteins and Other Macromolecules*; Springer-Verlag: New York, 1973.

Johnson, W. C., Jr. *Proteins: Struc. Funct. Genet.* **1990**, *7*, 205.

Klier, J.; Scranton, A. B.; Peppas, N. A. *Macromolecules* **1990**, *23*, 4944.

König, P.; Richmond, T. J. *J. Mol. Biol.* **1993**, *233*, 139.

Kothakota, S.; Dougherty, M. J.; Fournier, M. J.; Mason, T. L. *Macromol. Symp.* **1995**, *98*, 573.

Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 536.

Kothakota, S.; Yoshikawa, e.; Murphy, O. J.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Polym. Sci. Part A Polym Chem.* **1995**, *33*, 1267.

Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.: Mason, T. L.; Tirrell, D. A. *Science* **1994**, *265*, 1427.

Krylov, D.; Mikhailenko, I.; Vinson, C. *EMBO J.* **1994**, *13*, 2849.

Laemmli, U. K. *Nature* **1970**, *227*, 680.

Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759.

Lang, P.; Burchard, W. *Macromolecules* **1991**, *24*, 814.

Lau, S. Y. M.; Taneja, A. K.; Hodges, R. S. *J. Biol. Chem.* **1984**, *259*, 13253.

Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Royal Society of Chemistry: London, 1992.

Laue, T. M. *Meth. Enzymol.* **1995**, *259*, 427.

Lehn, J.-M. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1304.

Li, Y.-T.; Hsieh, Y.-L.; Henion, J. D.; Senko, M. W.; McLafferty, F. W.; Ganem, B. *J. Am. Chem. Soc.* **1993**, *115*, 8409.

Lim, F.; Sun, A. M. *Science* **1980**, *210*, 908.

Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F.; Eisenber, D. *Science* **1993**, *259*, 1288.

Lumb, K. J.; Kim, P. S. *Science* **1995**, *268*, 436.

Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.

MacArthur, I. *Nature* **1943**, *152*, 38.

Macsuga, D. D. *Biopolymers* **1972**, *11*, 2521.

Maret, G.; Wolf, P. E. *Z. Phys. B* **1987**, *65*, 409.

Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8898.

Martin, J. E.; Wilcoxon, J. P. *Phys. Rev. Lett.* **1988**, *61*, 373.

Mason, T. G.; Gang, H.; Weitz, D. A. *J. Mol. Struct.* **1996**, *383*, 81.

McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett* **1983**, *24*, 245.

McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.

McGrath, K. P.; Kaplan, D. L. *Mat. Res. Soc. Symp. Proc.* **1993**, *292*, 83.

McGrath, K. P.; Butler, M. M.; DiGirolamo, C. M.; Petka, W. A.; Kaplan, D. L.; Laue, T. M. **1997**, *submitted for publication.*

McLachlan, A. D.; Stewart, M. *J. Mol. Biol.* **1975**, *98*, 293.

Mendel, D.; Ellman, J.; Schultz, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 4359.

Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149.

Mitchell, P. J.; Tjian, R. *Science* **1989**, *245*, 371.

Monera, O. D.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1993**, *268*, 19218.

Monera, O. D.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1994**, *3*, 1984.

Monera, O. D.; Zhou, N. E.; Lavigne, P.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1996**, *271*, 3995.

Nisizawa, M.; Endo, R. *J. Polym. Sci.: Part A: Polymer Chemistry* **1995**, *33*, 2521.

Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182.

O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538.

O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Cell* **1992**, *68*, 699.

Pace, C. N. *Meth. Enzymol.* **1986**, *131*, 266.

Pathak, C. P.; Sawhney, A. S.; Hubbell, J. A. *J. Am. Chem. Soc.* **1992**, *114*, 8311.

Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci.* **1951**, *37*, 205.

Pauling, L.; Corey, R. B. *Proc. Nat. Acad. Sci.* **1951**, *37*, 241.

Pauling, L.; Corey, R. B. *Nature* **1953**, *171*, 59.

Peppas, N. A.; Langer, R. *Science* **1994**, *263*, 1715.

Perutz, M. F. *Nature* **1951**, *167*, 1053.

Peteranderl, R.; Nelson, H. C. M. *Biochemistry* **1992**, *31*, 12272.

Petka, W. A.; McGrath, K. P.; Kaplan, D. L.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polymer Preprints* **1994**, *35*, 452.

Pine, D. J.; Weitz, D. A.; Zhu, J. X.; Herbolzheimer, E. *J. Phys. France* **1990**, *51*, 2101.

Pines, M.; Rosenthal, G. A.; Applebaum, S. W. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5480.

Provencher, S. W.; Glöckner, J. *Biochemistry* **1981**, *20*, 33.

Rabenstein, M.; Shin, Y.-K. *Biochemistry* **1995**, *34*, 13390.

Ramachandran, G. N. *Adv. Prot. Chem.* **1968**, *23*, 284.

Rennert, O. M.; Anker, H. S. *Biochemistry* **1963**, *2*, 471.

Richardson, C. D. in *The Enzymes*; Academic Press: San Diego, 1981.

Sali, D.; Bycroft, M.; Fersht, A. R. *J. Mol. Biol.* **1991**, *220*, 779.

Sanger, F. *Annu. Rev. Biochem.* **1988,** *57*, 1.

Schuermann, M.; Hunter, J. B.; Hennig, G.; Müller, R. *Nucleic Acids Research* **1991**, *19*, 739.

Sodek, J.; Hodges, R. S.; Smillie, L. B.; Jurasek, L. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 3800.

Sorger, P. K.; Nelson, H. C. M. *Cell* **1989**, *59*, 807.

Sreerama, N.; Woody, R. W. *Anal. Biochem.* **1993**, *209*, 32.

Sreerama, N.; Woody, R. W. *J. Mol. Biol.* **1994**, *242*, 497.

St. Pourcain, C. G.; Griffin, A. C. *Macromolecules* **1995**, *28*, 4116.

Stauffer, D.; Coniglio, A.; Adam, M. *Advances in Polymer Science* **1982**, *44*, 103.

Stock, A.; Clarke, S.; Clarke, C.; Stock, J. *FEBS Lett.* **1987**, *220*, 8.

Stock, A. in *Advances in Post-Translational Modifications of Proteins and Aging*; Zappia, V. ; Galletti, P.; Porta, R.; Wold, F.; Eds., Plenum Press: New York, 1988.

Stockmayer, W. H. *Macromolecules* **1991**, *24*, 6367.

Stone, D.; Sodek, J.; Johnson, P.; Smillie, L. B. *Tropomyosin: Correlation of Amino Acid Sequence and Structure,* E.N.A. Biro. North Holland Publishing Co.: Amsterdam, 1975.

Su, J. Y.; Hodges, R. S.; Kay, C. M. *Biochemistry* **1994**, *33*, 15501.

Tabor, S.; Richardson, C. C. *Proc. Nat. Acad. Sci. USA* **1987**, *84*(14), 4767.

Turner, R.; Tjian, R. *Science* **1989**, *243*, 1689.

Van Kleef, F. S. M. *Biopolymers* **1986**, *25*, 31.

Venyaminov, S. Y.; Baikalov, I. A.; Shen, Z. M.; Wu, C.-S. C.; Yang, J. T. *Anal. Biochem.* **1993**, *214*, 17.

Vieth, M.; Kolinski, A.; Brooks III, C. L.; Skolnick, J. *J. Mol. Biol.* **1995**, *251*, 448.

Walborsky, H. M.; Lang, J. H. *J. Am. Chem. Soc.* **1956**, *78*, 4314.

Walton, A. G. *Polypeptides and Protein Structure*; Elsevier: New York, 1981.

Weber, P. C.; Salemme, F. R. *Nature* **1980**, *287*, 82.

Weis, W. I.; Brunger, A. T.; Skehel, J. J.; Wiley, D. C. *J. Mol. Biol.* **1990**, *212*, 737.

Weiss, B.; Jacquemin-Sablon, A.; Live, T. R.; Fareed, G. C.; Richardson, C. C. *J. Biol. Chem.* **1968**, *243*, 4543.

Weitz, D. A.; Zhu, J. X.; Durian, D. J.; Gang, H.; Pine, D. J. *Physica Scripta* **1993**, *T49*, 610.

Weitz, D. A.; Pine, D. J. *Diffusing-wave Spectroscopy*; Clarendon Press: Oxford, 1993.

Wendt, H.; Berger, C.; Baici, A.; Thomas, R. M.; Bosshard, H. R. *Biochemistry* **1995**, *34*, 4097.

Wendt, H.; Durr, E.; Thomas, R. M.; Przyblyski, M.; Bosshard, H. R. *Protein Sci.* **1995**, *4*, 1563.

Wilson, I. A.; Skehel, J. J.; Wiley, D. C. *Nature* **1981**, *289*, 366.

Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205.

Winter, H. H. *Polymer Engineering and Science* **1987**, *27*, 1698.

Woolfson, D. N.; Alber, T. *Protein Sci.* **1995**, *4*, 1596.

Wu, D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *PMSE Preprints* **1996**, *74*, 71.

Yale, H. L. *J. Med. Pharm. Chem.* **1959**, *1*, 121.

Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451.

Yanisch-Perron, C.; Viera, J.; Messing, J. *Gene* **1985**, *33*, 103.

Yoshida, R.; Uchida, K.; Kaneko, Y.; Sakai, K.; Kikuchi, A.; Sakurai, Y.; Okano, T. *Nature* **1995**, *374*, 240.

Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471.

Yu, Y.; Monera, O. D.; Hodges, R. S.; Privalov, P. L. *J. Mol. Biol.* **1996**, *255*, 367.

Zhang, G.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1992**, *25*, 3601.

Zhou, N. E.; Mant, C. T.; Hodges, R. S. *Pept. Res.* **1990**, *3*, 8.

Zhou, N. E.; Zhu, B.-Y.; Kay, C. M.; Hodges, R. S. *Biopolymers* **1992**, *32*, 419.

Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.* **1992**, *267*.

Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol.* **1994**, *237*, 500.

Zhu, B.-Y.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1993**, *2*, 383.