OVEREXPRESSION, PURIFICATION AND BIOPHYSICAL STUDIES OF THE CARBOXY TERMINAL TRANSACTIVATION DOMAIN OF VMW65 FROM HERPES SIMPLEX VIRUS TYPE 1

Overexpression, Purification and Biophysical Studies of the Carboxy Terminal Transactivation Domain of Vmw65 from Herpes Simplex Virus Type 1

By

Logan W.F. Donaldson, B. Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

September, 1992

Master of Science (1992)
(Biochemistry)

McMaster University
Hamilton, Ontario

TITLE: Overexpression, Purification and Biophysical Studies of the Carboxy Terminal Transactivation Domain of Vmw65 from Herpes Simplex Virus Type 1

AUTHOR: Logan William Frederick Donaldson, B.Sc. (Lakehead), Dip. Hons. Standing (University of Western Ontario).

SUPERVISOR: Professor John P. Capone

NUMBER OF PAGES: *xi*; 142 pp.

ABSTRACT

In order to faciliate a biophysical analysis of the carboxy terminal acidic transactivation domain (AAD) of Vmw65 from Herpes Simplex Virus Type 1 (HSV-1), an overexpression system in *Escherichia coli* was constructed and optimized to produce milligram quantities of this polypeptide. Purification of the polypeptide was facilitated by creating a fusion protein to glutathione S-transferase (GST) from *Schizosoma japonicum* using a commercially available vector. Upon thrombin digestion of the fusion protein, the carrier and AAD products were resolved by anion-exchange chromatography.

With typically 15 mg of AAD available from a 12 litre culture, several biophysical studies were initiated. Circular dichroism and fluorescence spectroscopy both described a polypeptide with an extended structure reminicent of a random-coil; that is, it did not possess substantial quantities of known elements of secondary structure such as α-helicies and β-sheets under physiological conditions. A new structure high in α-helical content was induced upon addition of trifluoroethanol to mimic a hydrophobic milieu. Ultracentrifugation data supported the spectroscopic observations by describing an extended, monomeric polypeptide. The ultimate goal of the study, a teritiary structure, was sought by attempting to crystallize AAD with popular salts and organic solvents.

Biologically, the described random-coil structure of AAD could be relevant to its role as a promoter and stablizer of the transcriptional pre-initation complex, the determining step in gene expression. A structurally labile domain would support AAD's ability to interact with several targets including TFIID and TFIIB, though not necessarily by similar mechanisms. The requirement for a drastic conformational change such as a random-coil to

α-helical transition currently remains unclear though observations made in this study of AAD in trifluoroethanol have shown that a conformational change is indeed possible. With a means of producing large quantities of AAD, the opportunity now arises to study its interaction with available cloned targets. The ensuing biophysical studies will then provide a greater understanding of AAD's important role in gene expression.

DEDICATIONS

This thesis is dedicated to my parents, Karen and Logan Donaldson and my aunt / uncle, Margaret and Donald McKessock, for their constant support and encouragement throughout my education.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. John P. Capone, for his guidance throughout the course of these studies and for helping me to appreciate the biological relevance in a project based largely on physical techniques. I would also like to thank my committee members, Dr. Daniel Yang and Dr. Richard Epand, for advice and particularly, Dr. V. Ananthananarayanan, for the generous use of his spectrophotometric equipment. Dr. Preston Hensley is also acknowledged for sharing his expertise and providing key data.

I don't think that I could find a laboratory with more comraderie and support — I wish you all the best —Geoff, Dan, Richard, Peter, Bilyana, Pat and Shirley. The coffee clique certainly wouldn't be complete without KC and Debbie. Barbara, Karen, Richard P. and Steve — I enjoyed being your other roommate.

The author is grateful to the National Sciences and Engineering Research Council of Canada for support through a graduate scholarship (1990-1992).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AAD	acidic activation domain of Vmw65 (synonym for CT)
ATME	N-acetyl-tyrosine-methyl ester
BAPNA	N-α-benzoyl-arginine-*para*-nitroanilide
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BSA	bovine serum albumin
CAT	chloramphenicol acetyl-transferase
CT	carboxy terminal domain of Vmw65 (synonym for AAD)
DMSO	dimethylsulphoxide
EDTA	ethylenediaminetetraacetic acid
GST	glutathione S-transferase
HCF	Host Cell Factor (see VCAF-1)
IPTG	isopropyl B-D-thiogalactoside
KIU	kallikrein inactivating units
MPD	methylpentanediol
NBT	nitroblue tetrazolium
NIH	National Institutes of Health
NOE	Nuclear Overhauser Effect
NP-40	Nonidet P-40 detergent
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PEG	poly(ethylene glycol)
PMSF	phenylmethylsulfonyl fluoride
SDS	sodium dodecylsulfate
TBE	Tris-Borate-EDTA
TPCK	tosylphenylchloroketone
VCAF-1	Vmw65 Complex Associated Factor #1
VIC	Vmw65 Induced Complex

Chapter 1

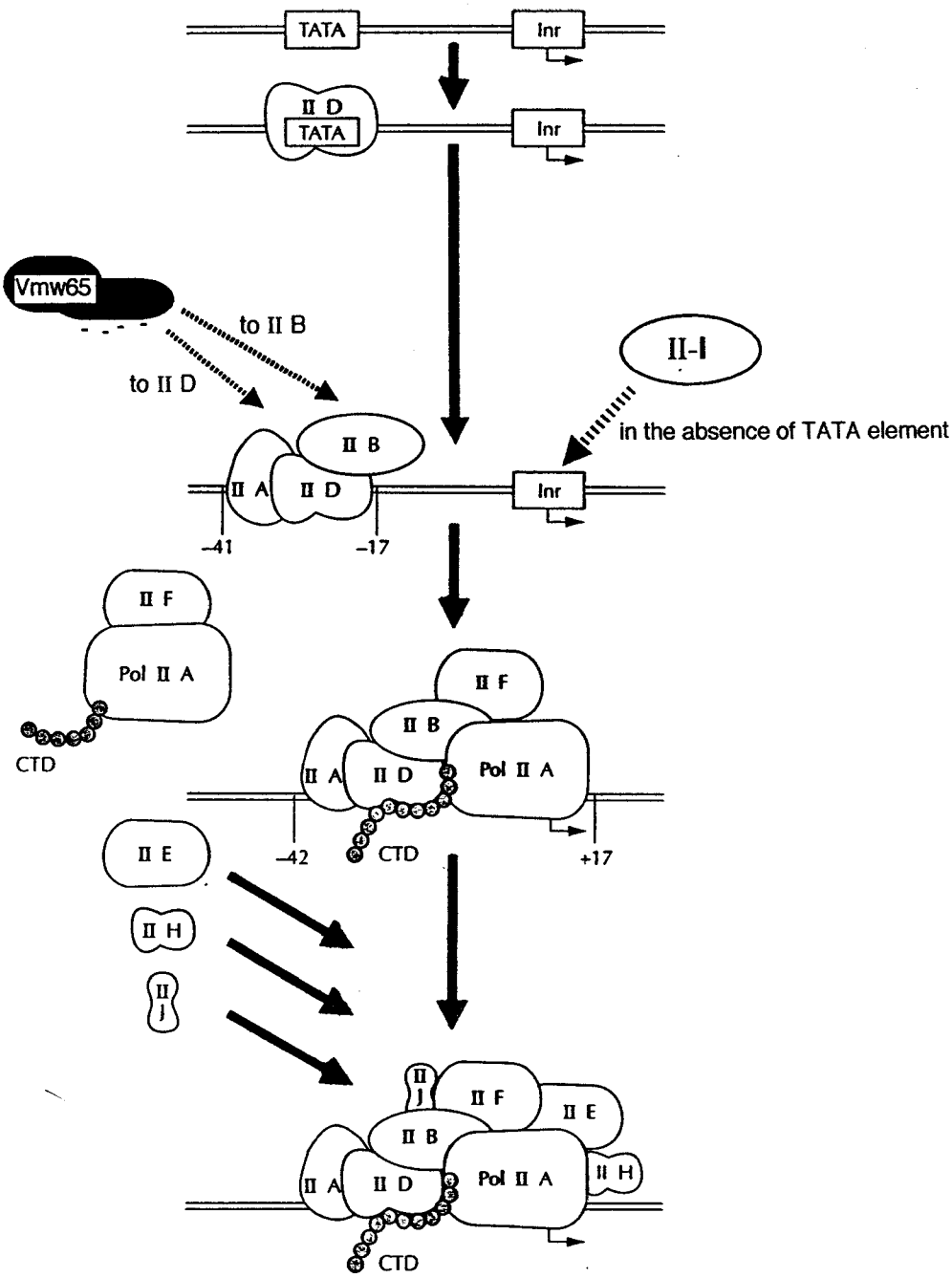
THE ROLE OF THE ACIDIC ACTIVATION DOMAIN IN GENE EXPRESSION

1.1. Introduction

Genes are transcribed at different rates according to how well they can present themselves to the transcriptional machinery; the limiting event being the formation of a functional pre-initiation complex. Included in the pre-initiation complex is RNA polymerase and a collection of core proteins termed general transcription factors.

Figure 1.1 depicts a current model (reviewed by Zawel and Reinberg, 1992) for the assembly of the RNA polymerase II based pre-initiation complex, involved in the synthesis of mRNA precursors. Seven transcription factors common to this pre-initiation complex are currently known. Each transcription factor is represented by a separate activity and may consist of one or more polypeptide chains. The remainder of this section will be devoted to

Figure 1.1. A Current Model of Pre-Initiation Complex Formation. Adapted from Zawel and Reinberg (1992), this model depicts the order at which the known transcription factors are believed to assemble. Complex assembly is initiated by the binding of TFIID to the TATA box which is stabilized by TFIIA. Next, TFIIB binds and provides a platform for RNA polymerase II (escorted by TFIIF). The roles of the remaining factors, TFIIE, TFIIH and TFIIJ are unclear though TFIIH has been implicated in disrupting an RNA pol II carboxy terminal domain (CTD) - TFIID interaction by multiple phosphorylations. Release of the CTD from TFIID is believed to be a key step in the transition to an elongation mode. In the absence of a TATA box, the initiator element (Inr) aids pre-initiation complex assembly by binding a set of specific proteins, including TFII-I, which then attract TBP and the other transcription factors. Vmw65, a member of the acidic class of transactivators, is known to specifically bind to TFIID and TFIIB.


TATA
Inr
II D
TATA
Inr
Vmw65
to II B
to II D
II-I
in the absence of TATA element
II B
II A
II D
Inr
-41
-17
II F
Pol II A
CTD
II F
II B
II A
II D
Pol II A
-42
CTD
+17
II E
II H
II J
II J
II F
II E
II B
II A
II D
Pol II A
II H
CTD

the description of each transcription factor in the order that they are believed to join the pre-initiation complex. Once the nature of the basal level transcription machinery has been established, other proteins reponsible for modulating gene expression can be introduced.

The primary event in the formation of the pre-initiation complex is the binding of transcription factor IID (TFIID) to the TATA box upstream of the initiation start site. Recently, TFIID has been shown to be a complex consisting of a TATA-box binding protein (TBP) and at least six other TBP-associated-factors (TAFs) in *Drosophila* (Dynlacht et al., 1991) whereas in human cells, there are at least ten TAFs (Pugh and Tjian, 1991) that are quite heterogenous in size (10-200 kDa). To date, TBP has been cloned from human, yeast, *Drosophila* and *Arabidopsis* and regardless of source, they can restore basal level transcription to a TBP-depleted HeLa cell extract *in vitro* indicating a similarity in overall function.

As shown by data from protein primary sequence comparison, (Hoffmann et al., 1990), TBP contains a highly conserved C-terminus and a divergent N-terminus. Surprisingly, experiments in yeast have indentified the C-terminal regions to be more important for conferring species specificity (Gill and Tjian, 1991). The N-terminal region of human TBP contains a stretch of 38 glutamine residues and five imperfect Pro-Met-Thr repeats. To date, no function has be ascribed to either component. The nucleic acid binding activity of TBP is conferred by two imperfect direct repeats separated by a basic region within the C-terminal domain (Horikoshi et al., 1992). Since TFIID binds as a monomer, it was suggested that the two direct repeats could function as a pseudo-dimer to accomodate the properties of known DNA binding proteins. Site-directed mutagenesis performed in the

basic region between the direct repeats neither reduced basal level transcription nor TATA binding suggesting a separate function (Lee et al., 1991).

The second factor to join the pre-initiation complex is TFIIA which consists of two subunits in yeast (Ranish et al, 1992). Upon binding to TFIID, it has been observed to stabilize the bound TFIID / promoter complex. In addition, indirect evidence for an anti-inhibitory role has been suggested with the discovery of an ATP-dependent, alkaline phosphatase reversible activity in yeast extracts (Auble and Hahn, 1992) which can dissociate TFIID from the TATA box but not after TFIIA has already bound. Through the systematic mutation of five lysine residues within the basic region of TFIID, the TFIID-TFIIA interaction was disrupted (Buratowski and Zhou, 1992) suggesting that this region was important for TFIIA binding. Since the genes encoding TFIIA have only recently been cloned, no mutants have been reported that likewise disrupt the corresponding TFIIA-TFIID interaction; however, a highly acidic region within the larger subunit is favored to be responsible.

Next, TFIIB, a single polypeptide of 33 kDa, binds to the TFIID-TFIIA complex. DNase protection assays have shown that the TFIIB footprint extends into the coding region suggesting that TFIIB may serve as a platform for RNA polymerase II to bind (Lin and Green, 1991). Since RNA polymerase II is dependent upon TFIIB binding, TFIIB is also believed to be the rate-determining step in the formation of the pre-initiation complex (Sharp, 1991). cDNAs from human, *Xenopus* and *Drosophila* have recently been obtained. Based on primary protein sequence analysis, TFIIB is found to be similar to TBP in that it contains a highly conserved C-terminus with imperfect direct repeats and two basic regions (Ha et al., 1991).

With the docking site established by TFIIB in place, TFIIF escorts RNA polymerase II to the pre-initiation complex. The RAP30 (Greenblatt et al., 1989) and RAP74 (Aso et al., 1992) components of TFIIF have recently been cloned. The availability of cloned products has allowed reconstitution assays to be performed and from this work, TBP, TFIIB, RAP30 and RNA polymerase II constitute a minimal factor requirement for transcription (Killeen et al., 1992).

The RNA polymerase II holoenzyme consists of 10 polypeptides in yeast with a total molecular weight of 550 kDa. Currently, a low resolution 16Å structure of yeast RNA polymerase II is known (Edwards et al., 1990).

Little is known about the role of the remaining three transcription factors, TFIIE, TFIIH and TFIIJ in the pre-initiation complex. TFIIH possesses a RNA polymerase II carboxy-terminal kinase activity which has been proposed by Laybourn and Dahmus (1990) to break the TBP tether through multiple phosphorylations and allow the elongation process to begin.

1.2. The Role of Transactivators

The genetic program of a cell requires genes to be transcribed at various rates and at specific times during its cycle. The general transcription factor can only accomplish basal level transcription; thus, other components are necessary for enhanced levels of transcription. Inclusion of promoter and enhancer sequences upstream to the initiation start site allow auxillary proteins termed transactivators to bind directly or indirectly to these sites, which in turn, interact with the pre-initiation complex to increase the rate of

transcription. The number and type of promoter sequences upstream of a gene provide a spectrum of enhanced rates over the basal level as numbers of transactivators can produce additive or synergistic effects. How a transactivator achieves transcriptional stimulation is unclear, though likely mechanisms would include (1) stablization of pre-initiation complex formation either by enhancing the rate of factor recruitment or by serving an anti-inhibitory role, and (2) promoting disassociation of the assembled pre-inititiation complex such that elongation may proceed. Of course, any stimulation of transcription will be further modulated by the inhibitory effects of repressor proteins and by the availability of the gene dictated by its position in the nucleosome.

1.3. The HSV-1 Transactivator, Vmw65

Herpes Simplex Virus (HSV) serves as an important system to study not only from its therapeutic importance but also as a simple model system to study transcription. Two serotypes (HSV-1, HSV-2) are highly homologous and they both contain a G+C rich double-stranded 150 kb DNA genome. The genome is enclosed by an icosahedral protein capsid which, in turn, is enclosed by a phospholipid membrane studded with eight types of glycoproteins (reviewed in Roizman and Sears, 1990). Between the membrane and capsid is the tegument which contains many proteins required upon infection.

The inclusion of various numbers and types of promoter sequences allows HSV-1 to express its >70 genes in three temporal classes, immediate-early (IE), early (E) and late (L). These classes are also known as α, β and γ. As illustrated in Figure 1.2, each class can be discerned by the promoters that they contain.

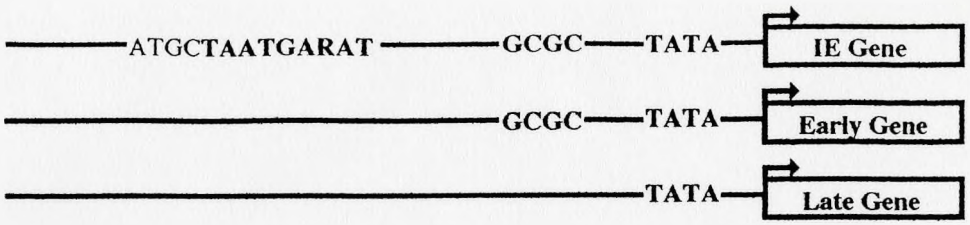


Figure 1.2. Generalized Upstream Regions of HSV-1 Genes. In the IE genes, the consensus TAATGARAT sequence is occasionally shared by the oct-1 ATGnTAAT consensus sequence.

Upon host-cell infection, five IE genes are initially expressed. Campbell et al. (1984) identified the gene responsible for the activation of IE gene expression as the late gene product, Vmw65 (also known as αTIF, ICP25, VP16). This polypeptide is manufactured as a late gene but it is delivered to the host cell upon infection as a component of the tegument region. It has been estimated that each virion contains 500-1000 Vmw65 molecules (Roizman and Spector, 1991).

Though Vmw65 manifests an electrophoretic mobility of 65 kDa, sequence analysis has shown that the predicted 490 residue product would have a molecular weight of 54 kDa (Dalrymple et al., 1985). Also known as VP16 or ICP25, it shares a 86% identity to its HSV-2 counterpart (Cress and Triezenberg, 1991a) and bears significant homology to the N-terminus of a Varicella Zoster Virus transactivator.

McKnight et al. (1987) showed that Vmw65 did not bind to its *cis*-acting site directly but through a cellular protein designated as αH1. Many laboratories were independently studying the same protein (OTF-1, OBP-100, NF-A1) which was finally identified as the ubiquitous cellular protein, Oct-1 (reviewed by Goding and O'Hare, 1989). Oct-1 is involved in the the transcription of many diverse classes of genes including immunoglobulins and small nuclear RNA's. Shortly after its identification, Oct-1 was

shown to specifically from a complex with Vmw65 (Gerster and Roeder, 1988; O'Hare and Goding, 1988; Stern et al., 1989). Subsequent mutational studies (Ace et al., 1988; Werstuck and Capone 1989ab; Greaves and O'Hare, 1989, 1990) have identified two Vmw65 (residues 141-185 and residues 317-403) that are responsible for complex formation.

The Oct-1 consensus binding sequence of ATGNTAAT overlaps the 5' end of the HSV-1 IE promoter. Oct-1 possesses a POU homeodomain (of a similar class as Pit-1 and Unc-1) which can be subdivided into a POU homeobox and a POU specific box (reviewed in Laughton, 1991). Homeodomains, in general, employ a helix-turn-helix recognition motif; for Oct-1, the POU homeobox recognizes the consensus sequence, TAATGT. The POU specific box recognizes the sequence ATGC upstream of the homeobox recognition site. The two POU subdomains together form a pseudodimer in Oct-1 to fulfill a dimerization requirement of the homeodomain class.

Vmw65 interacts with the Oct-1 homedomain in conjunction with an adaptor protein identified in HeLa cell nuclear extracts known as VCAF-1 (Xiao and Capone, 1990), C1 (Kristie and Sharp, 1990) or HCF (Stern and Herr, 1991). This adaptor binds to Vmw65 in the absence of DNA and participates in IE element recognition with Oct-1. At high concentrations, Vmw65 will bind to the IE element alone with particular specificity to the GARAT portion of the IE element which lies 3' to the Oct-1 recognition site.

In summary, binding of Vmw65 to promoter sequences is mediated through the ubiquitous factor Oct-1, and further assisted by an uncharacterized factor, VCAF-1. With

Vmw65 now conceptually anchored to its promoter, this discussion can now focus on the structural elements that confer the ability to stimulate transcription.

1.4. A Potent Activation Domain

By removing the C-terminal 79 amino acids, transactivation can be eliminated (Greaves and O'Hare, 1989; Cousens et al., 1989) without affecting complex formation (Werstuck and Capone, 1989). On the basis of primary protein sequence analysis, this region is highly acidicwith a preponderance of aspartic and glutamic residues. Sadowski et al. (1988) showed that this region could act as a discrete acidic activation domain (AAD) by fusing it to the DNA binding domain of yeast GAL4 (amino acids 1-147). Consequently, this fusion was able to stimulate the expression of a co-transfected CAT gene bearing a GAL4 promoter sequence. The promoter itself was highly position independent as high expression was observed if the promoter was placed at -110, -1180 or even at +1850 relative to the initiation site. Prior to this experiment, the 881 residue GAL4 polypeptide had been shown from deletion analysis to possess two acidic activating regions, (Ma and Ptashne, 1987) one internally positioned at residues 148-196 and another C-terminally positioned at residues 768-881. Since GAL4 and Vmw65 both contained AADs, they could be compared in domain-swap experiments (Sadowski and Ptashne, 1988). It became evident that acidity was not the only determinant of activity since the GAL4-Vmw65 chimera transactivated 100 times greater than wild-type GAL4 in their assay. Several other transactivators that contain AADs are known; these include yeast GCN4 with one internal domain (Hope and Struhl, 1986), yeast PUT3 with one internal and one C-terminal domain (Marczak and Brandriss, 1991) and human p53 with one N-terminal domain (Fields and

Jang, 1990). The Vmw65 AAD is generally chosen for studies since it is the most potent of its class.

Other regions in transactivators have been identified that are not acidic but are nevertheless peculiar in their primary sequence (reviewed in Mitchell and Tjian, 1989; Ptashne and Gann, 1990). *Drosophila* Antennapedia and human Sp1 (Courey and Tjian, 1988) for example, contain glutamine rich regions and are modular like the AAD. A proline-rich domain has been observed in the CTF gene product (Mermod et al., 1989). The mechanism of action of these three classes of transcriptional activating regions remains highly speculative at this time.

1.5. Potential Targets of the AAD

In a search for a transactivator target, pivotal components in the formation of the pre-initiation complex would be likely candidates. The binding of TFIID to the TATA promoter serves as the primary step in complex formation and is too obvious not be exploited by both positive and negative effectors. TFIIB is also a reasonable candidate since it is required for RNA polymerase II binding. As RNA polymerase II performs the actual work, one of its subunits could easily be a target. Of course, these are only the obvious targets; one must never rule out the many auxiliary factors that aid production of the pre-initiation complex. The aim of this section is to briefly summarize some of the work done towards the identification of targets for Vmw65 and to introduce some principles that may govern transactivation.

Like AAD, RNA polymerase II possesses regions of unusual primary sequence. Specifically, the yeast pol II gene, RPO21, was found to contain a 21 bp sequence that had been tandemly repeated 26 times which would correspond to a heptapeptide consensus repeat of (tyr-ser-pro-thr-ser-pro-ser) in the C-terminal doman (CTD) (Allison et al., 1985). After the discovery of several AADs, Sigler (1988) suggested that the acidic residues may interact ionically with the hydroxyl moieties of pol II. Assuming that AAD had no particular structure — a negative noodle, of sorts — Sigler proposed that AAD could perhaps tether itself to the polymerase in a conformationally indefinite manner. Though no data has been published to test this proposal, it still introduced an important principle of transactivation to consider; that is, genes could be transcribed faster if the pre-initiation complex was stabilized. The addition of more AADs could serve to anchor the polymerase, further explaining experimentally observed synergistic effects. Of course, the AAD-polymerase complex would have to be sufficiently weak to prevent the polymerase from being stuck at the initiation site leading to the idea that perhaps many weak interactions may be better than a few strong interactions.

Via affinity chromatography using coupled Protein A-AAD, TFIID was selectively retained and became the first transcription factor described to interact with AAD (Stringer and et al., 1990). On the basis of affinity chromatography alone, it could be argued that binding of TFIID could be non-specific since the protein has a basic pI which could interact avidly with the AAD. A follow-up study (Ingles et al., 1991) using a set of AAD mutants designed by Cress and Treizenberg (1991b) showed that TFIID would not bind to a AAD point mutant where a phenylalanine residue had be substituted with a proline residue suggesting that the AAD-TFIID interaction was specific. Like Vmw65, adenovirus E1A possesses an internal AAD which also binds to TFIID whose interaction has been

characterized beyond chromatographic methods (Lee et al., 1991; Horikoshi et al., 1991). The AAD of E1A was implicated in the formation of the E1A-TFIID heterodimer by the observation that the association was much stronger in the 13S product over the 12S product which lacks an AAD as a result of alternative mRNA splicing. Corresponding deletion mutants in TFIID that compromised E1A-AAD binding identified the basic region within the C-terminal conserved domain previously shown to interact with TFIIA (Lee et al., 1991).

Other research (Lin and Green, 1991; Lin et al., 1991; reviewed by Sharp, 1991) insisted that TFIIB made a higher affinity contact with AAD than TFIID when a similar method using a fusion of AAD to glutathione S-transferase instead of Protein A. Though TFIIB remained coupled to the column at 1 M NaCl which could superfically imply a strong interaction, there are instances where affinity cannot be correlated with the ionic strength (Formosa et al., 1991). Placement of the TFIID/TFIIB basic regions on a helical wheel revealed that the contributing basic residues of TFIID were dispersed over a larger surface than either of the two basic regions found in TFIIB (Yamashita et al., 1992). Since two different motifs (by secondary structure predictions, at least) are present, it is reasonable to assume that more than one recognition region may be present in AAD and that these recognition regions may operate independently or together to bind the two factors. TFIIB has only recently been cloned; hence, mutational studies have not yet been performed.

Other proteins collectively called adaptors (Berger et al., 1990; Martin et al., 1990) have been cited as targets of transactivators. These nebulous cellular factors could possibly be TAFs or uncharacterized components of the pre-initiation complex or could simply be

factors that play a transient, yet defined role in pre-initiation complex assembly. Only recently have partially purified adaptor containing fractions been reported (Davidson et al., 1992) that cooperate with Vmw65-AAD and other AADs to enhance transcription. Preliminary data from Berger et al. (1992) is particularly exciting as they have sequenced two genes from yeast that can complement mutants that have reduced GAL4-VP16 mediated transcription but normal basal level transcription.

Experimental evidence for multiple targets may be a necessary requirement for enhancement of transcription over the basal level. Lin et al. (1991) point out that at least two binding sites on the promoter are required for AAD and for weaker transactivators, as many as five binding sites may be required. Proposed mechanisms for any of the postulated targets still remain highly speculative.

1.6. The Pursuit of Structure

While the search for new targets continue, existing targets may be characterized structurally to determine if they possess common motifs. Though the pursuit of structure at this point may seem premature, the knowledge provided by such endeavors allows one to treat the various targets and factors as building blocks that can be rationally assembled.

Since many factors have only been recently cloned, structural information has only been obtained through mutagenesis and sequence comparison. Even at this rudimentary level, a degree of modularity appears. For example, TFIID and TFIIB are homologous in their primary structural motifs. Frankel and Kim (1992) have extended the idea of modularity to be a governing factor in the formation of transcription complexes. In their

model, specificity is gained as disordered, inactive modules undergo conformational transitions to the ordered, active modules. Since each module is dependent upon another module in a sequential fashion, a temporal aspect is gained which is missing in the current model of pre-initiation complex formation. Being an autonomous domain, AAD already possesses a degree of modularity. In conjunction with a unknown structure, AAD could be an interested candidate to test the modularity hypothesis.

1.7. Research Aims

Though molecular biological methods of structural analysis gives beneficial results in the context of an actual living system, a biophysical approach is required to directly determine the structure of AAD.

It is important to determine the structure of AAD for several reasons: (1) No structure of any transactivation domain is known. Since AAD is predicted to have many peculiar qualities, an unusual structure could potentially modify current opinons on protein structure. (2) Determination of its inactive and active structures could allow a mechanism to be derived which could be extended to other acidic transactivators and possibly other classes. (3) A structure could provide the basis for the design of rationally designed inhibitors or modifiers of function as a therapy for HSV-1 infection. In the widest perspective, all of these endevours would surely advance the understanding of gene expression.

To study the secondary structure of AAD, the techniques of circular dichroism spectroscopy, fluorescence spectroscopy and equilibrium ultracentrifugation were applied.

Crystallization trials were initiated as the crystallographic technique would provide the ultimate in analysis, a tertiary structure.

In order to perform the studies, milligram quantities of homogenous protein are necessary. Though a baculovirus system was available to synthesize Vmw65 (Capone and Werstuck, 1990), a bacterial system was sought to provide higher yields. Implicit in this decision was the development of an adequate expression system and a corresponding purification scheme.

This thesis was rationally divided into two chapters being the overexpression / purification of AAD and its biophysical study. A chronological order was deemed best to describe each study as it was undertaken. Since a wide range of techniques were used, sub-chapters have been installed to individually introduce the theory, materials/methods, results and conclusions of each technique or phase in the project.

Chapter 2

THE OVEREXPRESSION AND PURIFICATION OF VMW65 RELATED POLYPEPTIDES

2.1. Introduction

A prokaryotic expression system is a choice method for large scale protein production primarily due to the few technical demands it places on the investigator. This choice becomes even more desirable when one considers the numerous vectors that are now available to enhance experimental flexibility. Caution though, must be taken in the form of functional assays to ensure that if a eukaryotic protein is expressed, it is functional since *Escherichia coli* does not perform post-translational modification. If this problem is circumvented, many exciting possibilities arise such as structure determination, antibody production, cross-linking studies, affinity chromatography and reconstitution assays. The aim of this section is to mention some interesting overexpression systems that are currently

commercially available to the user and briefly discuss their attributes. Many excellent reviews are available as a collection of articles in Methods in Enzymology, vol. 185 (1990) and Current Protocols in Molecular Biology, vol. II (1989) and Sherwood (1991).

As a result of extensive study over several decades, bacterial vectors can provide many options to the researcher who wishes to use overexpression. The first consideration one makes is the localization of the overexpressed product. In the absence of a leader (or signal) sequence, the expressed protein will remain intracellular. If a leader sequence is present in the inserted gene or has been intentionally placed in the vector, the protein will be secreted to the periplasmic space. Secretion to the periplasm is advantageous when inclusion bodies (dense, aggregates of overexpressed protein) or a susceptibility to intracellular proteolysis is suspected. Lysis of the bacteria can be exchanged for gentler methods like osmotic shock but the fragility of these bacteria creates a risk of protein loss through leakage into the culture medium. If the facility exists to process media, complete secretion may be beneficial and can be accomplished by cloning in the coordinate sequence of a secreted bacterial protein. Though inclusion bodies are generally considered to be undesirable, their size can be exploited to achieve a higher initial state of purification by selectively centrifuging then resolubilizing them in the presence of a chaotropic salts. Removal of the salts does not guarantee, however, that the protein will refold into its native conformation. Degradation may be alleviated by using a strain deficient in the *lon* protease though it is more technically difficult to grow and process this strain. It is evident that design of of overexpression scheme will not only require an appreciation of the microbiology involved but also some investigative (trial-and-error) experiments.

To facilitate purification of the overexpressed protein, it can be cloned adjacent to a carrier protein creating a fusion (or chimera) that can be purified via affinity chromatography using the carrier-specific ligand. Table 2.1 lists some prokaryotic expression vectors that are available commercially. Upon inspection of this collection of vectors, there are more similarities than differences. The promoters used in these vectors are extremely strong and inducible by a non-metabolizable analog, heat shock or by addition of a foreign polymerase. To prevent expression prior to induction, a overexpressed repressor is available on the plasmid. Heterogenity lies in the carrier protein with the trend being to use smaller polypeptides which still retain a high specificity to their affinity ligand. Other traits of a well chosen (and recently, designed) carriers include low antigenicity for immunological purposes and high solubility within the bacterium since aggregates pose a serious threat to overexpressed protein's activity. Often for molecular biological and immunological uses, the overexpressed protein does not need to be liberated from its carrier but the need may arise in a structural study or if the carrier is obscuring an activity. In that case, a vector should be chosen that contains a highly specific protease sensitive site. One obvious requirement for a site would be to ensure that the site rests on an accessible surface or hinge-like region of the carrier. Though there are programs that are available to predict similar sites in the protein of interest, the best method is to experimentally determine if any exist since there may be cryptic sites that diverge from the consensus sequence.

2.2. Prior Cloning

It was originally hoped that a biophysical study of the acidic activation domain (AAD) would be expedited by the availability of two constructs in the laboratory. Both

Table 2.1 - Commercially Available Overexpression Vectors

Plasmid	Carrier [size]	Ligand	Promoter	Induction	Protease...	Supplier	Comments
pKT280	none	none	none	N/A	N/A	Clontech	for secretion via beta-lactamase signal
pKK233-2	none	none	*trc*	IPTG	N/A	Clontech	general expression
pKK233-3	none	none	*tac*	IPTG	N/A	Pharmacia	general expression
pMC-1871	B-galactosidase [116 kDa]	none	none	N/A	N/A	Pharmacia	Susceptible to proteolysis
pET-xx	T7 gene 10 [2-5 kDa]	rabbit antisera	T7 lac	IPTG	none	Novagen	uses lysogens of phage lambda DE3
pRIT2T	Protein A [31kDa]	IgG	lambda HS	42°C	trypsin	Pharmacia	Protein A tail is generally innocuous
pEZZ18	ZZ Synthetic peptide [14 kDa]	IgG	spa/UV5	N/A	N/A	Pharmacia	has Protein A sequence for secretion
pGEX2T	glutathione S-transferase [28 kDa]	glutathione	*tac*	IPTG	thrombin	Pharmacia	specific cleavage sequence
pGEX3X	glutathione S-transferase [28 kDa]	glutathione	*tac*	IPTG	Factor Xa	Pharmacia	alternative cleavage seq. to above
pMAL-x	maltose binding protein [40kDa]	amylose	*tac*	IPTG	Factor Xa	N.E.B.	kit contains carrier antisera
pRSET	T7 gene 10 [2-5 kDa] + poly(his)	metal affinity	T7	T7 pol.	enterokinase	Invitrogen	has OmpA signal to periplasmic space
pQE-xx	6 x histidine [0.7 kDa]	Ni-NTA chelator	T5	IPTG	none	Qiagen	can use DHFR tag for added stability
pFLAG-1	synthetic polypeptide [1-7 kDa]	monoclonal Ab	*tac*	IPTG	enterokinase	I.B.I	carrier designed to be very soluble

Entries are listed in a rudimentary order of complexity based on fusion protein capablility, ease of induction, possibility of carrier cleavage and molecular weight of the polypeptide. Abbrevations: IPTG, isopropyl β-D-thiogalactoside; lambda HS, λ phage heat shock promoter; N.E.B., New England Biolabs; T7 pol., T7 phage RNA polymerase; I.B.I., International Biotechnologies, Inc.

constructs were fusion proteins either to the Protein A binding domain from *Staphylococcus aureus* (Nilsson and Abrahmsén, 1990) or to the C-terminal portion of the Sj26 antigen from *Schitzosoma japonicum*, a glutathione S-transferase (GST) (Smith *et al.*, 1986).

Both the Protein A and GST based systems offer many advantages including ease of inducibility, high yields in the milligram per liter culture range and one-step purification using an affinity matrix. Protease sensitive sites also exist at the junction of the two fused proteins to facilitate cleavage, if necessary.

2.3. Overexpression of a Protein A-AAD fusion

2.3.1. Materials and Methods

2.3.1.1. Construction

The Protein A fusion, pRIT-CT (where CT = carboxy-terminus), was generated by excising a 500 bp *Sac* II/*Bam* HI fragment from an available Vmw65 containing plasmid and inserting it into the BamHI site of the general overexpression vector pRIT-2T (Uhlén et al., 1983) from Pharmacia LKB (Uppsala, Sweden) with the appropriate 5' *Sac*II/*Bam* HI linker such that the AAD was in-frame with Protein A. As the AAD polypeptide is approximately 9.3 kDa in size, the expected fusion was predicted to be 41 kDa total. Trypsin is used to cleave AAD from Protein A. Since trypsin will cleave after the basic residues arginine or lysine, three products of 89, 82 and 81 residues were expected dependent upon whether trypsin cleaved at arginine residue within the vector multiple cloning site or at either of the two arginine residues at positions 408 and 409 within the AAD. Figure 2.1 contains specific information regarding the nucleic acid and

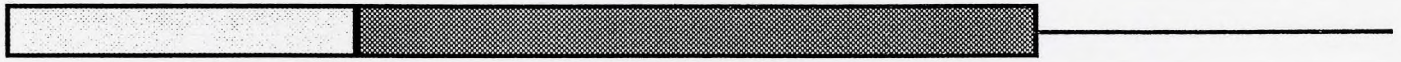
Figure 2.1: Constructs Used to Overexpress AAD. To simplify purification and to increase the likelihood of a soluble product within the bacterium, AAD was expressed either as a fusion to Protein A (pRIT-CT) or Glutathione S-Transferase (pGEX-CT, pGEX-CT/TGA). A graphical representation (not to scale) is provided to show where the carrier sequence terminates and the AAD sequence originates. Above the predicted nucleic acid sequence in the vicinity of the fusion junction, the translated protein sequence is given. Below the nucleic acid sequence, the numbers correspond to amino acid position in the wild-type Vmw65 sequence; those designated (---) include residues belonging to the carrier or new residues created during construction. Amino acids enclosed in (brackets) would result from translational suppression of the amber termination codon. An arrow depicts a predicted protease cleavage site either for trypsin (Protein A fusions) or for thrombin (Glutathione S-Transferase fusions). Restriction sites relevant to the cloning scheme employed are included for reference.

pRIT-CT


```
ser  arg  gly  ser  ala  gly  his  thr  arg  arg  glu  tyr  gly  gly
CCGCGTGGATCCGCGGGACACACGCGCAGA...GAGTACGGTGGGTAGGGGGCGCGACCG
          ---  ---  404  405  406  407  408  409  487  488  489  490  ***
```

Bam HI

pGEX-CT/TGA




```
pro  arg  gly  ser  ala  gly  his  thr  arg  glu  tyr  gly  gly
CCGCGTGGATCCGCGGGACACACGCGC...GAGTACGGTGGGTGATAACTGCAGCTG
          ---  ---  404  405  406  407  408  487  488  489  490  ***  ***
```

Bam HI, Rsa I, Pst I, Pvu II, Eco RI

pGEX-CT


```
pro  arg  gly  ser  ala  gly  his  thr  arg  glu  tyr  gly  gly  (gln) (gly) (ala) (arg) (pro) (asp)
CCGCGTGGATCCGCGGGACACACGCGC...GAGTACGGTGGGTAGGGGGCGCGACCGGAC
          ---  ---  404  405  406  407  408  487  488  489  490  491  492  493  494  495  496
                                                              ***
```

Bam HI

```
(pro) (his) (pro) (pro) (ser) (gly) (phe) (ser) (pro) (pro) (val) (thr) (gly) (ser) (tyr) (pro) (gln)
CCGCATCCCCCGTCTGGGTTTTCCCCTCCCGTCACCGGTTCGTATCCACAATAA
497  498  499  500  501  502  503  504  505  506  507  508  509  510  511  512  513  ***
```

corresponding amino acid sequence at the junction point of the fusion mentioned in this chapter.

2.3.1.2. Bacterial Strains

Overexpression of pRIT-CT is governed by the λ P_R promoter and is induced by temperature inactivation of an endogenous λ cI_{857} repressor. The specific strain used was *E. coli* N4830-1: [F^-, *su-*, *his-*, *ilv-*, *gal*KΔ8, Δ(*chl*D-*pgl*), λ, ΔBam, N+, cI_{857}, ΔH1].

2.3.1.3. Culture and Processing

Initially, the protocol similar to that of Werstuck and Capone (1989) was used to culture and purify the fusion protein. To 500 mL of 2-YT medium (1.5% peptone, 1.0% yeast extract, 0.5% NaCl) containing 50 μg/mL ampicillin (Boeringer-Mannheim), 20 mL of an overnight culture grown at 30 °C was inoculated. The temperature was maintained at 30 °C until the OD_{600} of the medium reached approximately 0.5 - 0.6. Expression of the fusion protein was initiated by adding an equal volume (500 mL) of 2-YT medium which had pre-warmed to 60 °C. The resulting temperature of the 1 L culture was approximately 42°C at this point. Expression was allowed to continue for 45 minutes at 37°C.

Cells were harvested at 4°C, as all subsequent steps, in a Beckman GSA rotor at 5000 rpm for 10 minutes. The pellet was washed once with 50 mM Tris-Cl, pH 7.6 at resuspended in lysis buffer (50 mM Tris-Cl, pH 7.6, 0.25 M NaCl, 0.01 M EDTA, 0.1 % NP-40, 13 % sucrose, 1 mM PMSF, 0.2 mg/mL lysozyme) and incubated (at 4°C) for 30 minutes. The viscous mixture was then carefully transferred into ultracentrifuge tubes and

centrifuged at 35000 rpm in a Beckman 50.2 vertical rotor at 4°C. The supernatant was removed, flash-frozen under liquid nitrogen and stored at -70 °C until required.

As this method was time/labour intensive, an alternative method was found to be more desirable to process the bacteria. After harvesting, cells were resuspended in 5 mL of 50 mM Tris, pH7.6 + 1 mM PMSF and lysed for 3 x 30 sec using a Braun Sonic 2000 sonicator. Lysis was deemed completed when the solution cleared slightly and became straw brown in colour. This mixture was then centrifuged at 17000 rpm for 15 minutes in a Sorvall SS-34 rotor. The supernatant was then flash-frozen under liquid nitrogen and stored at -70 °C until required.

2.3.1.4. Purification

A small 6 cm x 0.75 cm (ID) column (Bio-Rad) of IgG-Sepharose (Pharmica LKB) was prepared and equilibriated with 40 mL of TST buffer (50 mM Tris-Cl, pH 7.6, 150 mM NaCl, 0.05% Tween-20). The crude extract, thawed quickly under 30°C water and stored on ice, was applied and the column washed with several column volumes of TST until no protein could be detected by Bradford microassay (Bio-Rad) and then washed with 25 mL of 5 mM ammonium acetate, pH 5.0. The bound protein was eluted with 0.5 M acetic acid, pH 3.4 and concentrated with a Centricon 30 filtration unit (Amicon) to 2.5 mL. The concentrate was applied to a PD10 desalting column (BioRad) pre-washed with Buffer A (10 mM HEPES, pH 7.9, 50 mM NaCl, 1 mM EDTA, 1mM DTT, 10 % glycerol) and eluted with 3.0 mL of the same buffer.

2.3.1.5. Removal of Carrier Protein

The Protein A-AAD fusion was typically treated with a 1:50 fusion/trypsin ratio in 50 M Tris-Cl pH 8.0 at 37°C. The bovine pancreatic trypsin type XIII used was of high purity (Sigma, Stock T-8642) and TPCK treated to reduce any residual chymotrypsin contamination. Tryptic activity was confirmed by monitoring the evolution of *p*-nitroaniline at 410 nm in a freshly prepared mixture of 0.5 mM N-α-benzoyl-arginine-*p*-nitroanilide (BAPNA) (Sigma) in 50 mM Tris-Cl, pH 8.0. To prevent hydrolysis during storage, a 400 mM BAPNA stock solution in DMSO was kept at -20°C. The incubation was stopped with 2x SDS loading buffer (125 mM Tris-Cl, pH 6.8, 4 % SDS, 20% glycerol, 10 % β-mercaptoethanol) prior to loading the products onto a denaturing polyacrylamide gel or stopped by the addition of PMSF to 1 mM.

2.3.1.6. Analysis of Purification / Carrier Cleavage

To determine protein concentration, a Bradford microassay was employed against a BSA fraction V (Pierce) standard. For analysis, an appropriate quantity of desalted eluant from the IgG-Sepharose affinity column was run on a 10 % SDS-polyacrylamide gel using a BioRad minigel apparatus. BDH Electran 17-78 kDa protein standards or 14-94 kDa BioRad protein standards were used.

Western blotting was performed on selected minigels using a BioRad minigel module. Transfer to nitrocellulose was generally accomplished in 1.5 hours at 0.3 A using a standard Western transfer buffer (Sambrook *et al.*, 1989) supplemented with 0.5 % SDS. The nitrocellose was incubated for 15 minutes in TBST (10 mM Tris-Cl pH 8.0 , 150 mM NaCl, 0.5 % Tween-20) containing Carnation Instant Milk to block and then was washed twice for 15 minutes with TBST to clear excess milk. Rabbit polyclonal antisera raised

against affinity purified GST-AAD fusion protein was applied in TBST at a 1:2000 dilution. After two washes, goat anti-rabbit conjugated alkaline phosphatase (BioRad) was applied at a 1:3000 dilution. Development was initiated by adding BCIP/NBT (Promega) in 50 mM Tris-Cl, pH 9.0 and was stopped with 30% acetic acid.

2.3.2. Results

From a 1 litre culture, approximately 2 mg of fusion protein was obtained with much variation between cultures. SDS-PAGE (Figure 2.2) revealed two prominent bands of 70 kDa and 41 kDa in visually equal amounts with some minor bands attributed to degradative products. The 70 kDa protein was believed to be a bacterial protein that possesses a high affinity for IgG-Sepharose beads. The 41 kDa band was believed to be the Protein A-AAD fusion protein, though this was impossible to determine as the polyclonal antibody also reacted with the 70 kDa band (blot not shown).

As also shown in Figure 2.2, trypsinization (1 mg trypsin / 50 mg fusion protein) was very efficient as evidenced by the disappearance of the 41 kDa band even after as little as 15 minutes at 25°C. The 70 kDa band did not appear to be hydrolysed. However, an AAD product (Protein A was presumed to be hydrolysed into many small fragments) was not visually evident on the gel in a complex background. The gel was not run to completion; that is, the bromophenol blue dye front was not run off the gel so all products from the trypsinization should have been present. Autodigestion of the crude extract was negligible during the 60 minute digestion. A Western blot of a similar gel did not identify the presence of the AAD product.

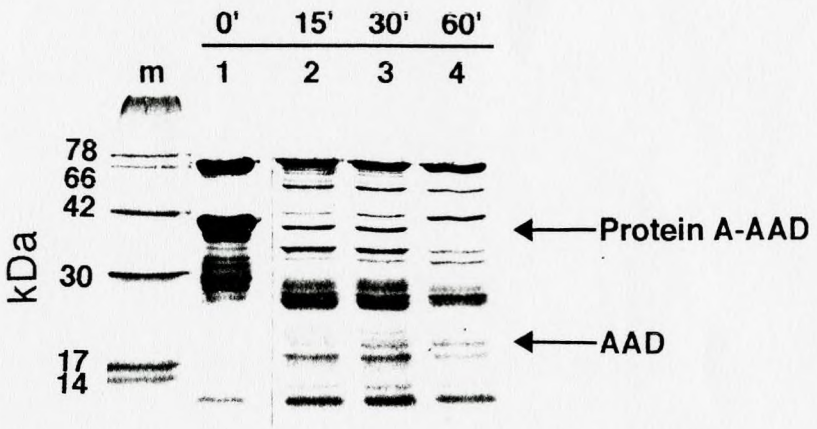


Figure 2.2: Protein A-AAD Overexpression and Trypsin Mediated Cleavage. 10 μg of the proteins eluted from a IgG-Sepharose column were treated with TPCK-trypsin for a period of 15-60 minutes at 25°C and subjected to 12% SDS-PAGE with Coomassie blue staining. *Lane m:* marker proteins. *Lane 2:* Self-incubation of eluate for 60 minutes. *Lanes 3-5:* Trypsin treatment for 15, 30 and 60 minutes.

2.3.3. Conclusions

The fact that no AAD was produced upon trypsinization was puzzling since there were only two arginines in the polypeptide and they were both very near the N-terminus. If the trypsinization problem was overcome and AAD was detected, there were still other problems to address. Though the yield of fusion protein was over 1 mg/L culture and was quite sufficient for molecular biological uses, this yield was very low for biophysical studies which require quantities in the tens of milligrams. The presence of a contaminant in similar amounts would still require the addition of one or several additional chromatography steps beyond affinity chromatography which would likely reduce yields further. It should also be remembered that the elution conditions of Protein A fusions are harsh and could potentially affect activity.

The fusion point between Protein A and AAD had not been sequenced; hence, a nonsense product may have be overexpressed. Since another overexpression vector, GST-AAD, had not yet been tested, an investigation of the Protein A-AAD anomalous behaviour was waived until the outcome of the GST-AAD fusion was known.

2.4. Overexpression of a GST-AAD fusion

2.4.1. Materials and Methods

2.4.1.1. Construction

Previously, the GST fusion was constructed by simply excising the *Bam* HI fragment from pRIT-CT and inserting it into the *Bam* HI site of the general overexpression vector pGEX-2T (Smith and Johnson, 1988) from Pharmacia LKB (Uppsala, Sweden) creating plasmid pGEX-CT.

2.4.1.2. Production

pGEX-CT was provided in the common strain, *E. coli* HB101. This vector is suitable for most strains since it contains the required *lac* repressor on the plasmid. Culturing conditions were very simple involving inoculation with an overnight culture and growth in 2-YT medium at 37°C supplemented with 50 μg/mL ampicillin up to mid-log phase (OD_{600} = 0.5). Induction was accomplished via addition of a 100 mM IPTG solution to 0.1 - 0.2 mM final concentration. Initial periods of inductions varied from 30-45 minutes. Harvesting, lysis by sonication and additional processing was similar to the procedure used for the Protein A-AAD fusion except the lysis buffer was PBS (10 mM phosphate, 150 mM NaCl, pH 7.4) supplemented with 1 mM PMSF. Initially, Triton X-100 was added to the opaque crude extract to 1% with the expectation of increasing the fusion's solubility but, in fact, the crude extract became cloudier and required an additional centrifugation to remove the precipitate.

2.4.1.3. Purification

At the initial scale used (250 - 1000 mL culture) small columns of a 1-4 mL were only required since the binding capacity of glutathione-Separose CL6B (Pharmacia) was specified at 5-7 mg/mL matrix. Glutathione (γglu-cys-gly), can be covalently attached to an activated matrix in several ways. S-hexylglutathione-agarose (Sigma #H7011) has been substituted without problems, on occasion.

If the crude extract had previously been stored at -70°C, it was recentrifuged and applied to a column washed with PBS containing 0.02% sodium azide as a microbial (unless otherwise stated, all column buffers contain sodium azide). The column was then washed until no protein could be detected by Bradford microassay. Elution was

accomplished by passing a 50 mM Tris-Cl, pH 8.0 solution containing 5-10 mM reduced glutathione to compete the fusion protein from the beads. The higher concentration of glutathione was more beneficial as the eluted protein did not tend to tail-off the column as much. The affinity column was regenerated with 3M NaCl in PBS and then extensively washed with PBS. The eluted protein was concentrated if required and stored at -20 °C.

Alternatively, a batch procedure could be employed rather than a column for small cultures (< 1000 mL). In a Corning or any tapered tube, a small quantity of beads (1 mL) was placed and washed with PBS. Beads were collected by centrifuging for 1 min at 3000 rpm in an IEC swinging bucket rotor. Care had to be taken when removing the supernatant since the beads would not pellet tightly. The protein sample was then mixed with the beads and rocked at room temperature for 5 minutes. The beads were then again collected by centrifugation and extensively washed (over 5 times) with PBS. Elution was accomplished by incubating the beads with 5 mM glutathione in 50 mM Tris-Cl, pH 8.0 for 5 minutes on a rocker. After centrifugation the supernatant was collected. The affinity beads were washed once with the free glutathione mixture and the supernatant pooled with the first. In summary, with the availability of 1-2 mL disposable plastic columns, this method is not recommended since the centrifugations are quite time consuming and there is always some bead loss.

2.4.1.4. Removal of Carrier Protein

Liberation of the GST portion of the fusion is accomplished by thrombin. As it is more specific having a consensus sequence of P4-P3-pro-arg-P1'-P2' where P3 and P4 are hydrophobic residues and P1', P2' are nonacidic residues *or* P2-Arg-P1' where either P2

or P1' are glycine residues (Chang, 1985), thrombin was only expected to cleave once in the vector multiple cloning site to produce an 89 residue product (consult Figure 2.1).

Thrombin can be purchased from several suppliers over a large range of purity. The thrombin (Sigma Stock T3010) was highly purified from human blood plasma and was supplied in concentrated form. Purity is expressed as a specific activity; in this case, 4430 NIH units/mg enzyme. As lots of this product do not vary significantly from this specific activity, I have used two conventions to describe the amount of thrombin added to a fusion protein preparation. The first is simply a mg/mg ratio of thrombin/fusion protein. The second method expresses the ration in NIH units of thrombin/mg of fusion protein. One can easily convert ratio to NIH units/mg by simply multiplying the former by 4430.

The optimum amount of thrombin required to cut the fusion over a 30 minute period at 30°C was determined by adding successive amounts of a stock solution of thrombin in 50 mM Tris-Cl, pH 8.0 (diluted from the parent concentrate) to tubes containing 10 μg of fusion protein. A quantity of GST-AAD was also digested with trypsin. Reactions were stopped with 2x SDS loading buffer and a quantity was run on a 10% SDS-polyacrylamide gel.

2.4.1.5. Analysis

Similar to the Protein A-AAD fusion, denaturing polyacrylamide gels were run and either stained with Coomassie blue or transfered to nitrocellulose and probed with polyclonal antisera raised against Protein A-AAD (specific to the AAD portion) or GST-PEA3 (specific to the GST portion). Silver staining (BioRad) was used according to manufacturer's instructions.

One novel stain called Stains-All (Eastman-Kodak) has the unique property of staining lipids yellow, proteins red and acidic biomolecules blue (Green *et al.*., 1973). Though RNA and DNA are the most obvious acidic biomolecules, phosphoproteins have also been shown to stain blue (Campbell *et al.*, 1983). Since the AAD is abundant in aspartic and glutamic acid residues, AAD might also stain blue.

Stains-All does not stain as intensely as Coomassie blue; thus, more protein was required in each lane on a gel. In early trials, this necessitated the use of a larger 16 cm x 16 cm Hoefer gel to assure enough protein was being loaded. Eventually, the staining protocol was optimized for BioRad minigels.

Before staining the gel must be fixed and have all detergent removed. If all of the SDS is not removed, the background will be a muddy yellow colour. Both requirements can be accomplished by shaking the gel in liberal amounts of 25% isopropanol at 50°C for 30 minutes. Once the gel is placed in the stain solution [0.01% Stains-all (from a 0.25% stock in formamide), 8% formamide, 25% isopropanol in 30 mM Tris-Cl, pH 8.8], staining can take up to 36 hours and must be done in the absence of light. Once exposed to light, the bands are prone to fading, often within minutes. The gel can be destained in distilled water that has been made alkaline by a few drops of 5 M NaOH. Poor contrast makes colour photography difficult while drying renders all of the proteins red.

2.4.1.6. Gel Filtration Chromatography

A Pharmacia FPLC medium pressure chromatography system was available to aid the resolution of product after GST-AAD thrombolysis. A 30 cm x 1.5 cm Superose-12 gel filtration column with a maximum loading capacity of 10 mg in 0.2 mL was used with Tris-

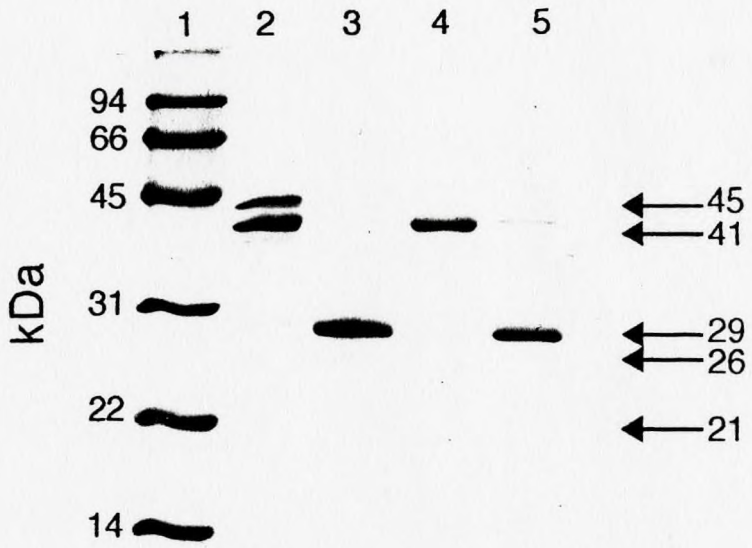


Figure 2.3: Comparison of Overexpression Products Derived From Plasmids pGEX-CT and pGEX-CT/TGA. Overexpressed proteins from bacteria harbouring the original GST-AAD fusion construct (pGEX-CT) and the GST-AAD fusion with the altered termination codon (pGEX-CT/TGA) were purified via affinity chromatography and subjected to 12% SDS-PAGE with Coomassie blue staining. *Lane 1:* marker proteins. *Lane 2-3:* the new construct before and after thrombin treatment with 1 NIH units/mg fusion protein. *Lanes 4-5:* the original construct before and after thrombin treatment. Arrows denote major products on the gel mentioned in the text.

Table 2.2: Analysis of GST-AAD fusion and thrombolysed products

Polypeptide	Thrombin Products	Trypsin Products	Stains-All Colour	Rx'n with anti-GST	Rx'n with anti-AAD
45 kDa	N/A	N/A	blue	yes	yes
41 kDa	N/A	N/A	blue	yes	yes
29 kDa	yes	yes	blue	no	yes
26 kDa	yes	no	red	yes	no
21 kDa	yes	yes	blue	no	yes

Glycine-dithiothreitol mobile phase. BioRad gel filtration standards were used for calibration.

2.4.2. Results

The yields of GST-AAD runs were much higher than Protein A-AAD ranging from 6-10 mg/L culture. SDS polyacrylamide gel electrophoresis showed two prominent bands of 42 kDa and 39 kDa in a with a minimal degradative background (Figure 2.3). The latter band was slightly more prominent. Both molecular weights were reasonable, if not somewhat high, for a Vmw65 fusion peptide as the expected molecular weight was 35382 Da. The corresponding Western blot, probed with either the GST specific GST-αPEA3 antisera or the AAD specific Protein A-AAD antisera hybridized with both prominent bands. Overexpression of GST alone via induction of bacteria habouring the parent vector resulted in one prominent band which ran at 28 kDa on the gel.

For a 30 minute digestion, a 1:5000 ratio of thrombin-to-fusion protein or 0.9 NIH units/mg fusion protein was used. Three bands were produced of 31 kDa, 28 kDa and 22 kDa. Trypsin digestion of the GST-AAD did not result in complete fusion hydrolysis reminicent of Protein A; rather, similar products remained. By consulting Figure 2.1, it can be seen that (Gly-Ser) would be appended to the N-terminus of AAD after thrombin digestion while like the Protein A-AAD fusion, AAD would be N-terminally truncated 7-8 residues. Prior to staining in Coomassie blue, the gel was stained with Stains-All to make a more critical assessment of the bands. It was found that the 31 kDa band and the 22 kDa band stained blue while the 28 kDa band stained red.Western analysis with the two types of antisera was performed a separate sample of cleaved fusion protein. The results of the

Table 2.2: Analysis of GST-AAD fusion and thrombolysed products

Polypeptide	Thrombin Products	Trypsin Products	Stains-All Colour	Rx'n with anti-GST	Rx'n with anti-AAD
45 kDa	N/A	N/A	blue	yes	yes
41 kDa	N/A	N/A	blue	yes	yes
29 kDa	yes	yes	blue	no	yes
26 kDa	yes	no	red	yes	no
21 kDa	yes	yes	blue	no	yes

Stains-All and Western analysis is presented in Table 2.2. FPLC gel filtration chromatography using a Superose-12 column (Pharmacia) failed to resolve the 31 kDa and 22 kDa proteins (data not shown).

Anticipating the large scale fermentation and purifications that will be required in the future, the GST overexpression system was superior to the Protein A overexpression system. Yields were in the range of 8 mg/L culture with little degradation and thrombolysis of the fusion was highly specific and complete.

In light of the results summarized in Table 2.2, it appeared that two fusions were being synthesized since both polypeptides reacted with each polyclonal antibody. Proteolysis was highly diagnostic as it showed that these proteinsdiffered in the carboxy-terminal Vmw65 portion and not in the amino-terminal GST portion since a unique molecular weight 28 kDa GST fragment was produced.

Though the 28 kDa band agreed with the predicted molecular weight of GST, the other 31 kDa and 22 kDa bands did not agree with the expected 9.3 kDa AAD. Vmw65 does possess an element causing aberrant electrophoretic mobility on an SDS polyacrylamide gels since its sequence molecular weight is 54 kDa (Dalrymple *et al.* 1985) and not 65 kDa. If the AAD was responsible, the effect would likely be exaggerated if this domain was electrophoresed alone. Other polypeptides that are rich in acidic residues such as pig heart calpastatin (Takano et al., 1988), nuclear histone-binding protein N1/N2 (Kleinschmidt et al., 1986) and the anti-oncogene p53 (Zakut-Houri et al., 1983) have sequence molecular weights below the predicted electrophoretic molecular weight.

A plausible explanation for the existence of two AAD polypeptides upon liberation by thrombin could be attributed to suppression of the amber (UAG) codon that terminates the Vmw65 gene. An inspection of the genotype of strain HB101 that harboured the overexpression plasmid revealed the presence of a *supE*44 suppressor. *SupE*44 is common in many strains of *E. coli* because it is believed to have arisen in the 1940's as a X-ray induced mutation in the glutamine tRNA gene, *glnV*, in *E. coli* K-12 strains (Lederberg and Tatum, 1946; reviewed by Bachmann, 1972; Eggertsson and Söll, 1988). This mutation can be traced in the *E. coli* K-12 pedigree to only three generations from the original K-12 strain isolated from a diptheria patient in 1922. UAG is used the least of the three natural stop codons in *E. coli* (Kohli and Grosjean, 1981) and has an intermediate read-through level much less than 1%. With respect to GST-AAD, the apparent read-through level was significantly higher that that by at least one order of magnitude by visual inspection of a gel.

Several hundred base pairs of Vmw65 3' non-coding sequence do exist in the pGEX-CT construct. Translation of this sequence beyond UAG to the next stop codon (ochre, UAA) would add 23 residues or an additional 2.3 kDa to the 9.3 kDa AAD (consult Figure 2.1). Initially, it would seem questionable to assume that a small increase in molecular weight could be responsible for a 9 kDa increase in mobility between the presumed 22 kDa AAD and its 31 kDa extended product. Indeed, an inspection of the extra 23 residues revealed a net charge of zero alluding to an increase that should have been more direct rather than so artificially high. However, proline residues have also been implicated with aberrant electrophoretic mobility. The acidic rich p53 anti-oncogene is also rich in proline residues while the antigen p28 of Human T cell Leukemia Virus (Iino et al., 1986) likely has its aberrant mobility entirely due to a proponderance of proline residues. With

respect to the intact fusion protein, the effect of AAD could have been attenuated due to the size (28 kDa) and charge of the slightly basic (pI = 8.53) GST carrier. Again, a similar argument can be made for the effect of AAD on the N-terminal 400 residues of Vmw65, though the relative effect is greater.

2.4.3. Conclusions

The suppression problem seemed to be the only barrier to this very efficient overexpression scheme. Unfortunately, gel filtration could not resolve the two peptides. Anion exchange would seem to be an unlikely candidate for resolution of the contaminating peptide since the pI's of the two AAD's would be very similar considering the background of the many acidic residues. The test of the suppression hypothesis would be to transform the plasmid into a non-suppressing strain (no *supE*, *supP*, *supD*, *supU*, *supF*, *supZ* type suppressors) or to alter the termination codon to ochre (UAA) or opal (UGA).

2.5. Overexpression of an Improved GST-AAD Fusion

To test the translation read-through hypothesis, the amber (UAG) stop codon was changed to an opal (UGA) stop codon . The parent vector contains UGA stop codons in all reading frames immediately 3' to the EcoRI site. This allows facile insertion of gene fragments that may or may not possess termination codons. The AAD fragment could be truncated and religated into the parent vector with some C-terminal residue loss. However, it remained important to preserve the integrity of the domain as best possible at both termini.

2.5.1. Materials and Methods

2.5.1.1. Construction

The indicated method of changing the stop codon while preserving the C-terminus was cassette mutagenesis. Initially, a *Bam* HI/*Eco* RI (all restriction enzymes were Boehringer-Mannheim) fragment encompassing the AAD gene was excised from pGEX-AAD and purified via phenol/chloroform extraction from low melting point agarose. This fragment was then cut with *Rsa* I at a position in the gene near tyr-488 (two residues from the C-terminus) to produce approximately a 5' 270 bp and a 3' 230 bp subfragment. The 270 bp subfragment was purified by similar means. The parent plasmid was linearized for AAD reinsertion via *Bam* HI/*Eco* RI digestion and purified by glass-milk under high ionic strength conditions (GeneClean II, Bio 101 Laboratories). A 23 base and a 27 base oligonucleotide (5'-ACGGTGGGTGATAACTGCAGCTG-3'/3'-TGCCACCCACTATTGACGTCGACTTAA-5') was annealed to produce a cassette with a 5' blunt end and a 3' Eco RI overhang. (all oligonucleotides were synthesized at the Central Facility of the Institute for Molecular Biology and Biotechnology at McMaster University). Ligation was accomplished by incubating the *Bam* HI/*Eco* RI cut vector, the *Bam* HI/*Rsa* I cut 270 bp fragment and the *Rsa* I/*Eco* RI cassette with 5 units of T4 ligase (Pharmacia), 1 mM ATP in 1x One-Phor-All buffer (Pharmacia) for 16 hours at 16°C. The bacterial strain used for transformations was *E. coli* DH10β [DH5α: F-, *rec*ABC, *sup*E44]. Cells were made competent by a standard RbCl protocol (Sambrook *et al.*, 1989). Introduction of a second *Pst* I site was basis for the primary screening of transformants. Secondary screening of potential positives was accomplished by *Nar* I/*Pvu* II digestion. To reflect the substitution of the opal termination codon, this construct was named pGEX-CT/TGA (refer to Figure 2.1 for comparison with pGEX-CT).

2.5.1.2. Nucleic Acid Sequence Verification

Plasmid DNA was denatured and prepared for sequencing by the method of Zhang et al. (1988). Sequencing was performed with a Sequenase kit (United States Biochemical) according to manufacturer's instructions with the non-coding strand of the cassette used in the cloning procedure and a synthetic primer 5'-AGCAAGTATTAGCA-3' located approximately 90 nt from the 3' terminus of the GST gene. A 90 nt distance was chosen to ensure legible sequence determination at the GST-AAD junction and to avoid a hairpin region near the 3' terminus of the GST gene. Annealing of the cassette primer proved its own existence in the plasmid. 45 nt of the 3' end of GST and nearly all of the AAD sequence was verified except for regions of compression presumably due to the GC rich nature of this Herpesvirus gene. Attempts to alleviate the compression by substituting dITP for dGTP in the sequencing reaction and running the sequencing gel at 55°C was not successful.

2.5.1.3. Determination of Optimum Overexpression Period

Previously, growth after induction at mid-log phase was limited to 30-45 minutes. In order to ascertain the optimum period the following procedure was used: A 100 mL culture (2-YT medium supplemented with 50 µg/mL ampicillin) was grown at 37°C to an OD_{600} of 0.5 and induced with 0.1 mM IPTG. A 1 mL sample was removed from the cuture at various intervals up to 3 hours, centrifuged and the bacterial pellet dissolved in 400 µL of 1x SDS loading buffer. Samples were boiled for 5 minutes and a quantity (ca. 10 µL) normalized to 0 minutes post-induction was loaded onto a 15 % SDS-polyacrylamide gel while still warm as the sample was quite viscous at room temperature.

2.5.1.4. Large Scale Overexpression of the Revised GST-AAD fusion

Many different quantities of bacteria were grown harbouring this plasmid from 1 L to 16 L per run. Though the volumes changed, the general parameters of the procedure did not change appreciably from those previously described for the pGEX-CT based fusion protein. Typically, inoculation of an overnight culture to 2-YT medium supplemented with 50 μg/mL ampicillin was adjusted such that induction would result approximately 2 hours later. The typical induction period with 0.1-0.2 mM IPTG varied between 90-150 minutes and was dependent mainly on equipment considerations during post-processing. The majority of overexpression runs were done by setting up multiple cultures in standard Pyrex culture flasks. To ensure even processing of up to six flasks, inoculations were staggered over a two hour period. Harvesting the bacteria proved to be the greatest limation to scale-up since the largest rotor (Beckmann GS3 rotor) could only accommodate 2.5-3.0 L of broth per centrifugation.

A New Brunswick Scientific 25 liter fermenter was refurbished and became the optimum method of growing large quantities of bacteria with high reproducibility. The unit contains two independent 12.5 fermenters in one housing. Temperature maintenance was accomplished by a closed water circulation system that could be heated. Intermittent cooling of the system providing temperature stabilization was accomplished by automatic opening the water circulation system to all cold tap water to flow through the stirring assembly. Air was provided by the in-house system with prefiltering through glass wool. Various ports included a septum to allow aseptic sampling, a thermometer housing, a thermocouple housing and a means of removing quantities of culture media while the unit was still operating. For 2-YT broth a stirring rate of 200 rpm and a air flow of 8 L/min was optimal. A stock of 5x 2-YT concentrate was autoclaved and mixed with glass distilled water *in situ*.

Though the broth was not sterile, it must be remembered that the total culture time from inoculation to harvesting was only five hours.

2.5.1.5. Processing and Purification

The cold water system also served as a means of rapidly cooling the broth as the induction neared completion. If flasks were used, they were transferred to ice to await processing. Bacterial pellets were resuspended in 75 mL PBS, 0.02% sodium azide, 1 mM PMSF, 200 KIU/mL trasylol (Miles Laboratories). Sonication, debris removal and storage was similar to previously discussed methods.

The availability of a BioRad EconoSystem for chromatography greatly enhanced processing as this unit contains an integrated pump, UV detector, fraction collector and chart recorder. The system worked well though it was prone to mechanical problems. Previously, fractions would be assayed for protein by Bradford assay or absorbance at 280 nm.

If the crude extract corresponding to 12 L of culture was to be processed, this sample was divided into two portions to prevent potential overloading of a 25 mL (5 cm x 2.5 cm (ID)) glutathione-Sepharose column. Elution and quantitation were via means previously described.

To the eluted fusion protein, 1 NIH unit/mg fusion was added directly and incubated for 30 minutes at 30°C. Thrombin is slightly inhibited by glutathione but no problems were observed. Some precipitation was noticed as fine wisps throughout the mixture which increased with time and was aggrevated by increased temperature. Some

precipitation was found in tubes which had remained at 4°C for several days. Addition of Triton X-100 or Tween-20 would not cause resolubilization. Unfortunately, salt could not be added since the increased ionic strength would interfere with planned anion exchange chromatography.

Resolution of the cleaved products could be accomplished by various means. One obvious method would be to dialyse the free glutathione that was still present in the cleaved mixture and run the mixture back over the affinity column. The method was tried but it was found that there was residual unbound GST in the flow-through. In addition, any degradative products and fragments from thrombolysis would still exist in the mixture. The pI of GST is 8.53 while the pI of AAD is 3.38. With such a broad difference, the obvious choice for separation was anion exchange chromatography. This method was insenstive to glutathione; no dialysis was required. Protein was loaded onto a 10 cm x 2.5 cm (ID) column of DEAE-Sephacel or DEAE-Sepharose (both by Pharmacia). All of the protein remained bound in the running buffer (50 mM Tris-Cl, pH 8.0). A linear gradient was programmed on the EconoSystem of 0 - 500 mM NaCl over 500 mL. It was important to set the UV absorbance maximum to low setting (ca. 0.1-0.2) since AAD only possess two tyrosine residues as absorbing species at 280 nm. Fractions containing AAD eluted at 350 mM NaCl. Quantitation of the NaCl content was accomplished via conductivity measurements. These fractions were pooled and concentrated with an Amicon ultrafiltration cell containing YM-3 membranes (cutoff - 3500 Da). For most runs, AAD was desalted via extensive dialysis in Spectrapor 4 membranes (cutoff - 3500 Da) over days to 1 mM ammonium bicarbonate, pH 8.0, flash-frozen under liquid nitrogen and lyophilised. Alternatively, AAD was desalted on a 40 cm x 1.6 cm (ID) Sephadex G-25SF column and

was stored frozen with 0.1% sodium azide as a preservative. By spectrophotometric means, about 15 mg of AAD was obtained from 12.5 L of culture.

Though the intact GST-AAD fusion was quite pure after affinity chromatography, possession of a highly purified protein was sought for future biophysical purposes such as crystallization. Homogenous GST-AAD was obtained by anion exchange chromatography (9 x 2.6 cm ID DEAE-sephacel) using a 200 mM to 550 mM NaCl gradient (in 10 mM Tris-Cl, pH 8.0) over 300 mL. GST-AAD eluted as a broad, yet distinct peak from other proteins, beginning at ca. 320 mM NaCl and extending to the end of the gradient. Fractions were pooled, concentrated, dialysed and lyophilised.

2.5.1.6. Determination of Purity

Coomassie Blue and silver staining (BioRad) of 15% SDS-PAGE was initially used to determine the extent of AAD purity.

In addition to SDS-PAGE, gel filtration was performed on a Beckman System Gold HPLC apparatus with a 7.5 x 300 mm Pharmacia TSK-G4000SW column with an accompanying Bio-Rad 7.5 x 100 mm TSK guard column using a mobile phase of 20 mM Tris-Cl, pH 8.0, 0.2 M NaCl and a flow rate of 1.5 ml/min.

Using the same elution system, the gel filtration column was calibrated according to BioRad gel filtration standards ranging from 670000 to 1350 Da. The choice of buffer system is important since ionic effects should be minimized. For this study, an aqueous based mobile phase was chosen containing 200 mM NaCl in 20 mM Tris-Cl, pH 8.0. In

lieu of retaining a high ionic strength, other systems have used sodium dodecyl sulfate in the mobile phase. For small peptides, a 36% acetonitrile / 0.1% trifluoroacetic acid mixture has been beneficial (Swergold and Rubin, 1983).

2.5.1.7. AAD Quantitation

Most commercial protein assays were deemed ineffective due to the unusual amino acid composition of AAD; thus, absorbance at 280 nm was used exploiting the two tyrosine residues that were available. Using an extinction coefficient of 1280 $M^{-1}cm^{-1}$ at 280 nm for the peptide gly-L-tyr-gly and the recommendations of Gill and Von Hippel (1989), protein concentrations were determined by mixing 200 µL of AAD with 800 µL of 7.5 M guanidine hydrochloride to yield a final Gu•HCl concentration of 6.0 M.

2.5.1.8. Amino-acid Composition

Approximately 50 nmol (5 ng) of purified AAD was hydrolysed at the HSC/Pharmacia Biotechnology Service Centre (Toronto, Ontario), PITC derivatized and chromatographed on a Picotag (Waters) column to determine amino acid composition against 250 pmol amino acid standards (BioRad). As a extra from the peak integrations, the concentration of AAD could be calculated, knowing the absolute quantities of amino acids exiting the column and the amount originally loaded onto the column.

2.5.2. Results

A sample of the new GST-AAD fusion containing an opal-substituted termination codon was overexpressed, affinity purified and thrombolysed from a 100 mL pilot culture.

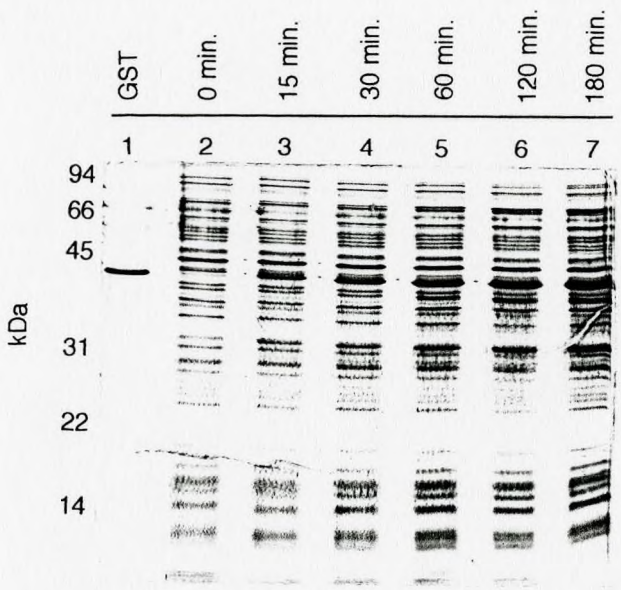


Figure 2.4: Overexpression Time Course of Plasmid pGEX-CT/TGA (revised GST-AAD fusion. Bacteria harbouring this plasmid were grown to mid-log phase and induced with 0.2 mM IPTG. At various times post-induction, cells in 1 mL of culture were dissolved in SDS loading buffer. The protein concentration was normalized to 0 minutes post-induction and a small fraction was subjected to 15% SDS-PAGE with Coomassie blue staining. *Lane 1:* purified GST-AAD. Lane 2-7: 0, 15, 30, 60, 120, 180 minutes post-induction.

From a comparision of the expressed and thrombolysed products between the old and new GST-AAD fusions, the suspected higher molecular weight contaminant was now missing on a SDS-polyacrylamide gel (Figure 2.3). From this initial observation, it appeared that translational read-through was likely the cause of the contaminant.

In the absence of the contaminant fusion protein, the overexperession protocol could now be optimized, starting with the overexpression period. Even in the bacterial protein background, the overexpressed GST-AAD product was very evident at 15 minutes of post-induction and continued to increase up to 120 minutes post-induction (Figure 2.4). No major degradative products were observed up to 180 minutes post-induction on the stained gel or a Western blot (not shown) of a similar gel probed with GST-AAD antisera. Generally, 120 minutes was chosen as a convenient time-point for harvesting.

In an overexpression scheme, solubility of the fusion protein within the bacterium is paramount. It was suggested that growth at a lower temperature might provide enhanced solubility of the fusion protein thus improving the yield (Dr. B. Weygood, personal communication). A similar procedure was performed at room temperature for a period of 24 hours post-induction. Up to 8 hours post-induction, the GST-AAD was stable though after 24 hours, it could not been detected visually on a Coomassie blue stained gel (not shown).

Consulting Figure 2.5, 1 NIH unit/mg fusion protein concentration of thrombin was deemed sufficient for liberating AAD from the GST-AAD fusion protein. Though higher concentrations (using a similar incubation period) of thrombin provided a more complete digestion, this enzyme was costly and thus required usage of a minimal amount.

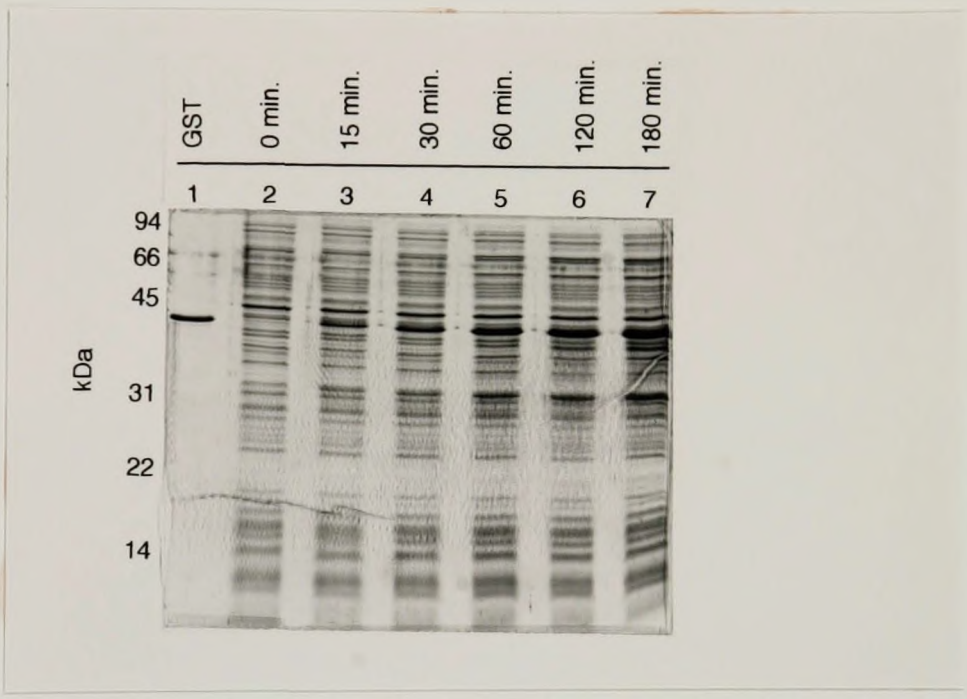


Figure 2.4: Overexpression Time Course of Plasmid pGEX-CT/TGA (revised GST-AAD fusion. Bacteria harbouring this plasmid were grown to mid-log phase and induced with 0.2 mM IPTG. At various times post-induction, cells in 1 mL of culture were dissolved in SDS loading buffer. The protein concentration was normalized to 0 minutes post-induction and a small fraction was subjected to 15% SDS-PAGE with Coomassie blue staining. *Lane 1:* purified GST-AAD. Lane 2-7: 0, 15, 30, 60, 120, 180 minutes post-induction.

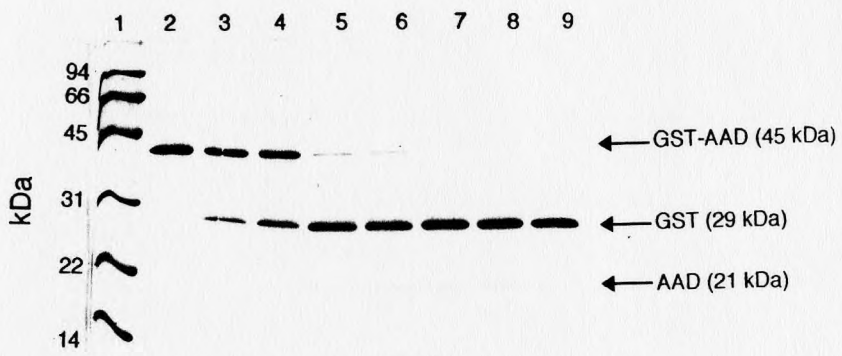


Figure 2.5: Determination of Optimum Conditions for Thrombin Cleavage. 5 μg of purified GST-AAD was digested with varying amounts of thrombin for 30 minutes at 30°C, subjected to 15% SDS-PAGE and stained with Coomassie Blue. *Lane 1*: marker proteins. *Lanes 2-9:* 0, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0 and 10 NIH units/mg fusion protein.

Separation of AAD from uncleaved GST-AAD, GST and other proteins was accomplished with a DEAE-based anion exchange column using a linear gradient of 0-500 mM NaCl. AAD eluted at 280 mM NaCl. Anion exchange chromatography could additionally be used if intact GST-AAD fusion was required. Figure 2.6 depicts the purification scheme from crude extract through anion exchange chromatography as seen by polyacrylamide gel electrophoresis.

From the Coomassie Blue stained gel in Figure 2.6, AAD was considerably pure. Silver staining of a similar gel was unsuccessful as AAD would not take up the stain but a very low background again described a high degree of purity (gel not shown).

HPLC based gel filtration provided another means of assessing AAD purity. From the elution profiles taken at 230 nm (peptide bonds) and at 280 nm (tyr, trp residues), the preparation of AAD used was >99% pure by peak integration (Figure 2.7). Calibrated against a set of gel filtration standards, a molecular weight of 54 kDa was predicted, far in excess of its true molecular weight of 9.3 kDa (Figure 2.8). This result could reflect the persistence of ionic interactions between AAD and the matrix or a highly extended structure that could not be adequately modelled by a set of predominantly globular standards. Refer to Section 3.7 for a further discussion of molecular volume based on a variant of this procedure.

The corellation between the predicted and experimental amino-acid compostitions was very good as shown in Table 2.3. Though estimation of aspartate and glutamate deviated slightly, both of these peaks overlapped each other early in the chromatographic

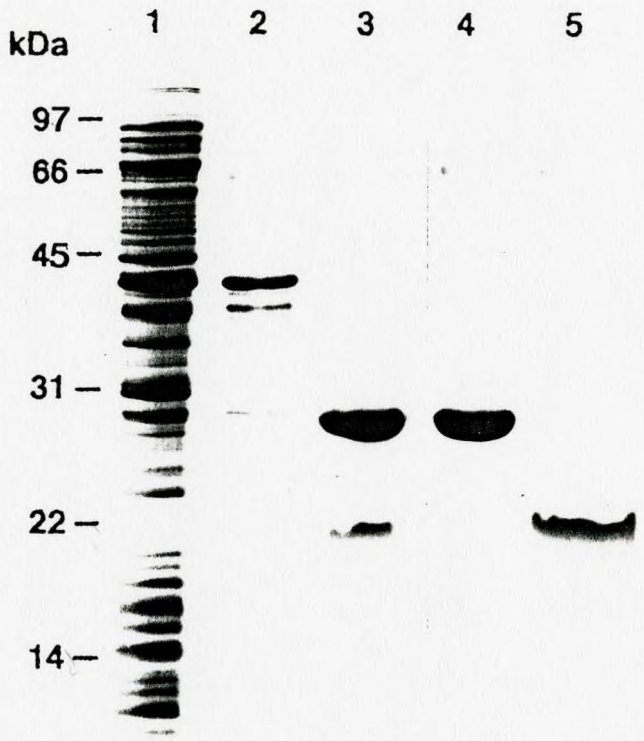


Figure 2.6: Purification of AAD. After affinity chromatography and carrier cleavage, AAD was purified to apparent homogenity by anion exchange chromatography. *Lane 1:* Crude extract. *Lane 2:* GST-AAD after affinity chromatography. *Lane 3:* GST, AAD after treatment with thrombin. *Lane 4:* GST fraction from ion exchange chromatography. *Lane 5:* AAD fraction from ion exchange chromatography. (15% SDS-polyacrylamide gel stained with Coomassie blue)

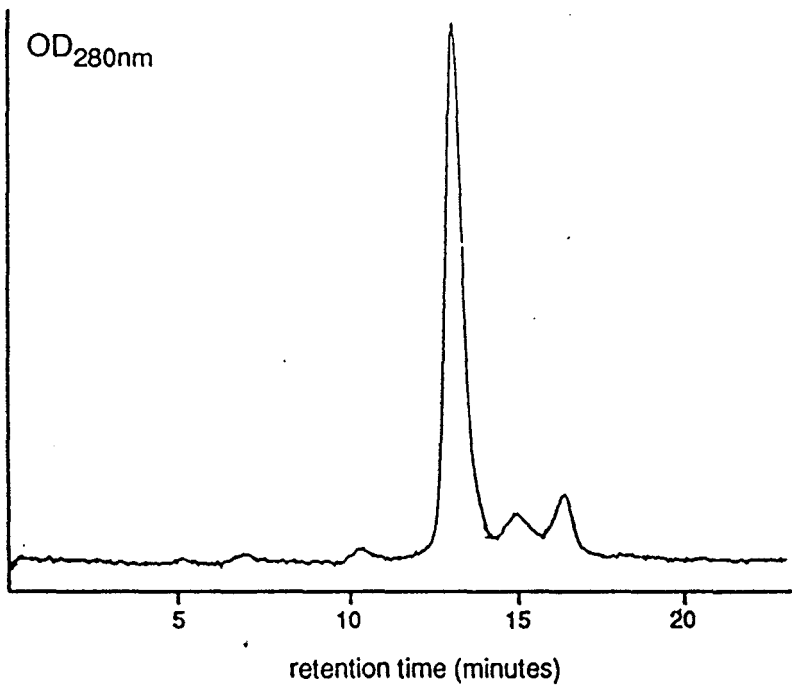


Figure 2.7: Determination of Purity by HPLC Gel Filtration. Chromatography was performed using a TSK-G4000SW column at moderate salt concentration (200 mM NaCl in 20 mM Tris-Cl, pH 8.0) to minimize ionic effects. Peak integration of the 280 nm elution profile showed that AAD was >99% pure.

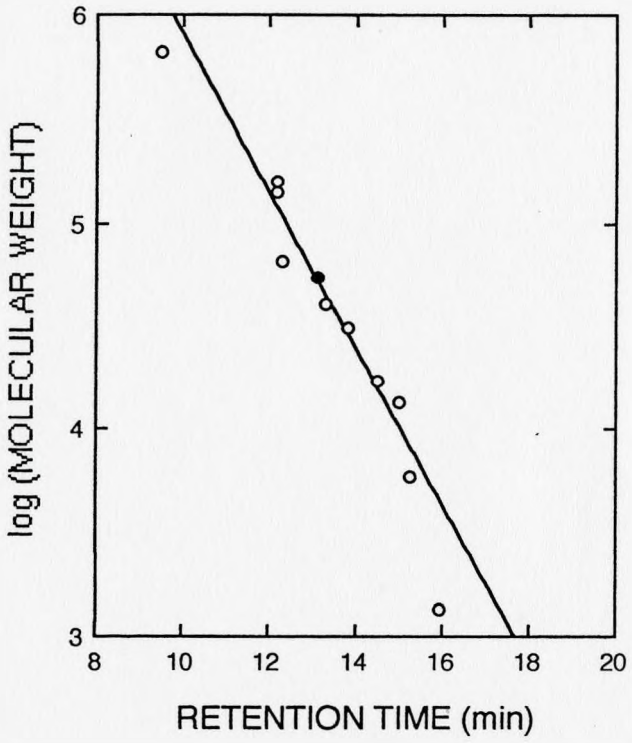


Figure 2.8. Gel Filtration of AAD. Chromography was performed on a TSK-G4000-SW column using a 20 mM Tris-Cl, pH 8.0, 0.2 mM NaCl mobile phase with the following standards (open circles): BSA Fraction V, 66 kDa (Pierce); alkaline phosphatase, 140 kDa (Pharmacia); RNase A, 13.7 kDa (Boeringer-Mannheim); DNase I, 31 kDa (Boeringer-Mannheim); insulin, 6 kDa; BioRad Gel Filtration standards - thyroglobulin, 670 kDa; γ globulin, 158 kDa; ovalbumin, 44 kDa; myoglobin, 17 kDa; vitamin B12, 1.35 kDa. Closed circle, AAD. From the graph, AAD is eluting as a protein with a molecular weight of 54 kDa.

Table 2.3: Amino Acid Hydrolysis Profile of AAD

Residue	Expected[a]	Observed[b]	
		Sample 1[c]	Sample 2
asx	16.	14.4	15.0
glx	6.	6.9	7.8
ser	5.	5.4	4.9
gly	12.	14.0	12.8
his	4.	3.7	3.3
arg	2.	2.5	2.3
thr	5.	4.6	4.7
ala	10.	10.0	9.9
pro	6.	6.5	6.4
tyr	2.	2.3	2.1
val	2.	2.2	2.2
met	4.	3.4	3.4
cys	0.	0.0	0.1
ile	1.	1.1	1.1
leu	9.	8.2	8.6
phe	5.	4.8	4.5
lys	0.	0.4	0.2
trp	0.	not done	not done

[a] Contributions according to the predicted 89 residue product, AAD, after thrombolysis

[b] AAD was analysed according to the procedure outlined in Section 2.5.8.

[c] 4.4 μg of AAD (0.5 nmol) was hydrolysed in this sample.

run likely making integration difficult. The amino acid composition was consistent over two independent preparations of AAD.

2.5.3. Discussion

Above a 10 L culture volume, features of bioengineering begin to surface. The fermentor solved the production problem that was encountered when culturing was done in multiple flasks spread over two laboratories. Harvesting still remained a problem since ten centrifugations in a 3 L rotor would be required to collect bacteria from 25 L of culture. A 4 L swinging bucket rotor (Sorvall HS-4) was used once to pellet bacteria but the pellet was spread over a large surface area in the flask and consequently, was not packed tightly. If a higher degree of scale-up was required, a continuous flow centrifuge or an Amicon mechanical cell harvester would be beneficial. Glutathione-Sepharose (Pharmacia) is expensive for large scale applications though recently (Spring, 1992) Sigma glutathione-agarose is now available in 50 mL quantities at 30% of the cost of the Pharmacia product. As ion-exchange resin is inexpensive, it does not pose a scale-up problem. Throughput may be increased at the chromatography stage by using higher flow rates, though this necessitates using better-quality columns that can withstand higher pressures. High pressures can be used, too, in wider bore columns.

AAD represents only 25% of the total fusion protein; thus, much of the protein generated during overexpression is wasteful. If the same yields were obtained with a small carrier peptide such as FLAG (consult Table 2.1), a 2-3 fold increase in amount of AAD would result. The FLAG system, however, incorporates a proprietary monoclonal affinity matrix which can cost a substantial amount when quantities in the order of 10's mL are

required. It would be preferable instead to switch to the poly(his)$_6$ system (Qiagen) since affinity matrix is an economical nickel chelating resin.

One possible refinement to the GST system would be to have a monoclonal antibody available to the carrier as other companies offer similar products. Such an oversight, though, is small for a overexpression system suited to many purposes.

Chapter 3

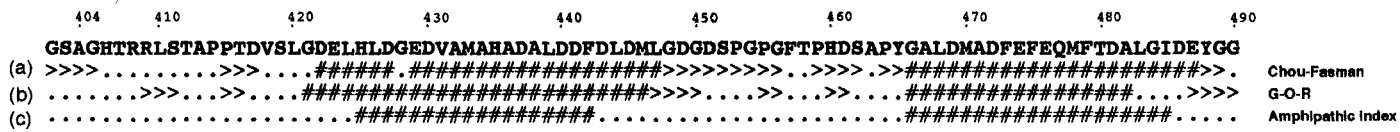
BIOPHYSICAL STUDIES ON THE VMW65 AAD

3.1. The Primary Structure of AAD

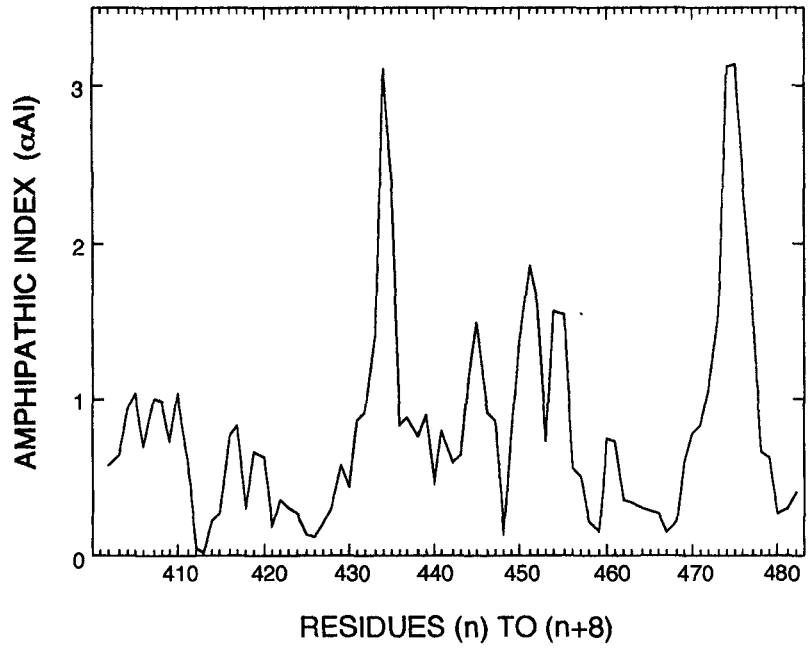
Even from a casual inspection of the AAD amino acid sequence in Figure 3.1, its special composition can be appreciated. The most immediately striking feature is the acidity; of the 89 residues listed, 24 bear a negative charge, provided mainly by aspartic acid. In such a acidic background, the two basic arginine residues near the N-terminus of the domain do very little to affect the extremely low pI of 3.3. In relation to function, some loss of acidity can be tolerated though it is more interesting to note the sensitivity of the hydrophobic residues to mutation, particularly, a mutation in phenylalanine-442 described by Cress and Triezenberg (1991b). Substitution of F442 with alanine and especially with proline severely compromised activity which could be partially restored with tyrosine. The function of this mutation in the context of the entire molecule *in vivo* remains to be observed. If a coarse sequence alignment of AAD with other transactivators, irrespective of

Figure 3.1: Secondary Structure Prediction. In the upper panel, the Chou-Fasman and Garnier-Osquthorpe-Robson methods were applied to the AAD sequence (labelled according to position in Vmw65) using programs in the University of Wisconsin GCG Software Package on the National Research Council CND/SND system (Ottawa, Ontario). (.) = random-coil or extended conformation. (#) = alpha-helical conformation. (>) = beta structures. Conformations were also assigned according to amphipathic index shown in the lower panel. Two prominent peaks were observed in the alpha-helix AI while determination of β-structure elements was more ambiguous.

A


404 410 420 430 440 450 460 470 480 490
GSAGHTRRLSTAPPTDVSLGDELHLDGEDVAMAHADALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYGG
(a) >>>>.....>>>....######.##################>>>>>>>>>..>>>>.>>####################>>.
(b) ....>>>...>>....#########################>>>>....>>...>>....#################....>>>>
(c) ....##################.....................##################.....
Chou-Fasman
G-O-R
Amphipathic index

B


AMPHIPATHIC INDEX (αAI)
0
1
2
3
410
420
430
440
450
460
470
480
RESIDUES (n) TO (n+8)

class is attempted, again it is the hydrophobics that contribute (Cress and Triezenberg, 1991b; addition homologies noted by J. Capone). Six proline residues are dispersed throughout the polypeptide with pro-pro occurring in one instance. Proline pairs are generally found in random-coil regions of protein, though they are also observed near the first turn of an α-helix (MacArthur and Thornton, 1991). Lysine, cysteine and tryptophan residues are conspicuously absent in AAD which is unfortunate since these amino acids contain functional groups suitable for covalent modification and chemical/instrumental analysis.

3.2. Prediction of Secondary Structure

The ultimate goal of protein chemists is the development of a method to determine higher elements of structure, particularly tertiary structure (Bowie et al., 1991) from the primary structure. Implicit to the algorithm is a understanding of protein folding which is far from complete though significant strides in protein engineering, like the development of a synthetic four-helix bundle (FELIX) have provided encouragement.

The Chou-Fasman (Chou and Fasman, 1979) and the GOR (Garnier et al., 1978) algorithms are widely used methods of secondary structure. The Chou-Fasman method uses a statistical approach based on X-ray structures coupled with empiral rules while the GOR method is purely statisical (reviewed in Fasman, 1990). Application of these methods to AAD revealed similar results shown in Figure 3.1 — the presence of two large α-helical regions (gly421-leu447 and gly466-ala482/glu486) flanked by regions of random coil and β-turns. Neither method predicted any β-sheet contribution.

Using an extensive computer-based matching algorithm on a massively parallel computer, Zhu et al. (1990) have found loose sequence similarities in many transactivators from both eukaryotic and prokaryotic origins. Based on structure prediction and statistical tendencies for an amino acid to reside in a particular structure, they proposed a consensus "acid-turn helix" motif.

3.3. The Role of Amphipathicity

A recurring theme in the transactivation model is the requirement for a random-coil to α-helix transition. Owing to the nature of the AAD sequence, any potential α-helix formed would likely be amphipathic. Visually, one can make a determination of the amphipathicity drawing a helical wheel where residues are applied to the spokes which are separated by 100° (5 helical-turns can be cast on a wheel before there is overlap) To probe this question further in a quantitative manner, programs were written to calculate the hydrophobic moment (Eisenberg et al., 1982, 1984) which has previously been effective at predicting the existence of transmembrane helicies (Lemire et al., 1989).

To obtain a hydrophobic moment, the helical wheel is transformed into a set of vectors that radiate from the centre of the helix. The magnitude of each vector can be related to the hydrophobicity of its corresponding amino acid. The magnitude of the resultant vector when all individual vectors are summed is the hydrophobic moment, P(ω) shown algebraically in Equation 3.1.

$$P(\omega) = \left[\sum_{k=0}^{l-1} h_k \cdot \cos(k\omega)\right]^2 + \left[\sum_{k=0}^{l-1} h_k \cdot \sin(k\omega)\right]^2 \qquad [3.1]$$

The value (ω) represents the period; for an α-helix, (ω) is approximately 100° or 3.6 residues/turn. The calculation proceeds for a window of *(l)* residues. Each residue at position (k) will have an associated hydrophobicity, (h_k). At least 36 scales of hydrophobic coefficients are available; the scale designated PRIFT was used on the recommendation of an extensive comparison performed by Cornette et al. (1987). The PRIFT scale was derived statistically from protein that are known to form amphipathic helicies. Other scales have used other criteria such as solubility, chromatographic mobility and a mathematical treatment of the side chains.

A preliminary investigation was performed to determine the optimum frequency to use by calculating a matrix of hydrophobic moments using angles 80-120° in 0.1° increments and over the entire AAD with a 9 residue window (2.5 turns). Results (not shown) were tabulated and displayed as a contour plot to aid interpretation. By scanning the database, Cornette et al. found that the dominant period peaked at 97.5° (3.7 turns) rather that the commonly accepted 100°. From their observations, they attributed the deviation to an openness in the C-terminus of a helix. AAD had peak period of 98.0°.

To determine the regions in AAD that had the highest hydrophobic moment, an an α-helix amphipathic index (αAI) or β-sheet amphipathic index (βAI) were calculated over a 9, 12, 15 and 18 residue window across the polypeptide using the functions described by Equation 3.2. To obtain the amphipathic index, the average hydrophobic moment over a realistic range of periods for a particular structure is calculated and then normalized against all the periods.

$$\alpha AI[P(\omega)] = \frac{\frac{1}{25^\circ}\int_{85^\circ}^{110^\circ} P(\omega)d\omega}{\frac{1}{180^\circ}\int_{0^\circ}^{180^\circ} P(\omega)d\omega} \qquad \beta AI[P(\omega)] = \frac{\frac{1}{20^\circ}\int_{160^\circ}^{180^\circ} P(\omega)d\omega}{\frac{1}{180^\circ}\int_{0^\circ}^{180^\circ} P(\omega)d\omega} \qquad [3.2]$$

From the results shown in Figure 3.1, two regions were markedly amphipathic with respect to an α-helix at a window of 9 residues. Using larger windows, the location of the amphipathic regions did not change. The peak magnitudes above the background were not as appreciable for the 12, 15 and 18 residue windows as they were for the 9 residue window (data not shown). This could suggest that the amphipathic regions do not extend for considerable distances in AAD. At any window size, it was difficult to select regions according to the β-sheet amphipathic index (data not shown); hence, a further study of this data was not performed.

Though data from computational methods including those described in this section must always be interpreted with extreme caution, the data presented nevertheless supports the existence of two amphipathic regions if this conformation exists *in vivo* and illustrates the ease at which this method can be implemented.

3.4. Circular Dichroism Spectroscopy

Commonly referred to as CD spectroscopy, this method is widely used as a simple, yet highly sensitive means of obtaining secondary structural information. Much of the information described in this section is summarized in Donaldson and Capone (1992).

3.4.1. Theory

In CD spectroscopy, left and right circularly polarized light is used instead of plane polarized light. By passing left/right circularly polarized light through an optically active sample, elliptically polarized light will exit. In the context of circularly polarized light, either the left or right circular component has been preferentially absorbed. The value, ellipticity, is commonly used to quantitate this phenomenon and is defined as the arc tangent of the ratio of the minor axis to the major axis of the ellipse (reviewed in Cantor and Schimmel, 1980; Yang et al., 1984). The CD spectropolarimeter functions by alternating left/right circularly polarized light and measuring the absorbance change. Absorbance change and ellipticity can be interconverted by a mathematical relationship.

For reasons of standardization, ellipticity can be converted to molar ellipticity. To further standardize against all proteins regardless of length, mean residue ellipticity is used and is defined as the molar residue ellipticity on a per residue basis.

Secondary structure from CD spectroscopy is typically measured in the range 190-250 nm though some specialized vacuum instruments can extend this range farther into the UV spectrum. Past 190 nm though, many technical problems occur due to water and oxygen absorption and even due to the cuvette material. In this wavelength range, absorptions in the polypeptide are due to backbone amide chromophores; specifically, these include $n \rightarrow \pi^*$ transitions at 210 nm and $\pi \rightarrow \pi^*$ transitions at 190 nm (Johnson, 1988). Secondary structure will affect the overall electronic nature of the molecule producing spectra that are visually quite different between α-helix, β-sheet and

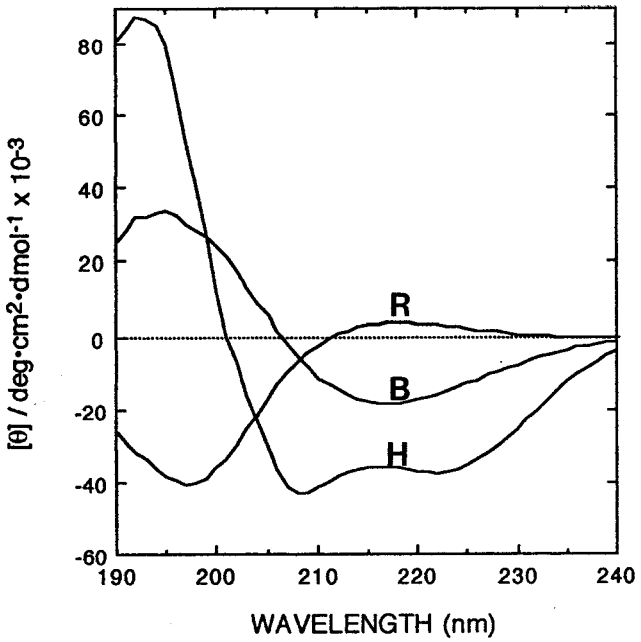


Figure 3.2: CD spectra of poly-L-lysine. Depending on physical conditions, poly-L-lysine can exist in three distinct conformations illustrated in the above spectra. R - random coil form. B - beta-sheet conformation. H - helical conformation. The α-helical conformation (at pH 11.1, 22°C) is the simplest conformation to describe and quantitate since is markedly different from the other two conformations, exhibiting distinct minima at 208 and 222 nm. The β-sheet conformation (at pH 11.1, polymer heated to 52°C and cooled to 22°C) has a broader minimum near 216 nm whereas the random coil conformation (at pH 5.7, 22°C) exhibits a slight maximum in this region. Spectra such as these have formed the basis for programs designed to deconvolve an unknown spectrum into the three secondary structural components. Interpretation must be made with caution since: (1) Chain length of a particular secondary structural motif can greatly affect the magnitude of its given spectrum and (2) There are deviations in the spectra obtained from other peptide homopolymers. *Plotted from data published by Yang et al. (1984).*

random coil. The term random coil is used to describe a structure that is neither α-helix nor β-sheet. It is difficult to predict what a β-turn spectrum would look like since it is impossible to make a solution of β-turns. Figure 3.2 shows the characteric spectra of poly-L-lysine in its various conformations for comparison. It is possible to assign a secondary structural composition to a given spectrum through a variety of means which will be discussed later in this section.

3.4.2. Materials and Methods

3.4.2.1. Instrumentation

Jacketed cells of 0.1 and 0.2 cm for secondary structure analysis and 2.0 cm for tertiary structure analysis were used, with typically 10 spectra being accumulated and averaged and base-line corrected on a Jasco J-600 spectropolarimeter interfaced to an IBM personal computer running Jasco software. Calibration of the machine was routinely performed with D(+)-10-camphorsulfonic acid at 290 nm. With the exception of temperature trials, all spectra were acquired at 25°C maintained by a Brinkmann-Lauda RC8 water bath. For conversion to mean residue ellipticity [θ], a mean residue weight of 104.8 g/mol for AAD was used.

The AAD used was derived from several overexpression runs. After chromatography, purity was established by SDS-PAGE. Protein was desalted and lyophilised prior to use. Reconstitution into the necessary buffers was follow by filtration through 0.2 μm sterile filter to ensure a transparent solution.

TFE is hygroscopic which could create misleading results. The TFE (Sigma) used for the CD study was taken from a new (sealed) bottle; hence, much water was not expected to be retained. In retrospect, it may have been preferable to store a small quantity over molecular sieves. When required, TFE was added to the protein solution and immediately scanned in a capped cell.

3.4.2.2. Determination of Protein Concentration

Ellipticity is dependent on protein concentration. The Bradford and Lowry assays are useful for routine biochemical work providing that the protein being quantitated is similar in composition to the protein used to standardize the assay. Frequently, this is not the case resulting in potentially large errors in excess of 50%. Though, CD spectroscopy is largely qualitative, the calculation of secondary structure contribution is one case where the ellipticity must be accurately known, often to within 3%. Owing to the error in the method of quantitation itself, a higher uncertainty in protein estimation only compounds the problem.

The most common method of protein quantitation used for spectroscopy is a spectroscopic method relying on the absorbance of aromatic residues in the near ultraviolet which was previously discussed in Section 2.5.1.7. For the CD investigations, this method was not used since the values obtained were 2.6-2.8x greater in two instances than the values obtained from peptide hydrolysis and integration on HPLC. An alternative method is the Kjeldahl method which determines total nitrogen. Unfortunately, even the micro-Kjeldahl protocol still requires milligram quantities of protein.

3.4.3. Results

3.4.3.1. A Random-Coil to Helix Conformational Change

The first priority in a CD study is to determine if there is a concentration or ionic strength effect since any multimerization or aggregation will result in a false interpretation of the spectra. Centered around a typical protein concentration of ca. 40 μg/mL, no change in the spectrum of AAD was observed over an eight-fold range (not shown). This observation was supported by equilibrium sedimentation data (Section 3.6) and gel filtration data (Section 2.5.6). Neither the position of the AAD peak in the gel filtration profile changed nor did any new peaks arise over a 1000-fold range (10 mg/mL to 10 μg/mL concentration, 10 μL injected into the column). Aggregation was also dismissed by equilibrium ultracentrifugation (see Section 3.6.). No change in the AAD was observed over a NaCl range of 0-150 mM. At high concentrations, chloride ion obscures the spectrum in the far ultraviolet. Though ionic strength was not explored further, perchlorate ion or iodide ion may have been alternatively used to circumvent these problems.

The spectrum of AAD at physiological conditions is shown by Figure 3.3(*a*). Consulting Figure 3.2 for reference, the spectrum of AAD is reminicent of a random coil. As AAD is involved in protein-protein interactions, it may not be experiencing an aqueous environment *in vivo*; thus, spectra were also collected in the presence of TFE to similate a more hydrophobic milieu. Results of Lehrman et al. (1990) are exemplary in justifying the use of TFE by showing a correlation of predicted secondary structure to known secondary structure using peptide fragments dissolved in TFE solutions. AAD adopted a more α-helical conformation with increasing concentrations of TFE (Figure 3.3(*b-f*)) and saturated at 50% TFE.

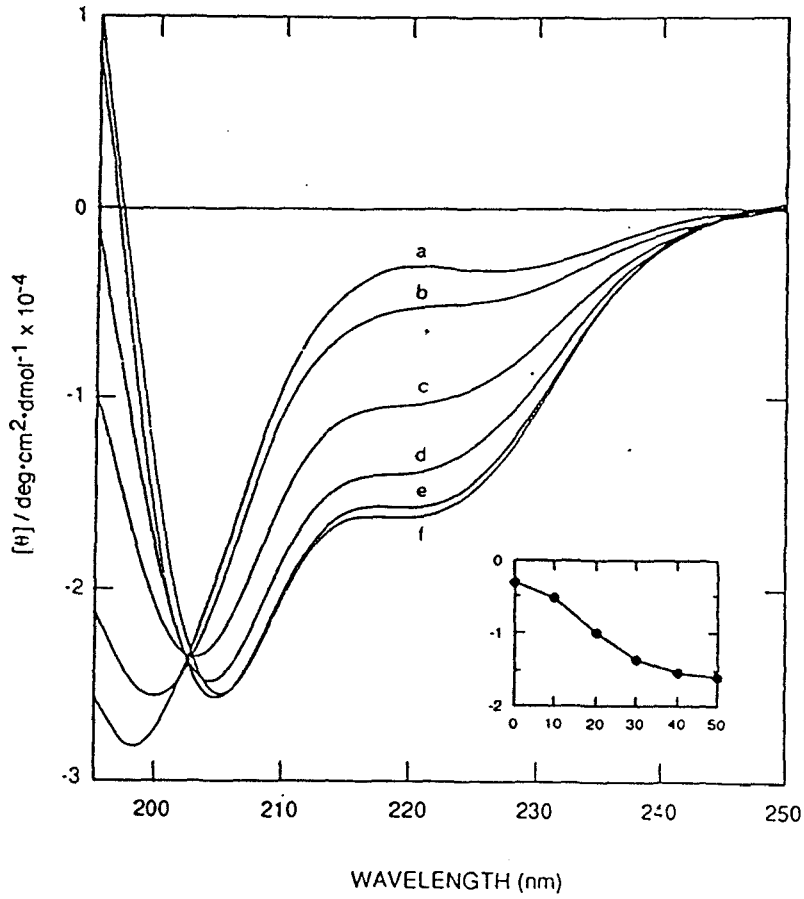


Figure 3.3: Far-UV CD Spectra of AAD. Spectra were acquired in 10 mM MOPS, pH 7.1, at 25°C. Protein concentration was 38 μg/mL and path length was 0.1 cm. Spectra *(a-f)* were obtained in the presence of 0, 10, 20, 30, 40 and 50% TFE, respectively. *Inset*, titration of $[\theta]_{222nm}$ (in deg•cm^2•dmol^{-1}) versus percent TFE (v/v). $[\theta]_{222nm}$ is indicative of α-helical structure.

At nearly one magnitude higher in AAD concentration, spectra in the near-UV range (300-250 nm) were acquired at physiological conditions and in the presence of 50% TFE (Figure 3.4). Residues that absorb in this range include phenylalanine (270-250 nm), tyrosine (290-270 nm), cysteine and tryptophan. In the near-UV, these fluorophores are sensitive to tertiary structure rather than secondary structure. Since AAD does not possess any cysteine or tryptophan residues, only the contributions of two tyrosine and five phenylalanine residues were observed. Though definite assignments of structure cannot be made from such spectra, the spectra did indicate that the aromatics were experiencing an asymmetric environment and that the corresponding structure did markedly change upon addition of TFE. It should not be concluded that the far-UV data contradicts the near-UV data as molten-globule forms of proteins may also exhibit near-UV CD bands (Goto et al., 1990).

3.4.3.2. Estimation of Structure Contributions from CD Spectra

To bring CD spectroscopy from a qualitative to a quantitative technique, many algorithms have been developed to estimate the individual contributions of each secondary structure class. The first program used in this study was PROSEC, a IBM-PC based program packaged with the Aviv spectropolarimeter. Since the data format differed between the Aviv and the Jasco machine used in this study, a short program was written and compiled in Microsoft Quickbasic for Macintosh to port the data to PROSEC. PROSEC is based on the work of Chang et al. (1978) who derived a set of "basis spectra" — four synthetic spectra representing random, β-sheet, β-turn and α-helical structures — from 18 spectra of "real" proteins with known three-dimensional structures from crystallographic analysis. Data is provided to the program in 1.0 nm increments from 240-190 nm. In operation, the program uses an iterative method to closely match as possible the input

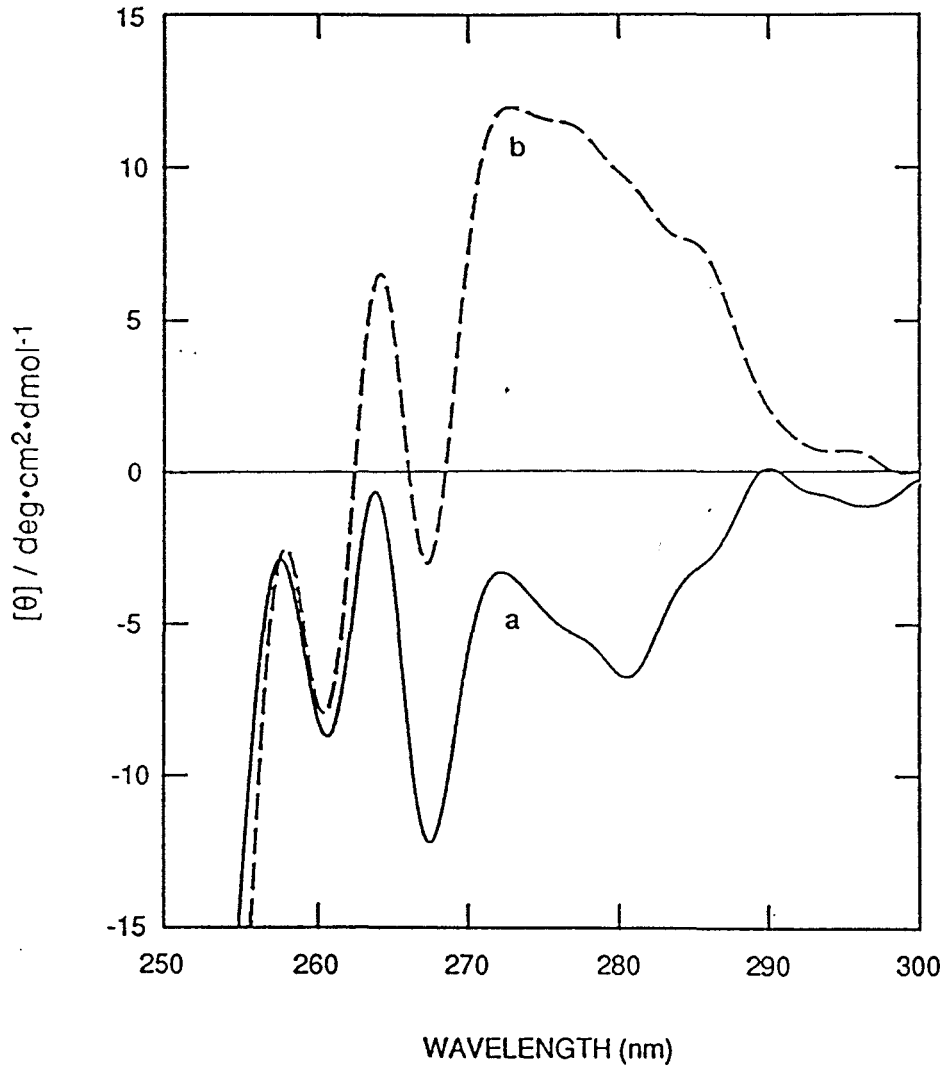


Figure 3.4: Near-UV CD Spectra of AAD. Spectra were acquired in 10 mM MOPS, pH 7.1, at 25°C. Protein concentration was 0.46 mg/mL and path length was 2.0 cm. Spectra (*a* and *b*) were acquired in the presence of 0 and 50% (v/v) TFE, respectively.

spectrum with fractional multiples of the basis spectra. Deviations are reported as an arbitrary error value. One interesting feature of this algorithm is that the fractional contributions of each basic spectrum can be added back together to compare how well the synthetic spectrum matches the input spectrum; this function is provided by the accompanying program, BLEND. For the spectra shown in Figure 3.3, error values ranged from 4600 to 900. When the fractional contributions of each spectrum in Figure 3.3 and others were BLENDed back to complete spectra and compared, a threshold error of ca. 400 was defined as the maximum allowable error for reliable estimation (data and BLENDed spectra not shown). As none of the AAD spectra fell within this range (4600 - 900), PROSEC was deemed unreliable for quantitation.

A second program residing on the mainframe computer, CONTIN, provided acceptable results. CONTIN is an extensive general purpose statistical program written in FORTRAN that was adapted by Provencher and Glockner (1981) to process CD spectra. To present the Jasco data in an acceptable form to the program, it was processed with the program `tocontin.exe`, written and compiled in Turbo C v5.0 for the IBM-PC (program listing in Appendix C). Visual inspection of the synthetic spectra correlated with the input spectra but it must be remembered that this program will fit nearly any input well. Structure contributions obtained from the 50% TFE spectrum were similar to those predicted by Chou-Fasman and GOR methods. Since the α-helical spectrum is the most different both in shape and magnitude from the other structure spectra, this fractional contribution can be relied on the most. Visually, the 0% TFE spectrum did not appear to contain any α-helix yet CONTIN predicted an 18% contribution (Table 3.1)

Table 3.1: Comparison of Secondary Structure Predictions from the AAD Sequence and CD Spectra

Method	alpha-helix	beta-strand	beta-turn	Residual
Chou-Fasman	53% [a]	0%	27%	20%
Garnier-Osquthorpe-Robson	48%	0%	19%	33%
CONTIN Prediction				
50% TFE [b]	53±1%	13±3% [c]		34±4%
0% TFE [d]	18±3%	19±5%		63±7%

[a] Percentages are expressed as summations of each structure contribution divided by 89 residues.

[b] Data for CONTIN calculation provided by Figure 3.3f from 240 to 195 nm.

[c] CONTIN does not make a distinction between β-strand and β-turn structures.

[d] Data for CONTIN calculation provided by Figure 3.3a from 240 to 195 nm.

3.4.3.3. Effect of Temperature on 2° Structure

Temperature is a simple and informative parameter to vary. At low temperatures, a variety of effects are observed from cold-denaturation to a transition to a more helical state. At high temperatures, denaturation can be acheived without the addition of chaotropic reagents. The effect of temperature on AAD secondary structure was monitored both in the presence and the absence of 50% TFE. As there can be considerable deviation from the temperature observed in the water bath to the temperature observed in the cell, a thermocouple was installed in the water line adjacent to the output port of the cell.

In the absence of TFE, the magnitude of trough near 195 nm varied inversely with temperature (Figure 3.5). At 5°C, a spectral bump indicative of a random-coil was observed near 220 nm. To varying extents, a random structrure predominated in all of the spectra. An isodichroic point at 208 nm suggests a two-state transition between structures. From a visual comparison of the reference spectra in Figure 3.6, the 80°C spectrum could be the product of a contribution of β-structure since the 195 nm trough increased and the 220 nm decreased in magnitude. To graphically illustrate the conformational change, the relative change in ellipticity (relative to 25°C) was plotted against temperature using data derived from the regions of highest change (Figure 3.7). For AAD in the absence of TFE, data points at 199 nm were plotted. For AAD in the presence of TFE, data points at 222 nm (indicative of the amount of α-helix present) were plotted. From the plot, no discrete conformational transition point could be detected; rather, a linear decrease in relative ellipticity at both wavelengths was evident.

In the presence of TFE (Figure 3.6), it was suprising to see that the helical structure could not be abolished even at 82.2°C. This data could suggest that TFE was promoting a

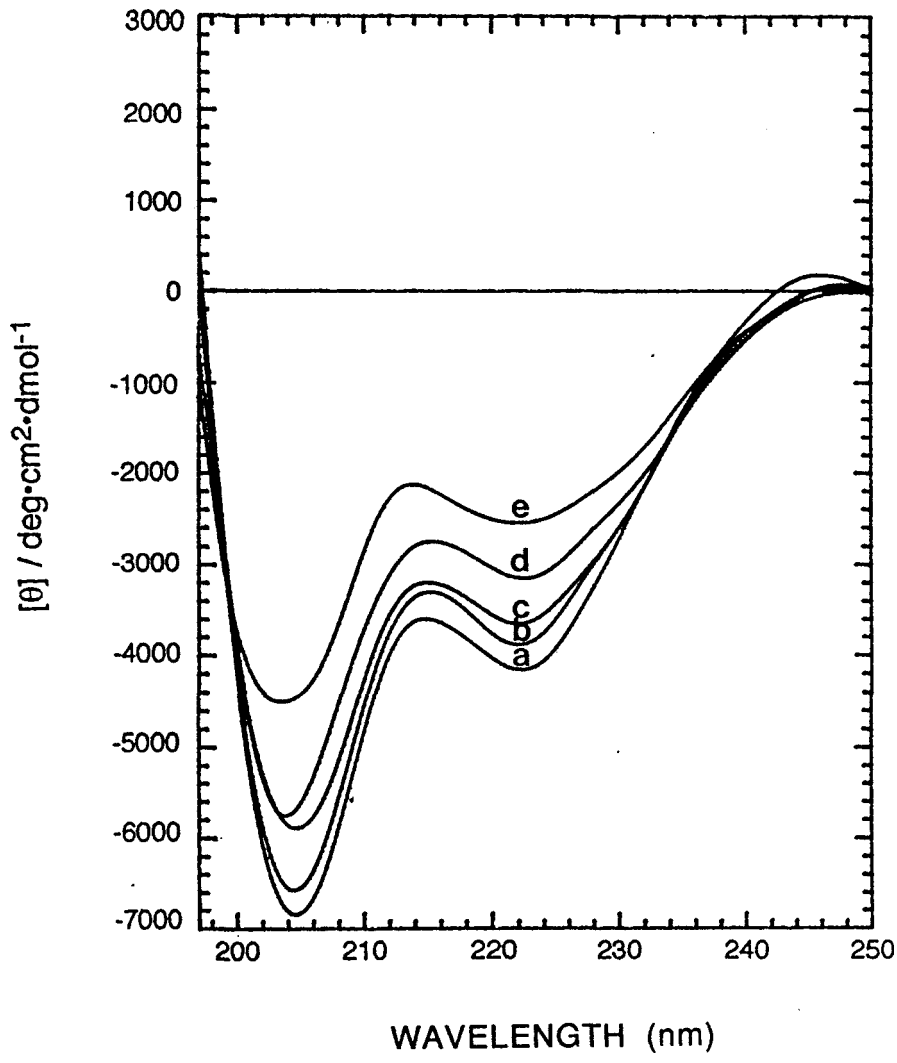


Figure 3.6: Effect of Temperature on AAD (+TFE). Spectra were acquired in 10 mM Tris-Cl, pH 8.0 + 50% (v/v) TFE. Path length was 0.1 cm. Spectra (*a* - *e*) were acquired at 25.0°C, 34.7°C, 41.8°C, 72.2°C and 82.2°C.

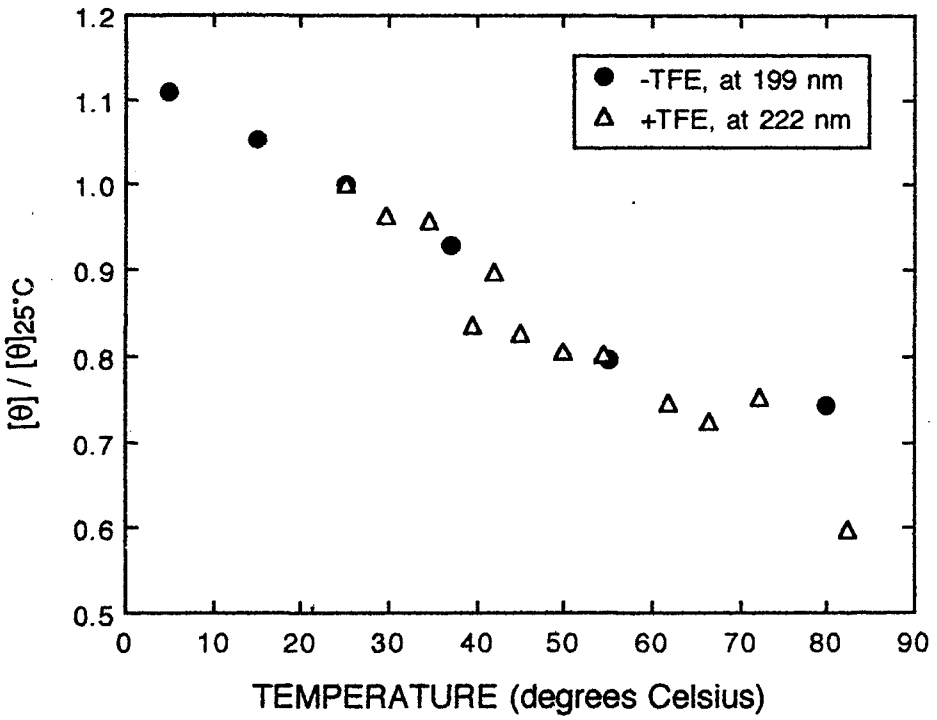


Figure 3.7: Plot of Temperature Effect on AAD. From the CD spectra depicted in Figure 3.5 and Figure 3.6 showing the effect of temperature on AAD at physiological pH, the relative change in ellipticity ($[\theta]_{25°C} = 1.0$) was plotted from points at 199 nm (in the absence of 50% TFE) or at 222 nm (in the presence of 50% TFE).

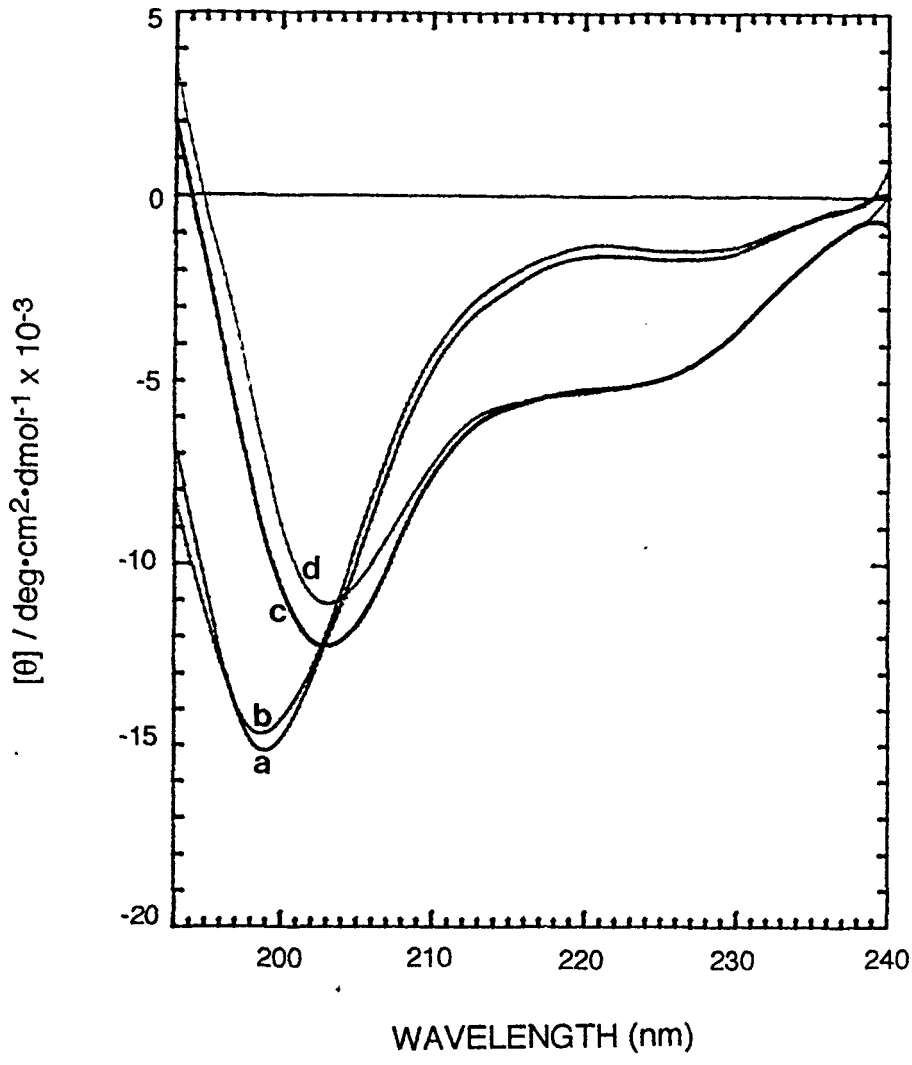


Figure 3.8: Effect of Divalent Cations on AAD. Initially, spectra were acquired using a solution of 5 mM sodium phosphate, pH 7.05, 30% TFE (v/v), 5 μM AAD (50 μg/mL) in the presence of (*a*) 0.25 mM Ca^{++} and (*b*) in the absence of Ca^{++}. To determine if divalent cations could stabilize a helical structure, a 30% TFE (v/v) 4x stock solution containing similar ratios of the above components was diluted to 7.5% TFE (v/v) in the (*c*) presence of Ca^{++} and (*d*) in the absence of Ca^{++}. This procedure was repeated substituting Mg^{++} and Zn^{++} with similar results (data not shown).

conformational change that is artifical in a sense, since one popular model of transactivation requires the conformational change to be quite labile to the enviroment. Perhaps a different effect would be observed at lower (non-saturating) TFE concentrations. However, even this observation would be difficult to correlate with a relevant biological activity.

3.4.3.4. Effect of Divalent Metals on 2° Structure

Incorporation of a transition metal for many proteins is essential means of maintaining a conformation to serve both modulatory and catalytic roles. However, the presence of salt (mono- or divalent cation) in the medium can also influence a protein's stablility by altering its solvent shell (Arakawa et al., 1990). If the salt causes preferential hydration—exclusion of salt from the solvent shell to create an increase in surface tension— the tendency to aggregate in the absence of denaturation will be greater. This is the basis of salting-out. Alternatively, if the cation can bind to the protein, an increase in solubility will be traded for a higher risk of denaturation. In the scope of this study, however, salt concentrations were not used in the molar range required to observe a solubility effect.

Though no motifs for metal binding have been identified from the sequence, the protein does contain many carbonyl groups that could interact with the coordinating shell of a divalent metal.

To test the ability of common divalent cations Ca^{++}, Mg^{++} and Zn^{++} to direct a conformational change and/or stabilize a helical conformation, the following experiments were performed at 25°C (shown in Figure 3.8): First, AAD conformation at neutral pH was simply assessed by adding a predetermined volume of a 1 M solution of the corresponding

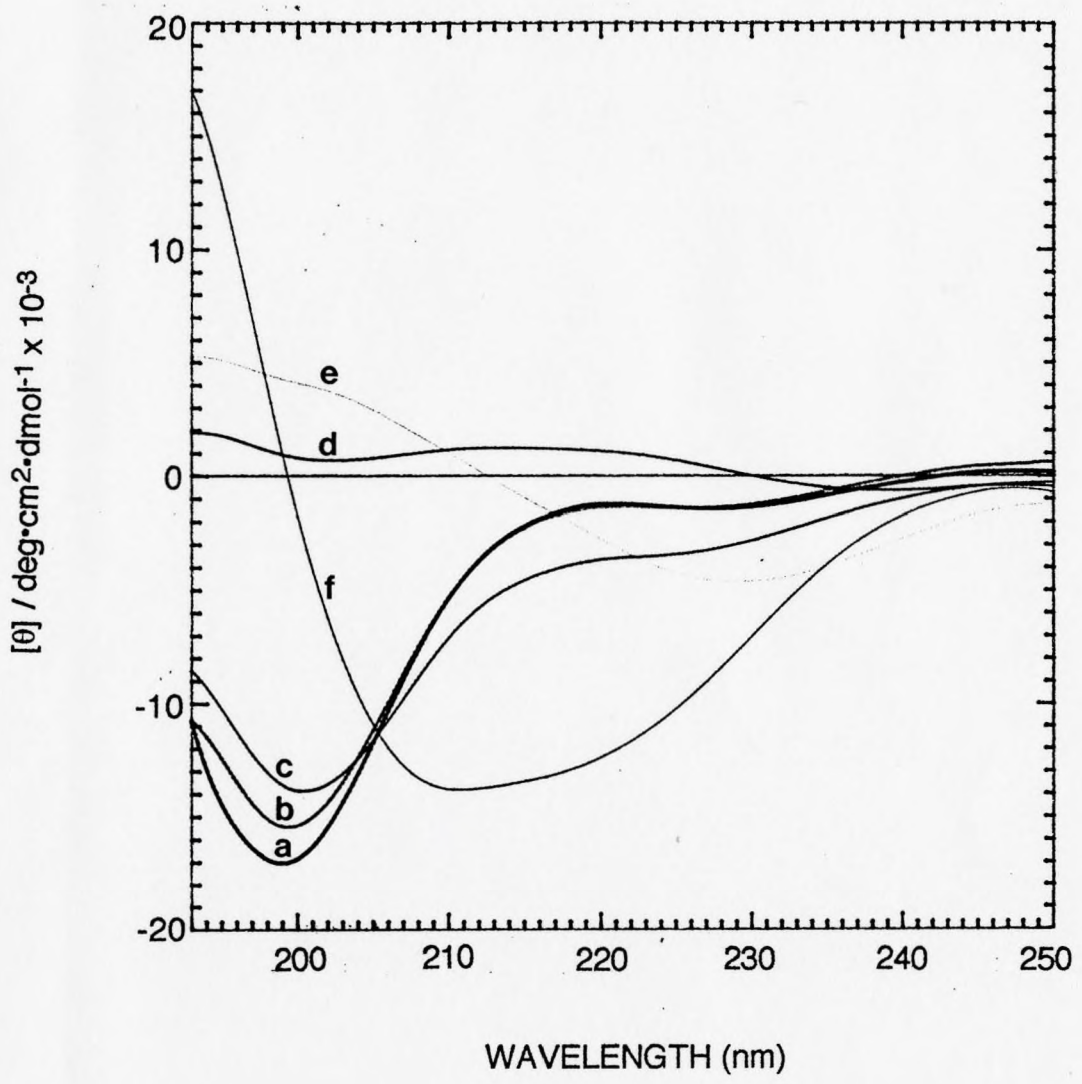


Figure 3.9: Effect of pH on AAD. Spectra were acquired in 4 mM buffer solutions suited to the particular pH required containing 50 µg/mL protein. Solution pH was determined using a microprobe before acquisition. Spectra (*a-f*) were acquired at pH 7.3 (MOPS), pH 5.65 (succinate), pH 4.72 (succinate), pH 3.97 (citrate), pH 3.25 (citrate), pH 2.85 (phosphate)

chloride salt to a final concentration of 1 mM (a 200-fold molar excess) and comparing spectra. No conformational change was observed for either salt. In the second experiment, a helical spectrum was initially induced by addition of 30% TFE to a 20 μM solution of AAD in 20 mM sodium phosphate, pH 7.05 with 1 mM of either divalent metal salt to potentially support the stabilization of the helix. This solution was then diluted 4-fold with a 30% TFE solution to a 5 μM protein concentration and a 250 μM salt concentration (a 50-fold molar excess; 2.4-fold acidic residue excess) to maintain a helical structure, or the solution was diluted 4-fold with water to produce a TFE concentration (7.5%) that could not support a helical conformation in the absence of a metal. However, even the presence of metal could not stablize the helical conformation at this TFE concentration. It cannot be discounted that at another TFE concentraton, an effect is observed. At 30% TFE, there is some deviation in the 200-205 nm region of the spectrum but this could be attributed to a generally greater inaccuracy of measurement in the far ultraviolet.

3.4.3.5. Effect of pH on 2° Structure

With scarsely few basic residues, AAD's calculated pI of 3.3 is dominated by aspartic acid residues ($pK_1 = 2.9$). From mutations designed to decrease the overall acidity of AAD (Cress and Triezenberg, 1991b), a degree of flexiblity is observed and cannot be attributed to the protonation of a specific residue.

To a 5 μM solution of AAD in distilled water, the pH was either adjusted by a small volume of dilute HCl or by the addition of a concentrated buffer solution (MOPS, succinate, citrate or phosphate) to 4 mM. Immediately prior to acquistion at 25°C, the pH was read with a micro-probe calibrated from pH 4-7.

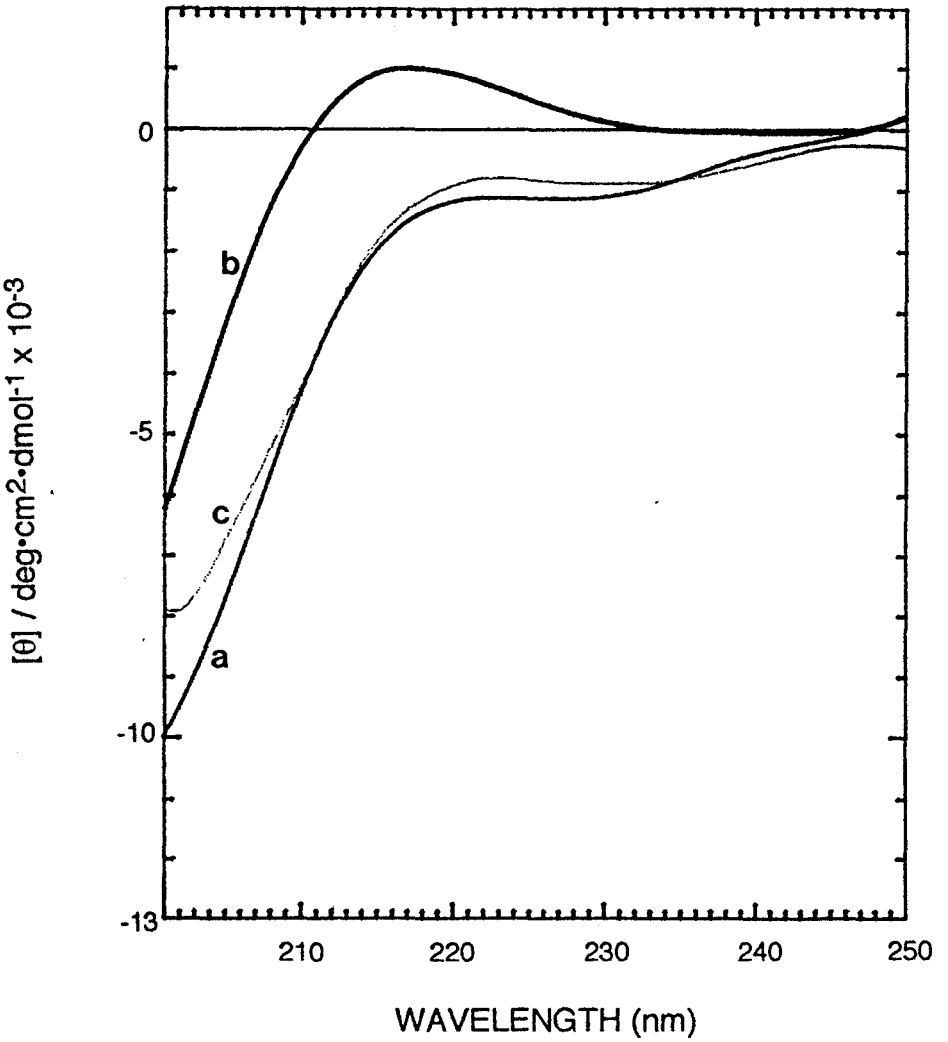


Figure 3.10: Effect of poly-Lysine on AAD. Spectra were acquired of a 50 μg/mL AAD solution in 5 mM HEPES, pH 7.4 in the (*a*) absence and (*b*) presence of 83 μg/mL poly-L-lysine (230 residues/polypeptide). The concentrations of AAD and the homopolymer were equimolar on a per residue basis. To illustrate any interaction between the two polypeptides, a difference spectrum (*c*) was plotted.

From pH 7.33 to pH 4.72, the magnitude of the trough at 198 nm steadily decreased though the spectra were still reminiscent of a random-coil conformation (Figure 3.9). Upon lowering the pH further from 4.0 to its isoelectric point, two reproducible spectra were obtained that could not be assigned to a known conformation. The composition of these spectra also could not be predicted by CONTIN with an acceptable degree of accuracy. The pH 4 spectrum was very peculiar since almost no circular dichroism was observed. It should be noted that precipitation was observed when HCl or buffer was initially added though the solution upon complete mixing was not visibly turbid in the quartz cell. Near the pI, solubility is at a minimum; perhaps an aggregation effect was observed. As the pH was decreased from the pI to 1.9, spectra reminicent of a helical conformation were observed and could be saturated in the pH range 2.5 to 1.9. The spectra at low pH differed from those produce by TFE in that the minima at 208/222 nm were broader yet still comparable in magnitude. Perhaps a population of AAD at the lower pH still exists in an aggregated form which upon addition to a typical helical profile produces the resultant spectrum.

3.4.3.6. Effect of Cationic Species on 2° Structure

To mimic a non-specific ionic interaction of AAD with its target, various cationic species were added. Sodium dodecyl sulfate (SDS) has been previously used by Kubota and Yang (1986) to simulate a membrane-like environment for the 26-mer, mellitin, whose mechanism of action involves integration into a membrane. A range of alkaline pH (this polypeptide is basic) from 7-12 and a range of SDS concentration up to 10 mM was studied; at neutral pH and 1 mM/10 mM SDS, mellitin was observed to adopt a helical conformation. A 1.0 % (34 mM) SDS solution containing 5 μM AAD did not show any conformational change (spectrum not shown). Similar results were obtained with 0.5% and

0.1% solutions of the polycation, CTAB, containing 11 μM AAD in 10 mM MOPS, pH 7.1.

To test the potential requirement for a polycation that resembles a protein in size and structure, poly-L-lysine with a mean residue length of 230 was used. Assuming a molecular of ca. 40 kDa, poly-L-lysine was added to a 50 μg/mL solution of AAD such that an equimolar ratio on a per residue basis was preserved. Spectra in the presence and absence of the homopolymer were acquired (Figure 3.10) and subtracted to indicate any change in conformation. Unlike mellitin, which was converted to a helical form upon addition of poly-L-glutamic acid but not poly-D-glutamic acid (Takeda and Moriyama, 1991), AAD exhibited no conformational change. This experiment was also repeated at a 2-fold higher AAD / homopolymer concentration with similar results.

3.4.4. Summary

CD spectroscopy of the AAD describes a domain which is flexible and sensitive to its environment with respect to both hydrophobicity and protonation. This would be consistent with mutational studies of the Vmw65 AAD which demonstrate that not only charge but also the nature and pattern of the hydrophobic residues are of important consequence for transactivation (Cress and Triezenberg, 1991b). An α-helical conformation for AAD did not predominate as predicted through computer algorithms; however, a random-coil description is not suitable either since tertiary interactions do exist and are influenced by hydrophobicity. To conclusively illustrate a necessary conformational change in AAD will require identification and purification of its target protein(s).

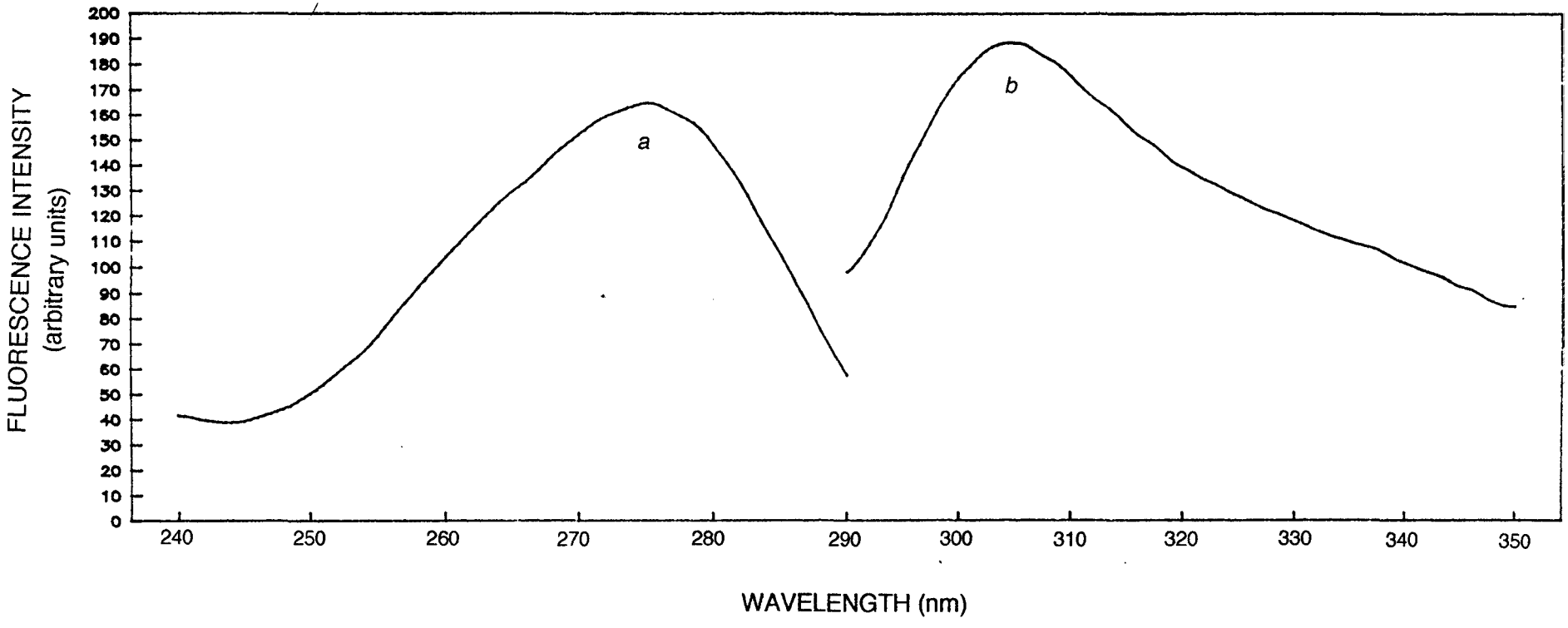


Figure 3.11: Excitation and Emission Spectra of AAD. An excitation spectrum (*a*) was acquired at 315 nm of a 30 μg/mL AAD solution in 50 mM Tris-Cl, pH 7.6. A 278 nm peak wavelength was selected for the emission spectrum (*b*) and for subsequent spectra in this study. A peak was observed at 305 nm was observed in the emission spectrum.

There are many interesting experiments that could be performed if target(s) are known. By using small peptides from regions in the target(s) that are important for AAD binding from mutational studies, a subtractive CD study similar to poly-lysine addition could be performed. A more sophisticated experiment would use 2D-NMR transfer NOE methods to determine new contacts made between AAD and its peptide. From the knowledge of specific interactions, a "rational" approach can be undertaken to design therapeutic compounds that could interfere with AAD-target binding.

The CD results presented are similar to those of Van Hoy et al. (1992) on an AAD from the yeast transactivator, GCN4. This group supplemented their CD data with a 2D-NMR study though the absence of intra-strand NOE's made a tertiary structure assessment impossible. Unlike Vmw65-AAD, their polypeptide was chemically synthesized and much shorter (19 aa *versus* 89 aa) but still retained modest activity.

An interesting phenomenon with respect to small peptides was cited by Dyson et al. (1988) after a CD and 2D-NMR investigation on a 19 residue basic polypeptide from the C-terminus of myohemerythrin. Like the studies on Vmw65-AAD and GCN4-AAD, a random-coil structure in water, observed by CD, could be induced into an α-helical conformation by TFE. However, NMR described a dynamic structure that contained a population of rapidly folding β-structures that fluctuated through a random-coil intermediate. Upon addition of TFE, only the dynamic regions in the polypeptide assumed an α-helical conformation. This phenomenon was dubbed a "nascent α-helix". In relation to their polypeptide, the authors suggested that the ability to fold was necessary for antibody presentation. A similar motif has been implicated for the basic region adjacent to the leucine zipper of GCN4 (Saudek et al., 1991) and recently for AAD by 1D and 2D-

NMR methods (O'Hare and Williams, 1992). The 1D-NMR ^1H chemical shifts were reminiscent of a predominantly random-coil structure (Wishart et al., 1992) and the subsequent 2D spectra similarly showed few intra-strand NOE's and evidence of transient β-turns.

3.5. Fluorescence Spectroscopy of AAD

A fluorescence study was initiated to exploit the two tyrosine residues in AAD resident at position 465 (near the C-terminus) and at position 488 (at the extreme C-terminus). The fluorescent properties of tyrosine with respect to peptides and proteins has been recently reviewed by Dewey (1991) and Lakowicz (1991). Tyrosine exhibits two absorption bands due to two $\pi \rightarrow \pi^*$ transitions near 220 and 280 nm which produce a broad fluorescence emission near 305 nm. In this mode, the phenolic group will donate its proton to form hydrogen bonds with the solvent. Formation of the tyrosinate anion at high pH reveals another mode where the fluorophore is poised to form hydrogen bonds by accepting a proton from the solvent. The ability to form hydrogen bonds in the interior of the protein or with the solvent affects not only the absorption / emission maxima but also the quantum yield. Taken alone, the solvent will influence tyrosine fluorence through its polarizability and its ability to form hydrogen bonds (polarity). The difficulty in attributing various physical parameters from solvent mixtures and environmental effects due to location in a polypeptide have resulted in a set of general rules to explain tyrosine fluroscence. Where applicable, these rules or trends have been used to describe features of the following experiments.

3.5.1. Materials and Methods

Fluorescence measurements were performed on a Perkin-Elmer LS-50 Fluorimeter interfaced to an IBM-PC running Perkin-Elmer PECCS software. 1.00 cm quartz cells equipped for use with a magnetic flea were used for all experiments. Temperature was maintained by a Brinkmann-Lauda RC6 circulating water bath. For experiments in which temperature was varied, a thermocouple was installed in a separate quartz cell adjacent to the sample (in a four-chambered cell holder). A constant intensity light source over all wavelengths used was supplied by the unit using an internal control. Spectra were acquired using the optimum slit width (a compromise between providing an intense signal and eliminated unnecessary bandwidth). Generally, the excitation / emission slits were set to 3-5 nm. One acquistion was deemed sufficient at a scan rate of 100 nm/min. AAD used for the experiments was from a single production and was deemed >95% pure by SDS-polyacrylamide electrophoresis (data not shown). Reagent-grade L-tyrosine (Sigma) was used without further purification. For quenching experiments, molecular biology grade acrylamide (Bethesda Research Laboratories), optical grade CsCl (Sigma) and reagent grade KI, KCl (Sigma) were dissolved in double distilled, deionized water and clarified through a 0.2 μm filter. The KI solution was used soon after preparation before a characteristic yellowing of the solution occurred due to the formation of I_3^-.

3.5.2. Results

3.5.2.1. Fluorescence Spectra of Tyrosine, its Derivatives and AAD

The optimum excitation wavelength of 278 nm was determined by exciting a 30 μg/mL AAD solution in 50 mM Tris-Cl, pH 7.6 and monitoring the emission at 315 nm. A peak at ca. 305 nm was used for subsequent experiments (Figure 3.11).

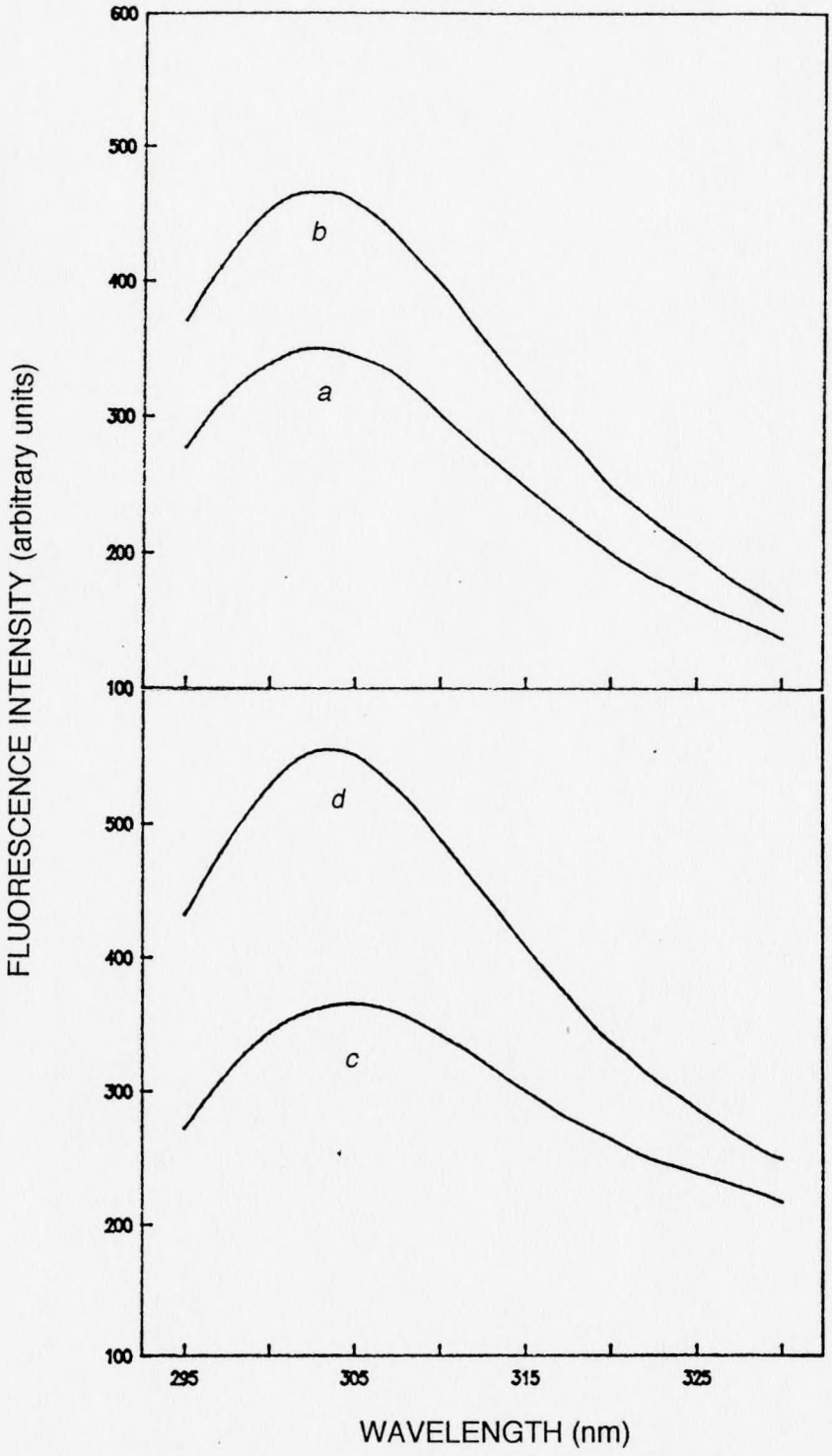


Figure 3.12: Emission Spectra of Tyrosine and AAD. Emission spectra (excitation wavelength = 278 nm) were acquired of a 6.6 μg/mL L-tyrosine solution in (*a*) distilled water and (*b*) 50% (v/v) TFE and of a 30 μg AAD solution in (*c*) 50 mM Tris-Cl, pH 7.6 and (*d*) 50% (v/v) TFE. λ_{max}(L-tyr) = 303.0 nm; λ_{max}(L-tyr + TFE) = 302.5 nm; λ_{max}(AAD) = 305.0 nm; λ_{max}(AAD + TFE) = 303.5 nm.

Using the previously determined excitation wavelength, spectra of AAD and a 6.6 μg/mL L-tyrosine solution were acquired in the absence and presence of 50% (v/v) TFE. Shown in Figure 3.12, L-tyrosine experienced an increased quantum yield upon addition of TFE with a λ_{max} of 303 nm consistent with accepted values (Cantor and Schimmel,1980). A λ_{max} of 305 nm was observed for AAD at physiological conditions that blue-shifted slightly to 303.5 nm upon addition of TFE. Unlike tryptophan, tyrosine experiences small spectral shifts in response to its enviroment. This blue-shift could indicate that one or both of the tyrosine residues in AAD could have experienced a conformational change that caused their migration to the surface of the protein. A necessity for a conformational change is supported by the lack of blue shift for free tyrosine and marked spectral change noted from circular dichroism spectroscopy.

3.5.2.2. Quenching Analysis of AAD

A quenching analysis using the Cs^+, I^- and the non-ionic quencher, acrylamide, was performed with N-acetyl-tyrosine-methyl ester (ATME) used as a control. Results are shown in Figure 3.13. Of the two ionic species, Cs+ was an ineffective quencher under any condition. Iodide ion is a widely used anionic quencher due to its high efficiency. In the absence of TFE, the ATME fluorescence was appreciably quenched, unlike the AAD fluorescence. Acidic residues adjacent to the fluorophores in AAD could be contributing to the weak quenching by electrostatically shielding the tyrosine residues from I^-. Due to its decreased efficiency, the effect of Cs^+ on AAD fluorescence may not be noticable. Acrylamide, another highly efficient quencher, should be immune to the electrostatic effects and indeed, considerable quenching of AAD was observed in the absence of TFE. The presence of TFE impeded the ability of all three species to quench tyrosine fluoresence; hence, contributions due to an alternative AAD conformation cannot be discerned.

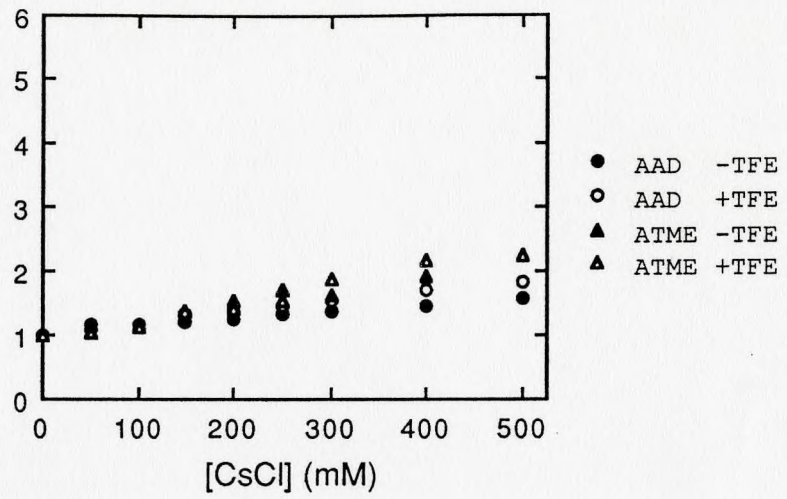
The Stern-Volmer describes the contributions of dynamic quenching (collsional quenching that occurs after formation of the excited state) and static quenching (formation of a ground state complex with a quenching agent before excitation) through the mathematical relationship shown below:

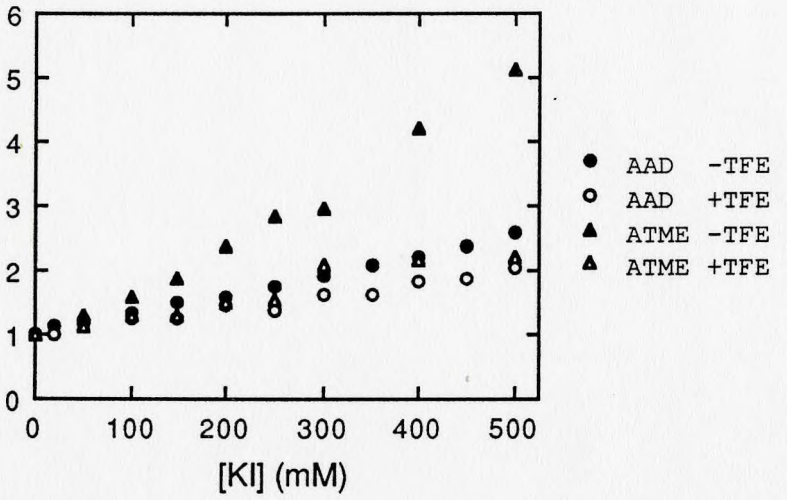
$$F_o / F = (1 + K_{sv} [Q]) \cdot (1 + K_a [Q]) \qquad [3.3]$$

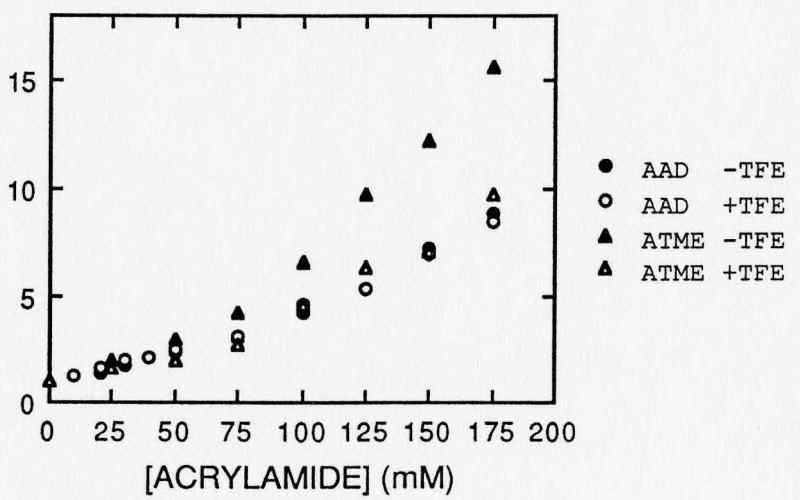
If the static quenching is negligible, the corresponding rate constant, K_a, will be small making the second term in Equation 3.3 approach unity which graphically should be depicted as a straight line of slope K_{sv}, the dynamic rate constant. Additionally, K_{sv} represents $k_q \cdot t_o$; the product of the quenching rate constant and the lifetime of fluorophore in the absence of quencher (since a period will occur between excitation and collisional deactivation of the excited state). In Figure 3.13, some static quenching is observed since the plots though in many cases, inefficient quenching makes it difficult to determine whether some plots are linear, or not. To calculate the quenching rate constants from the curve, the lifetimes of the two fluorophores must be experimentally derived. Experimental procedures are available for determining a fluorescence decay profile and deconvolving the contributions of various flurophores (Knutson et al., 1982).

It should be noted that other quenching effects endogenous to AAD may be occurring. These can include an interaction with the peptide backbone, excited-state proton transfer to the many available carboxylic acid side chains and tyrosine-tyrosine energy transfer since an NMR study showed the two flurophores to be in close proximity (O'Hare and Williams, 1992). Interactions with carboxylic acid side chains are usually negligable since they are competing with the aqueous solvent. Aside from studying tyrosine-tyrosine

Figure 3.13: Quenching Analysis of AAD. The fluoresence at 305 nm of a 30 μg/mL AAD solution or of a 0.25 μg/mL N-acetyl-tyrosine-methyl ester (ATME) solution in 50 mM Tris-Cl, pH 7.6 ± 50% TFE was monitored in the presence of quenching agents at 25°C. *Panels A,B:* Stern-Volmer plots of the ionic quenchers, Cs^+ and I^-. A constant ionic strength of 0.5 M maintained by the addition of KCl. *Panel C:* Stern-Volmer plot of the neutral quencher, acrylamide.


6
5
4
3
2
1
0
0
100
200
300
400
500
[CsCl] (mM)
AAD -TFE
AAD +TFE
ATME -TFE
ATME +TFE


6
5
4
3
2
1
0
0
100
200
300
400
500
[KI] (mM)
AAD -TFE
AAD +TFE
ATME -TFE
ATME +TFE


15
10
5
0
0
25
50
75
100
125
150
175
200
[ACRYLAMIDE] (mM)
AAD -TFE
AAD +TFE
ATME -TFE
ATME +TFE

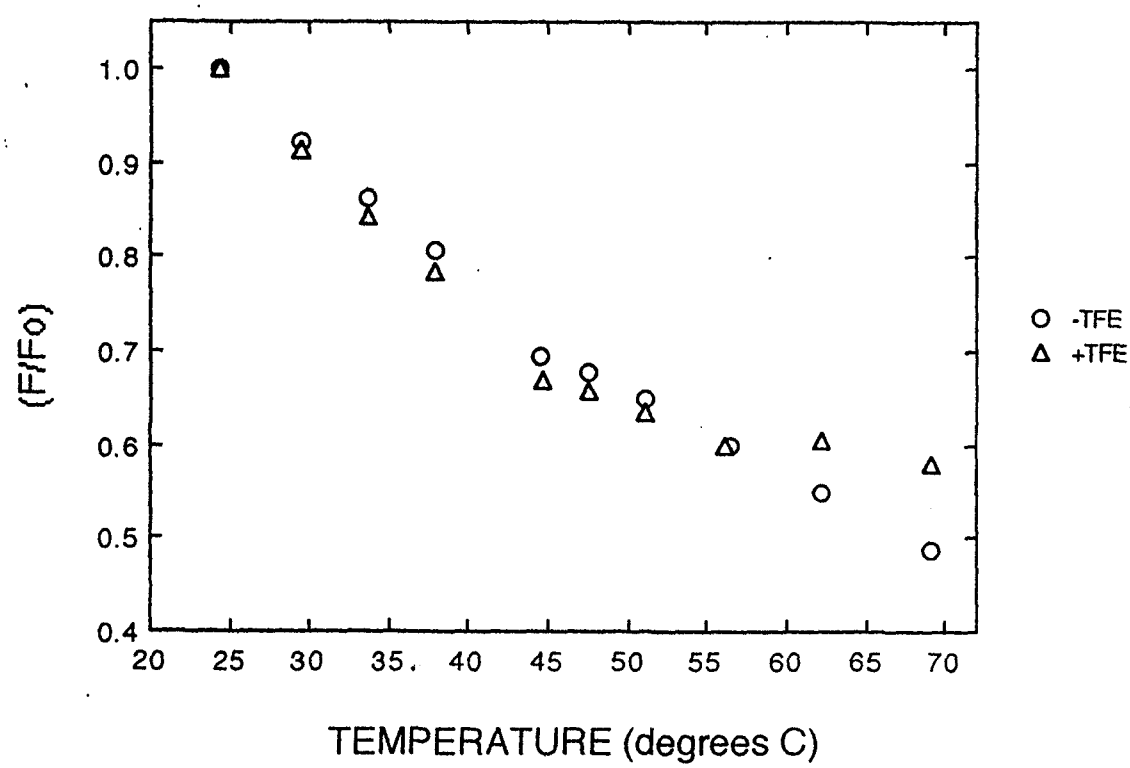


Figure 3.14: Effect of Temperature on AAD Fluorescence. A 30 μg/mL solution of AAD in 50 mM Tris-Cl, pH 7.6 was excited at 278 nm and the fluorescence followed at 305 nm at various temperatures. Plotted is fluorescence intensity relative to the flourescence intensity at 24.4°C.

energy transfer, mutagenesis to tryptophan could be performed or alternatively, a synthetic fluorophore such as 7-acetoxycoumarin-3-carboxylic acid, succimidyl ester (Molecular Probes, Eugene, OR) could coupled to a free amino group. This fluorophore absorption spectrum overlaps the emission spectrum of tyrosine.

3.5.2.3. Effect of Temperature on AAD Fluorescence

A experiment monitoring AAD fluorescence at varying temperatures in the absence or prescence of 50% (v/v) TFE was performed similar to an earlier circular dichroism experiment. From the results shown in Figure 3.14, an abrupt change in slope is observed at 35-40°C. In contrast, circular dichroism spectra of AAD over this range previously visually showed little conformational change (consult Figures 3.5-3.7) and no abrupt ellipticity change when plotted. In general, decreased fluorescence can be attributed to a greater occurence of static quenching mechanisms dependent on activation energy and variation in the electronic state of the fluorophore. Nevertheless, a pronounced change occurred upon heating and could potentially represent an subtle alteration in the conformation of one or both of the tyrosine residues.

3.5.3. Conclusions

Fluorescence spectroscopy can be an invaluable tool for probing protein conformation due to a flurophore's sensitivity to its surroundings. For AAD, this form of spectroscopy did not provide any definitive information on its conformation in water or 50% TFE except that a conformational change did occur in the presence of TFE. Perhaps, a redesign of the experiment incorporating energy transfer (tyr*->trp, tyr*->probe) or a detailed determination incorporating tyrosine fluorescence lifetimes would be fruitful.

Sequence changes in AAD, however, will require an assay such as the production of a suitable mutant virus to verify that any modification has influenced the activity. A suitable *in vitro* assay will likewise be required for any chemical modifications of AAD.

3.6. Equilibrium Sedimentation Ultracentrifugation

The results described in this section were kindly provided by Dr. Preston Hensley (SmithKline-Beecham Pharmaceuticals, King of Prussia, PA). The basis for this technique rests on the velocity at which the sample is centrifuged. The velocity chosen should be inefficient to pack the sample against the bottom of the cell (reviewed in Cantor and Schimmel, 1980). The resulting centrifugal force on the solute molecules will be opposed by diffusive forces ultimately creating a characteristic equilibrium distribution. By incorporating quartz windows, the absorbance of the solution can be measured throughout the cell. Two important parameters, namely molecular weight and molecularity, can readily be derived from sedimention equilibrium data. At equilbrium, the concentration at any point in the cell can be described by Equation 3.4 below which in turn, readily conveys the parameters to consider in such an experiment.

$$c_r = c_o \cdot \exp\left[M \cdot (1-\nu\rho) \cdot \omega^2 \cdot (r^2-r_m^2) / 2RT)\right] \qquad [3.4]$$

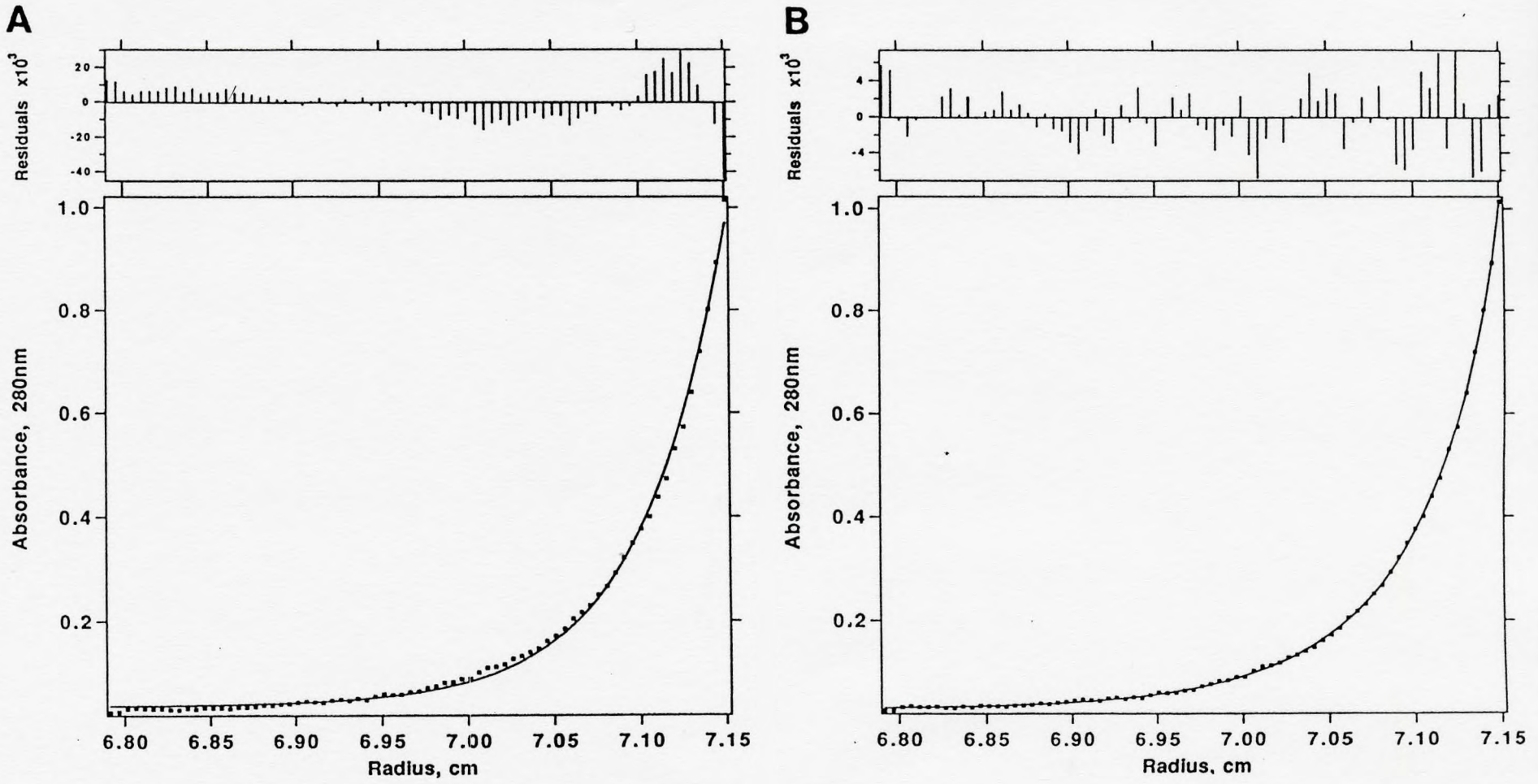
c_r represents the concentration at an arbitrary position and c_o represents the concentration at the meniscus. M is the molecular weight; for AAD, it is 9330 Da. The partial specific volume, ν, can be calculated (Zamyatnin, 1972), as well as the solution density, ρ. The angular velocity, ω, the distance from the centre of rotation, r, and the position of the meniscus, r_m, are given parameters.

An experiment was designed as follows: AAD was prepared in 40 mM sodium phosphate buffer, 150 mM NaCl, pH 7.3 (PBS buffer) at 20°C. The solution density was calculated at 1.007 gm/mL. AAD was provided at 0.89 mg/mL based upon an OD_{280nm} of 0.25. The partial specific volume of AAD was calculated to be 0.701 mL/gm. The experiment was performed using a Beckman Model E analytical ultracentrifuge equipped with a photoelectric scanner and a temperature control system. Three different data sets were collected and analyzed at each of three rotor speeds (24 000, 34 000 and 44 000 rpm), ca. 12 hours after reaching equilibrium.

Shown in Figure 3.15 is a representative data set acquired for AAD at physiological conditions. Initial analysis of the data to a monomeric species provided a molecular weight of 11002 Da ± 340 Da using a 65% confidence limit. This molecular weight was slightly higher than the sequence molecular weight of 9330 Da for AAD so the data was analysed further assuming the presence of AAD multimers. The data was best fit to a monomer-tetramer equilibrium evidenced by a random distribution of residuals. The calculated monomer contribution was 98.9% based on a calculated equibrium constant, K_{14}, of 8.64 x 10^{-11} M^3.

The corresponding ultracentrifugation experiment of AAD in 50% TFE was not performed due to time and equipment restraints. An ellipsoidal volume (with the missing data supplied by a diffusion coefficient from a sedimentation velocity experiment rather than a gel filtration determination of a Stokes radius) and the aggregation state would have be useful in the interpretation of this structure.

Figure 3.15: Equilibrium Ultracentrifugation of AAD. Shown is an absorbance profile for a 0.1 mM AAD solution in phosphate buffered saline, pH 7.3 fitted to a 100% monomeric (*Panel A*) or a monomer / tetramer distribution (*Panel B*). Consult Section 3.8 for particulars. Residuals are provided above the plots.


A
Residuals x10^3
20
0
-20
-40
1.0
0.8
0.6
0.4
0.2
Absorbance, 280nm
6.80
6.85
6.90
6.95
7.00
7.05
7.10
7.15
Radius, cm
B
Residuals x10^3
4
0
-4
1.0
0.8
0.6
0.4
0.2
Absorbance, 280nm
6.80
6.85
6.90
6.95
7.00
7.05
7.10
7.15
Radius, cm

3.7. Determination of Stokes Radius and Molecular Volume

The Stokes' radius is a value used to describe the volume of a macromolecule assuming it is spherical. This value is derived experimentally from gel filtration chromatography. Using available proteins of known Stokes radius (Siegel and Monty, 1966), a TSK-G4000SW column was run on a Beckman System Gold HPLC apparatus using a mobile phase of 0.2 M NaCl in 20 mM Tris-Cl, pH 8.0. The void volume, V_o, and the maximum retained volume, V_t, was obtained by running acetone and blue dextran through the column, respectively. The fraction of the volume of the gel that is available for each substance, $K_{av} = (V_e - V_o) / (V_t - V_o)$, (Laurent and Killander, 1964) was calcuated for each standard retention time converted to elution volume, V_e, and plotted against known Stokes radii shown in Figure 3.16. The observed K_{av} of AAD corresponded to a Stokes radius of 28 Å.

The Stokes radius was then used to calculate (f / f_o), the ratio of the frictional coefficient of AAD to the frictional coefficient of a solvated sphere. Prerequisite to this calculation, the partial specific volume, ν, was calculated to be 0.701 g^{-1}•mL based on amino acid composition (Zamyatnin, 1972) and δ, the degree of hydration, was set to a typical value of 0.35 g water per gram of protein. The input Stokes radius produced a (f / f_o) of 2.00 corresponding to major/minor prolate ellipsoidal axis ratio (a / b) of 20.1 (Tinoco et al., 1978). This ratio could be simply related to a ellipsoid of volume $4/3 \cdot \pi \cdot a \cdot b^3 = (M \cdot \nu)/N_o$. Input values into the latter equation produced dimensions of 6.04 Å x 101 Å. Calculations performed assuming an oblate ellipsoid produced improbable dimensions of 1.85 Å x 37.3 Å.

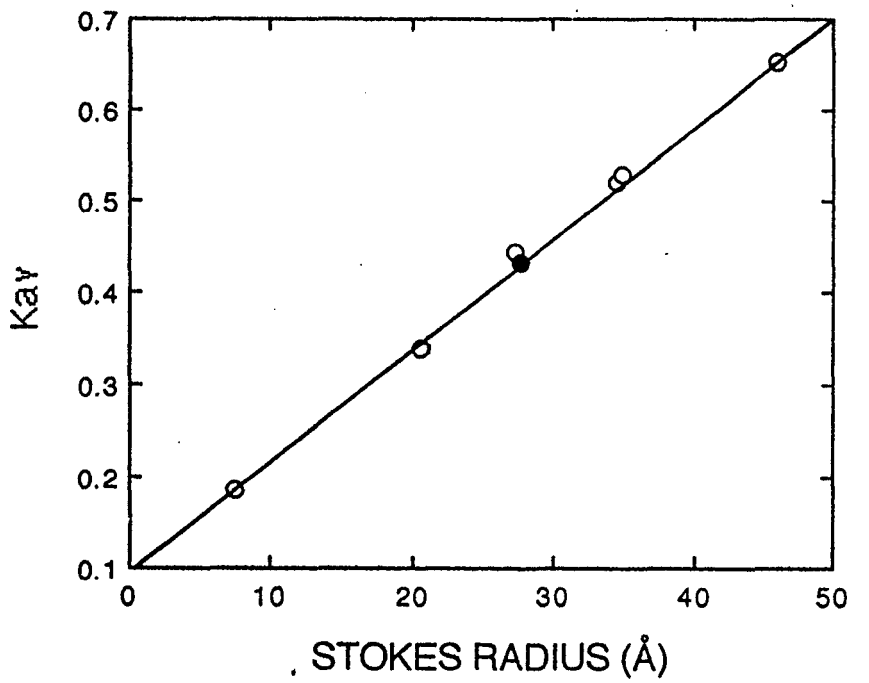


Figure 3.16: Stokes Radii Data from Gel Filtration Chromatography. Gel filtration was carried out using the following standards (open circles) for which Stokes radii are known: vitamin B_{12} (7.5 Å), myoglobin (20.7 Å), ovalbumin (27.3 Å), calf intestinal alkaline phosphatase (34.6 Å), bovine serum albumin (35 Å) and yeast alcohol dehydrogenase (46 Å). The observed K_{av} of AAD corresponded to a Stokes radius of 28 Å (closed circle).

3.8. Crystallization Trials

Two methods, NMR spectroscopy and crystallography, exist to determine the tertiary structure of a protein at the molecular level. Though NMR methods were not exploited in this project, it could prove to be a valuable option in the future especially with the advent of 4D-NMR to unambigously assign protons (Boucher et al., 1992). Crystallography is current better for larger molecules but regardless of size, there is always a requirement to obtain a crystals that will diffract well and be amenable to heavy-metal derivatization.

It is important to ensure that the AAD used was free of contaminants to prevent the formation of mosaic crystals that would diffract poorly. Generally, isoelectric focusing is used to separate any closely related degradative products. Isoelectric focusing is not suitable for further AAD purification since the lowest pI 3-5 ampholytes available would not provide adequate resolution. Concentrations of 5-10 mg/mL were typical for experiments and were acheived by solubilzing the lyophilized product in a small volume of water or buffer. Not all of the AAD would completely dissolve; hence, solutions were 0.2 μm filtered prior to use. Sodium azide to 0.02% was added as an anti-microbial. A physiological pH of the AAD solution was maintained with either a sodium phosphate or Tris-HCl. In retrospect, it may have been beneficial to use a lower pH to lower the solubility. There were structural changes, however, from the CD spectra which could yield a structural with no biological significance though at any pH, the biological significance would be indeterminate in the absence of target molecules.

If Chryschem brand crystallization containers were used for trials, ca. 20 μL of protein would be placed into the inner chamber and 1 mL of precipitant would be placed in

the outer chamber. A closed atmosphere was accomplished by sealing the exposed wells with clear packaging tape. Most trials utilized the hanging drop method since less protein solution was required. Here, 1 mL of precipitant is initally placed in a 24 well tissue culture plate and the edges of the wells coated with silicone grease. Next, 5 μL of protein solution is placed onto a siliconized 18 x 18 mm cover slip which is then carefully turned upside-down over the well containing the precipitant. Both methods discussed were incubated at 4°C or room temperature and monitored daily using a low power binocular microscope equipped with two polarizing filters situated above and below the crystallization container. If crystals were present, they would refract light against a dark background created by aligned polarizers.

Prior to setting up a matrix of conditions, a preciptation point was determined for each precipitant solution used. A 20 μL drop of concentrated AAD would be placed on a depression slide to which small amounts of concentrated precipitant were quickly added until a limit was defined. A common precipitating agent, PEG-8000 (McPherson, 1985), was ineffective at precipitating AAD at concentrations up to 25%. Ethanol was also ineffective up to 64%. AAD was found to precipitate in MPD (a common organic solvent used for crystallizations) at 60% and in ammonium sulfate concentrations of 18-19%.

Trials varied around the MPD and ammonium sulfate precipitation points in 2% increments. No crystals were observed in any of the solutions or concentrations. Similar trials with the cation, spermidine trihydrochloride, were also unsuccessful.

Since AAD does not possess a distinct structure, it may not crystallize under any condition. Additional trials using MPD and ammonium sulfate were performed using 6.7

mg/mL intact GST-AAD fusion. By leaving the carrier on, it was hoped that it would impart some regularity to the overall structure and assist in packing. These trials were also unsuccessful. Hampton Research (Riverside, CA) markets a kit called "Crystal Screen" which contains over fifty common reagents used in crystallographic trials which could be useful in future attempts.

3.9. Commentary and Future Experiments

The biophysical data presented in this chapter from circular dichroism, fluorescence spectroscopy and equilibrium ultracentrifugation suggest a random-coil or extended structure for AAD that is monomeric in aqueous solution. The random-coil classification does not suggest that there is no structure; rather, that the conformation that AAD does adopt is non-standard by the absence of α-helical or β-sheet character. Data provided by studies on other AAD's and from NMR studies corroborate and augment the findings presented in this thesis. The NMR data suggests that a very dynamic polypeptide that is able to flucuate through many β-turn intermediates though in a longer time-frame, AAD would appear to be random.

Addition of TFE can induce a coil-helix transition as observed by circular dichroism spectroscopy. A requirement for such a transition has been both experimentally shown to be necessary (Giniger and Ptashne, 1987) and unnecessary (Cress and Triezenberg, 1991b). To show that an amphipathic α-helix was unnecessary, Cress and Triezenberg designed 21 mutants in which one or more aspartic acid residues was replaced with asparagine with the intention of systematically removing acidic residues from the face of the putative α-helix. No correlation was observed upon analysis of *in vivo* activities from a

thymidine kinase reporter gene. It is unfortunate, though, that an AAD construct truncated to position 456 was used since missing residues could have played a role in determining conformation. Also, another amphipathic α-helix was predicted in this region. Giniger and Ptashne took a different approach by fusing two artificial polypeptides to the GAL4 DNA binding domain. Each polypeptide had the same composition but only one had a distribution that could be termed amphipathic and only it could transactivate a GAL1-*lacZ* fusion gene. This results could have been more plausible if there was evidence that these synthetic peptides could indeed form α-helicies. In light of this fact, the results of Giniger and Ptashne remain doubtful.

Support of the modularity hypothesis of Frankel and Kim (1991) requires a conformational change from an inactive to an active state and has been shown to be important for fos/jun (Patel et al., 1990), GCN4 (Weiss et al., 1990) and HIV-tat (Loret et al., 1991). A conformational switch would be important for AAD to tightly control this powerful transactivating domain. The conformational change in AAD needed to exert its effect may be quite subtle but upon combination with other weak effects (more AAD molecules or other transactivators) it may be able to provide the necessary influence on its target. A weaker conformational change may be benefical to the dismantling of the pre-initiation complex once AAD has performed its function. Finally, a labile domain could adapt to its multiple targets more readily than a domain that assumes only one catalytic state.

Future experiments will likely focus on a more strigent mutagenesis study of not only AAD but also its targets similar to a study by Lee and Berk (1991) on the interaction between the E1A activation domain and the basic repeat within TFIID. Though the basic

region of TFIID could be excised and expressed as an independent polypeptide followed by a biophysical study using a technique sensitive to secondary structure such as circular dichroism spectroscopy, it would not be possible to determine if either polypeptide were undergoing a conformation change upon a stable interaction. Unlike AAD which comprises a large, autonomous domain, the basic region of TFIID is postulated to be a molecular hinge between the two direct repeat lobes that comprise pseudo-monomers of a TATA-binding domain (Yamamoto et al. 1992); thus, the hinge conformation would likely be dependent on the conformation of the entire conserved carboxy-terminus. DNA bending is also known to occur at the TATA box (Horikoshi et al., 1992) which would preclude a protein conformational change and potentially affect interactions with other proteins. Before a biophysical study, an approach using the far-western technique (Lee and Berk, 1991) or a conventional assay using antibodies to block any potential interaction between AAD and its target could be initially performed. If a *bona fide* interaction exists, mutagenesis could then continue to further map the crucial areas. As new targets of AAD are identified and characterized, a database may be compiled to predict a motif. The biophysical techniques to definitively show a interaction would involve co-crystallization or NMR using intra-chain NOEs. Both of these highly complicated methods would benefit from a knowledge of the independent structures of each component. Of course, these experiments could also be extended to the basic regions of TFIIB.

If the experimental evidence for an AAD-TFIID interaction is coupled with the dependence on RNA pol II phosphorylation to initiate elongation, a new mechanism of AAD transactivation may be postulated. Both AAD and a phosphorylated RNA pol II carboxy-terminus have a high negative charge. If the AAD can compete successfully with

unphosphorylated RNA pol II for the TFIID binding site, AAD will essentially perform the function of TFIIH by dislodging RNA pol II to begin elongation.

Some of the most recent information regarding a mechanism for Vmw65 has come from White et al. (1992) using a GAL4-AAD chimera. They report that GAL4-AAD binds quickly to its promoter and stays bound over many rounds of transcription. They also report that GAL4-AAD increases the number of preinitiation complexes formed without affecting their rate of formation or enhancing their stability. Additionally, neither TFIID, TFIIA nor TFIIB were recruited to the transcriptional complex by GAL4-AAD. How can more complexes them form? Two possibilities include: (1) Binding of GAL4-AAD to TFIID causing the recruitment of auxillary factors such as TAFs or adaptors that could further stimulate binding of other transcription factors or (2) Binding of GAL4-AAD to the transcription complex causing it to be activated in a way that recruits more transcription factors or even additional targets of AAD.

In conclusion, the study of AAD will likely continue to be exciting and yield important information not only on transcription but also on the more global problems of molecular recognition and protein structure.

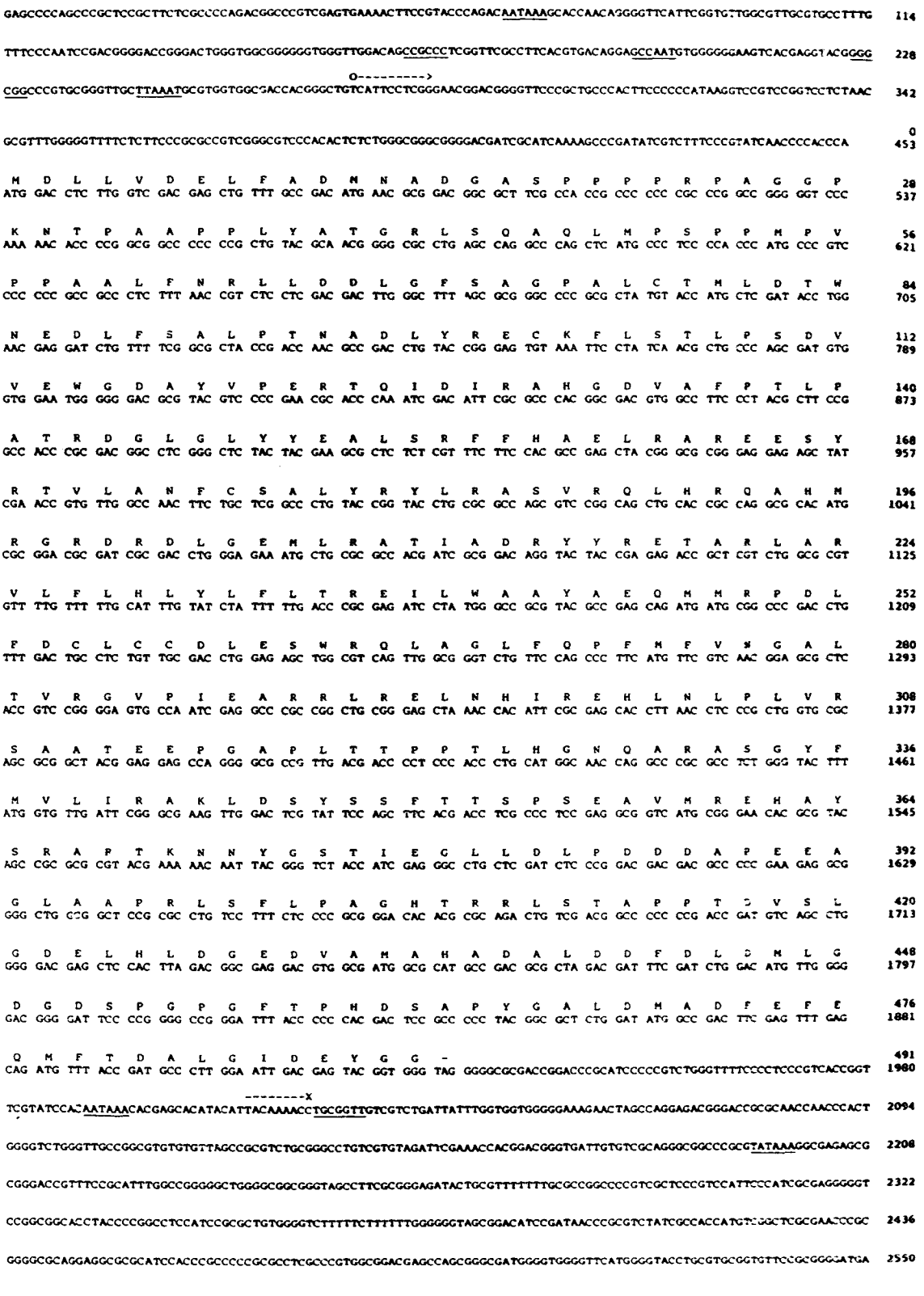
Appendix

A. Complete Sequence of Vmw65

B. Other Vmw65 Mutants as Fusions to Glutathione S-Transferase

C. Program Listing of <tocontin.exe>

Appendix A. *Complete Sequence of Vmw65.* Included are the 5' and 3' non-coding sequences and the corresponding amino acid sequence. The 490 amino acid product encodes a 54 kDa polypeptide (Dalrymple et al., 1985) and shares 86% identity and 93% similarity to its HSV-2 homolog (Cress and Triezenberg, 1991). Reprinted from Dalrymple et al. (1985).



Appendix B
Other Vmw65 Mutants as Fusions to Glutathione S-Transferase

Introduction:

Since the GARAT portion of the HSV-1 immediate-early promoter is necessary for transactivation (Spector et al., 1990), it is reasonable to assume that Vmw65 may interact with this region upon binding to Oct-1. Stern and Herr (1991) demonstrated the converse stating that Vmw65 will only bind to Oct-1 in the presence of GARAT. From experiments using a gel mobility shift assay, concentrations of Vmw65 suitable for complex formation in the presence of nuclear extract (containing Oct-1 and other necessary factors) are not sufficient to observe a Vmw65-promoter complex independent of nuclear extract. However, by adding a large excess of Vmw65 (Kristie and Sharp, 1990), some binding is observed and is specific to the GARAT sequence. To explain the Vmw65-GARAT interaction at high concentration, a certain population of molecules could perhaps rest in an alternate conformation that favors contact. This explanation is supported by observations of enhanced binding after denaturation/renaturation (Kristie and Sharp, 1990).

The region(s) of Vmw65 that contact the immediate-early promoter have not yet been determined though non-specific DNA binding has been observed using a synthetic polypeptide corresponding to amino acids 160-184, a region shown to be capable of interfering with wild-type transactivation (Werstuck and Capone, 1989a). Since the overexpression of the Vmw65-AAD as a fusion to glutathione S-transferase was successful, three other Vmw65 mutants were constructed to help deliniate the region responsible for the nucleic acid binding. Similar to AAD, high yields accompanied by a

facile purification were expected to provide ample quantities of Vmw65 mutants for the gel mobility shift assays. If a desirable mutant was found, an overexpression scheme would already exist to produce milligram quantities suitable for crystallization trials. Unlike AAD, such a mutant would likely possess a defined structure more amenable to successful crystallization.

Materials and Methods:

All plasmids were constructed with the intent to minimize the addition of foreign sequence at either termini and to include an opal termination codon. Unless otherwise noted, common molecular biological techniques previously discussed in earlier sections were used. All constructs were transformed into *E. coli* DH10β. Restriction enzymes and T4 DNA ligase were purchased from Boehringer-Mannheim, New England Biolabs and Pharmacia Molecular Biology.

Plasmid pGEX65/TGA was constructed by ligating a *Bgl* II/*Sph* I fragment containing most of the Vmw65 coding sequence from pSPUTK65 (the pSPUTK-xx series contains a more extensive mutiple cloning site than the original pGEX-xx series) into pGEXCT/TGA which had been previously cut with *Bam* HI/*Sph* I. Transformants were screened with *Nco* I/*Eco* RI and *Pst* I digestions. The predicted cleavage product would result in the addition of Gly-Ser-Pro to the N-terminus. Other bacterial Vmw65 constructs exist in the laboratory including GST-Vmw65 (amber termination codon) and Protein A-Vmw65.

Plasmid pGEXsal was constructed by excising a *Sal* I/*Eco* RI fragment from plasmid pGEX65 and inserting a *Sal* I/*Eco* RI cassette (5'-GTCGACGTGATCAGAATTC-

3' / 3'-CAGCTGCACTAGTCTTAAG-5') in its place. Transformants were screened with a *Nco* I/*Eco* RI digestion. This cassette added an opal termination codon and preserved the C-terminal residues. The predicted cleavage product would result in the addtion of Gly-Ser-Thr to the N-terminus. A similar construct existed in the laboratory as a fusion to Protein A. (PrA-sal).

Plasmid pGEXe47 was constructed by excising a 382 bp *Eco47* III fragment from plasmid pRITsal and inserting it into vector pGEX-2T which had been linearized with *Sma* I. The *Eco47* IIIfragment was purified by elution from a 5 % polyacrylamide gel containing 0.5x concentration of Tris-Borate-EDTA electrophoresis buffer. Transformants were screened with *Pst* I/*Kpn* I and *Bal* I digestions. The predicted cleavage product would result in the addition of Gly-Ser-Pro to the N-terminus and the Pro-Gly-Ile-His-Arg-Asp to the C-terminus.

Overexpression was induced by the addition of 0.2 mM IPTG to the culture medium at an OD_{600nm} of 0.5. Purification, thrombolysis and resolution of cleaved products by anion exchange chromatography have been previously discussed for the AAD fusion.

The gel retardation assay (gel mobility shift assay; bandshift assay)used was the optimized method of G. Werstuck (Werstuck and Capone, 1990). In a total reaction volume of 15-20 µL, 2-5 µg of protein extract (in 50 mM Tris-Cl, pH 8.0, 10 mM glutathione), 3 µg of poly(dI-dC) single-stranded competitor DNA, 5 µg of BSA and HeLa nuclear extract are incubated for 5 minutes at room temperature. Omission of the nuclear extract will allow the study of direct nucleic acid binding. 0.1 pmol of ^{32}P labelled DNA

probe (30000 counts) is then added and the mixture incubated for an additional 20 minutes at 30 °C. It had been previously shown that the order in which substituents were added had a significant effect on the outcome of the assay (Demczuk et al., 1991). A solution containing glycerol, bromophenol blue and xylene cyanol is added and the reaction is immediately loaded onto a pre-run (3 hours at 230 V) 3.5 % non-denaturing polyacrylamide gel containing 0.25x TBE and electrophoresed for 3 hours at 230 V.

Results:

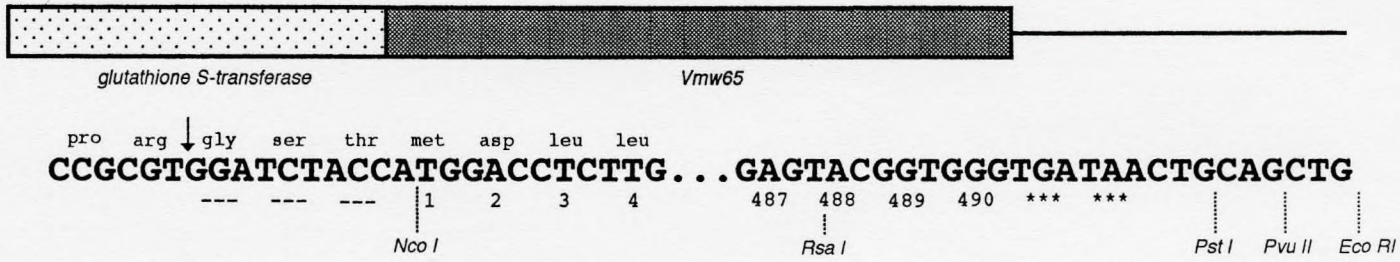
Figure B1 provides a detailed description of the three fusion proteins constructed. Briefly, a full-length (all 490 residues) Vmw65 fusion protein designated pGEX65/TGA, an AAD deleted (N-terminal 412 residues) fusion protein designated pGEXsal, and an internal region (residues 153-278) fusion protein designated pGEXe47 were constructed and overexpressed. The mutant, pGEXe47, was constructed because this region of Vmw65 contained many basic residues deemed necessary for a nucleic acid binding.

The largest culture of GST-sal was 12 L while overexpression of the GST-e47 and the GST-Vmw65 fusion was restricted to 1 L. In all cases, these bacteria were fermented, harvested and processed according to the GST-AAD protocol.

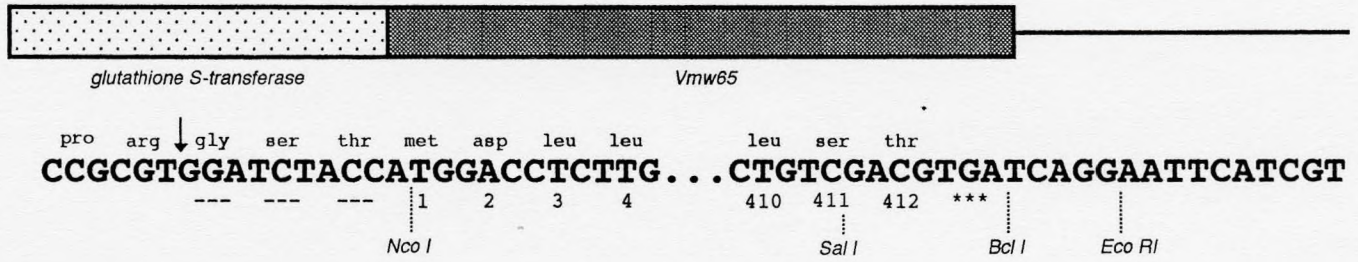
All three constructs were purified by affinity chromatography. From the gel shown in Figure B2, it can be seen that all suffered considerable degradation. It is interesting to note that no degradative products < 28 kDa were observed. One explanation for this observation is that the Vmw65 derived sequences are unstable while the GST carrier is quite stable. When GST is expressed alone, little degradation is observed. If the full-length

Figure B1: Mutants of Vmw65 as fusions to GST. The successful overexpression of the Vmw65 acidic activation domain prompted the construction of three other Vmw65 related fusion proteins. Below the graphical representation (not to scale) is the predicted nucleic acid sequence accompanied by the translated amino acid sequence near the GST-Vmw65 fusion junction and near the C-terminus of the fusion protein. Residues indicated by numbers correspond to residue positions in wild-type Vmw65, those indicated by (---) correspond to foreign residues introduced in construction. Arrows denote the cleavage site for thrombin. Restriction sites relevant to the constructions are indicated for reference.

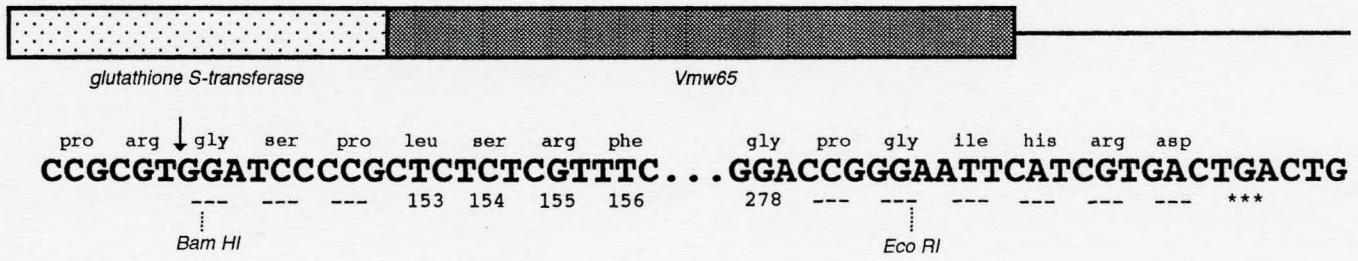
pGEX65/TGA



pGEXsal



pGEXe47



Vmw65 fusion has not folded into properly, it would likely be very susceptible to proteolysis. As there is no transcriptional or squelching assay available for Vmw65, verification of a functional protein remains elusive.

The highest overexpression was observed by the GST-sal construct. This was fortunate since it would serve as a complement to AAD. Overexpression of GST-sal was not as high as GST-AAD with yields in the range of 4 mg fusion protein/L bacterial culture. After elution from the affinity column, this fusion was unstable; preciptates could be noticed after 24 hours at 4°C or after 15 minutes at 37°C. Such instability made thrombolysis of GST-sal very difficult and in one instance, no sal fragment was recovered from an ion exchange column. On an instance where the sal fragment was recovered, it eluted before but very near the GST carrier. Since this construct lacked the acidic terminus, this observation was not unreasonable. Electrophoresis of the sal fragment eluted from the ion exchange column showed that it was not homogenous like its GST-AAD counterpart but still substantially pure.

Thrombolysis of the GST-Vmw65 and GST-e47 fusion proteins was not successful. In both cases, the predicted products could not be observed on a gel though the GST carrier was apparent (Figure B2). A faint band may have corresponded to Vmw65 but even if it was, its proportion to the rest of the proteins in the thrombin treated mixture was extremely low. It is puzzling that the Vmw65 derived sequences in this case could be so extensively digested while the sal fragment was resistant (A purified Protein A-sal fusion was also shown to be resistant to thrombin treatment, data not shown). A similar explanation could be offered for the thrombin susceptiblity of GST-Vmw65 as observed in the bacterium, that misfolding is causing more thrombin sensitive sites to be exposed.

Since the e47-fragment represented an internal section of Vmw65 sequence, the polypeptide could have been excised and expressed out-of-context of its normal environment (perhaps it was a sequence that spanned an internal region of Vmw65).

Mutational studies indirectly ascribe functions to various regions of Vmw65; however, the only available means to study their significance in the transcription complex has been to perform a gel retardation (or bandshift) assay. In its simplest form, the assay requires a radiolabeled DNA sequence containing the IE promoter, TAATGARAT, the cellular octamer binding protein, oct-1, and other nuclear factors supplied by a nuclear extract. Run on a non-denaturing polyacrylamide gel, any complexes formed will be seen on the autoradiogram as being retarded in mobility from the naked DNA probe. Any differences in the mobility of a complex is a function of its size though a quantitative determination cannot be made; hence, determination of the substituents in the complex is reliant upon systematic reconstitution and substitution. One disadvantage of this technique is that not all of the factors that contribute to a VIC (Vmw65 Induced Complex) are known though fortunately, complex formation correlates with transcriptional activation (Haigh et al., 1990; O'Hare et al., 1988; Stern et al., 1989).

Gel retardation analysis of the three constructs that had been partially purified by affinity chromatography was performed with/without nuclear extract added to the reaction (Figure B3). In the prescence of nuclear extract, a weak Vmw65 Induced Complex (VIC) was observed for the Protein A-sal control. None of the GST fusions including the GST-sal analog of the Protein A control produced a VIC. However, when the the GST carrier was removed from the GST-sal fusion protein, the liberated 412 residue product did form a VIC. In the absence of nuclear extract, binding of the Vmw65 fusions to the naked template

Figure B3: Gel Retardation Analysis of GST-Vmw65 Constructs. A ^{32}P labelled DNA probe containing a TAATGARAT promoter sequence was incubated without (Panel A) and with (Panel B) HeLa nuclear extract and various fusion proteins that had been paritially purified by affinity chromatography. Above each lane is the fusion's designation: GST-sal (Δ1-412); GST-Vmw65 (1-490); GST-e47 (Δ153-278); GST (glutathione S-transferase carrier); NE (nuclear extract); sal, purified Vmw65(Δ1-412) from ion exchange chromatography). Complexes TRF (TAATGARAT recognition factor) and VIC (Vmw65 induced complex) are denoted by arrows.


TRF

Protein A-sal
GST-sal
sal
GST-Vmw65
GST-e47
gst-ΔCT
GST
NE
Protein A-sal
GST-sal
sal
GST-Vmw65
gst-ΔCT
GST
GST-e47
A
B
VIC

was assessed. At higher protein concentrations, only the Protein A-sal fusion bandshifted though a GST-Vmw65 bandshift has been spuriously observed (P. Xiao, personal communication). Neither the GST-sal or liberated sal fragment would bandshift on the naked template unlike the results observed in the presence of nuclear extract.

Discussion:

The most startling result from the gel retardation analysis was a potential carrier effect between GST and Protein A. Though the two carrier proteins are approximately the same molecular weight, only the GST carrier inhibited VIC formation perhaps by steric effects or by interactions with the surface of Vmw65. This putative carrier effect could have been greatly supported by a similar observation in the absence of nuclear extract. T he liberated sal fragment may still bind DNA at higher protein concentrations or under refined conditions. To further explore a carrier effect, the sal fragment could be fused to other proteins such as maltose binding protein, a synthetic FLAG peptide or poly(his)$_6$ (consult Table 2.1 for others).

The gel retardation analysis also supports the dispensability of the AAD domain for VIC formation and DNA binding. Since AAD does not possess a defined structure, crystallization of Vmw65 may be facilitated by its removal. Such a construct was represented by pGEXsal. Unfortunately, the large scale production of GST-sal was hindered by degradation and instability of the partially purified protein. Due to the quantities of protein needed for a crystallation, these expression difficulties must be corrected before this avenue can be explored.

The fact that Vmw65 contacts the DNA in a highly specific manner implies the existence of a nucleic acid recognition motif other than an a helix-turn-helix or zinc finger (motifs reviewed by Harrison, 1991). This motif could be a discrete sequence or consist of several key residues throughout the polypeptide that together provide a specific set of DNA contacts. The HMG-box orginally discoverded in the high-mobility-group class of non-histone proteins (Reeves et al., 1991) and GA binding motif orginally discovered in the ets-1 oncogene are two examples of nucleic acid binding domains that remain to be strucuturally characterized. Recently, Kim (1992) reviewed the discovery of a new recognition motif termed the β-ribbon and emphasized that other motifs in the future will surely be discovered. Though the location and nature of the Vmw65 nucleic acid recognition sequence remains indeterminate, an extensive mutagenesis study may yet prevail.

Appendix C

```
/*  Jasco-to-CONTIN File Convertor for IBM-PC  */
/*  By: Logan Donaldson, Nov. 1991.            */

/*  Execute as:


    tocontin.exe |input.asc| [-c]|conc.factor| [-s]|split wavelength| [>]|output.txt


    This program accepts the input file at the command line a Jasco file in the 'ascii' form and converts
    it to a text file that CONTIN will accept on the VAX. The file must be in ascii form. If no output
    file is given, output will be redirected to [a.out] in the current directory. This program happens to
    be my first ever C program.

    The [-c] switch is to accomidate a change in protein concentration which is expressed a
    multiplication factor, of sorts. This switch is optional and will default to 1.00.

    The [-s] switch is to divide the input spectrum into two parts to adjust for the higher error
    observed at lower wavelengths (ca. 200 nm and below). CONTIN will process each part of the spectrum
    with bias towards the data with wavelenghts higher than the value given in the [-s] switch. This
    switch is optional and will default to no splitting of the spectrum. See the CONTIN manual for more
    info.
                                                                                    */
/* This was written in Borland Turbo C++ v5.0                                     */

/* Please refer to the following references for a detailed description of the
   CONTIN program for VAX.

   Provencher SW. 1982. CONTIN: A general purpose constrained regularization program for inverting noisy
       linear algebraic and integral equations. Computer Physics Communications. 27: 229-242.

   Provencher SW and J Glockner. 1981. Estimation of Globular Protein Secondary Structure from Circular
       Dichroism. Biochemistry 20: 33-37.  */

/*  If data translation is desired at the level of the mainframe, the FORTRAN program packaged in
    cdconvert.exe (P. Ala, Dep't of Biochemistry, McMaster U.) can be alternatively used.   */

/*  To pass data to the program contin.exe, the converted data file must set as the header file FOR005.
    Three files are output: one long and one short data file specificed in FOR006 and FOR007,
    respectively and one file suitable for plotting. For reliable results, the protein concentration
    (reflected in the input ellipticities) must be known within 3%.

/* ------------------------------------------------------------------------ */

/* Includes */

#include <stdio.h>
#include <string.h>
#include <stdlib.h>

/* Function Prototypes */

void inputLine(FILE *in);
char *midstr(char *str, int start, int size);
void waste(int n, FILE *in);

extern char line[80], str[80];


main(int argc, char*argv[])
    {
    char fileOption1, fileOption2, fileName[10];
    float concFactor, wasteFactor, step, ruser15, value;
```

```
int i, splitWavelength, start, end, points, remainingPoints,
   iuser14, iuser15, ruser14, ruser16, outputFlag;

if (argc <2 || argc >5) {
    printf("Error: Not enough arguments to run program \n");
    return(); }

FILE *in, *out;



if ((out = fopen(argv[1],"r")) == NULL) {
    printf("Error: Problem opening the input file \n");
    return(); }

/* process the switches */

for (i=1; i < argc; i++)   {
    fileOption1 = argv[i][0];
    fileOption2 = argv[i][1];

    if (fileOption1 == '~') {
        if fileOption2 = 'c'
            concFactor = atoi(midstr(argc[i],3,strlen(argc[i])));
        else if fileOption2 = 's'
            splitWavelength = atoi(midstr(argc[i],3,strlen(argc[i])));
        else
            printf("Error: Illegal switch character '%c' \n", fileOption2);
            fclose(in);      return();
        }
    else if (fileOption1 == '>')  {
        if (outputFlag)  {
            printf("Error: Output file declared already \n");
            fclose(in);  fclose(out);        return();
            }
        outputFlag = 1;

        if  ((out = fopen(argv[i], "w")) == NULL)  {
            printf("Error: Problem opening the output file \n");
            fclose(in);  fclose(out);        return();
            }
        }
    else  {
        printf("Error: Illegal switch passed to program \n");
        fclose(in);      fclose(out);        return();
        }
    }
}

if (!outputFlag)
    fopen("a.out", "w");  /* no check for a.out problems */

/* get the pertinent variables */

inputLine(in);
fileName = midstr(line,15,strlen(line)-15);
waste(8,in);

inputLine(in);   start = atoi(midstr(line,18,5));

inputLine(in);   end = atoi(midstr(line,18,5));

inputLine(in);   points = atoi(midstr(line,18,5));

inputLine(in);
inputLine(in);   step = atoi(midstr(line,18,5));

waste(14,in);
```

```
    /* this program has a data requirement of 0.2 nm increments */

    if (step > 0.2) {
        printf("Error: Step increment must be 0.2 nm or less \n");
        fclose(in);   fclose(out);
        return(); }

    printf("Processing the original Jasco file: %s\n",fileName);

    fprintf(in,"ORIGINAL FILE WAS %s\r\n",fileName);
    fprintf(in,"IFORMY\r\n");
    fprintf(in,"(7F9.0)\r\n");
    fprintf(in,"IWT\r\n");
    fprintf(in,"IUSER                    5\r\n");

    /* calculate the user constants */

    wasteFactor = (0.2 / step); /* will either be 0.1, 0.04, 0.01 nm */
    iuser15 = 0;

    if (!splitWavelength) {
        iuser15 = -1;
        iuser14 = (start - splitWavelength)*5 + 1 }

    ruser14 = 1;
    if (!concFactor)  ruser14 = concFactor;

    ruser15 = 0.03; /* 3% error in protein. conc. assumed */
    ruser16 = 500;   /* this never changes, either         */
    remainingPoints = (start - end)*5 + 1;

    /* print the rest of the header */

    fprintf(in,"IUSER    14%15.0d\r\n",iuser14);
    fprintf(in,"IUSER    15%15.0d\r\n",iuser15);
    fprintf(in,"RUSER    14%15.0d\r\n",ruser14);
    fprintf(in,"RUSER    15%15.0f\r\n",ruser15);
    fprintf(in,"RUSER    16%15.0d\r\n",ruser16);
    fprintf(in,"END\r\n");
    fprintf(in,"NSTEND %4d %15.0d %15.0d\r\n",remainingPoints,start,end);

    /* print out the data */

    for(i = 0; i<points; i+=7)  {
        j = ((points-i)>7) ? 7 : (points %% 7);
        while (j--)  {
            inputLine(in);
            value = atol(midstr(line,2,9));
            fprintf(in,"%9.0f",value);
            }
        fprintf(in,"/r/n");
        }

    fclose(in)   fclose(out);
    printf("File Conversion Complete\n");
    return();
    }

char *midstr(char *str, int start, int size)
    {
    char *pa;   pa = str[0];  /* my global string variable */

    pa += start;
    while (size--)  *pa++ = *str++;

    *pa = '\0'; return (pa);
    }
```

```
void inputLine(FILE *in)
    {
    char *pb;    pb = line[0];
    while ((*(pb++)=fgetc(in) != '/n');
    }

void waste(int n, FILE *in)
    {
    while (n--)    inputLine(in);
    }
```

References

Allison LA, M Moyle, M Shales and CJ Ingles. 1985. Extensive Homology Among the Largest Subunits of Eukaryotic and Prokaryotic RNA Polymerases. *Cell* 42: 599-610.

Arakawa T and SN Timasheff. 1985. Theory of Protein Solubility. *Meth. Enzymol.* 114: 49-76.

Aso T, HA Vasavada, T Kawguchi, FJ Germio, S Ganguly, S Kitajime, Y Yasukochi and SM Weissman. 1992. Cloning and Characterization of cDNA for the Large Subunit of the General Transcription Initiation Factor TFIIF. Abstract R-202. *J. Cell. Biol.* Supplement 16E. p. 138.

Auble DT and S Hahn. 1992. An ATP-Dependent Inhibitor of TFIID Binding to DNA. Abstract R-100. *J. Cell. Biol.* Supplement 16E, p.126.

Bachmann BJ. 1972. Pedigrees of some mutant strains of *E. coli* K-12. *Bacteriol. Rev.* 36: 525-557.

Baldwin JP. 1992. Protein-nucleic acid interactions in nucleosomes. *Curr. Opin. Struct. Biol.* 2:78-83.

Berger SL, WD Cress, A Cress, SJ Treizenberg and L Guarente. 1990. Selective Inhibition of Activated but Not Basal Transcription by the Acidic Activational Domain of VP16: Evidence for Transcriptional Adaptors. *Cell* 61: 1199-1208.

Berger SL, B Pina, N Silverman, G Marcus, J Agapite and L Guarante. 1992. Cloning of Adaptors Required for Transcriptional Activation Using Genetic Selection in *S. Cerevisiae. J. Cell. Biol.* Supplement 16E. p. 191.

Boucher W, ED Laue, S Campbell-Burk and PJ Domaille. 1992. Four-Dimensional Heteronuclear Triple Resonance NMR Methods for the Assignment of Backbone Nuclei in Proteins. *J. Am. Chem. Soc.* 114: 2262-2264.

Bowie JU, R Luthy and D Eisenberg. 1991. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. *Science* 253: 164-170.

Buratowski S and H Zhou. 1992. Trascription Factor IID Mutants Defective for Interaction with Transcription Factor IIA. *Science* 255: 1130-1132.

Campbell KP, DH MacLennan and AO Jorgensen. 1983. Staining of the Ca^{2+}-binding Proteins, Calsequestrin, Calmodulin, Troponin C, and S-100, with the Cationic Carbocyanine Dye "Stains-all". J. Biol. Chem. 258: 11267-11273.

Campbell MEM, JW Palfreyman and CM Preston. 1984. Identification of Herpes Simplex Virus DNA Sequences which Encode a trans-acting Polypeptide Responsible for Stimulation of Immediate Early Transcription. *J. Mol. Biol.* 180: 1-19.

Cantor CR and PR Schimmel. 1980. *Biophysical Chemistry: PartII: Techniques for the Study of Biological Structure and Function.* San Francisco: Freeman and Co. pp.409-433.

Capone JP and G Werstuck. 1990. Synthesis of the Herpes Simplex Virus type 1 trans-activator Vmw65 in Insect Cells Using a Baculovirus Vector. *Mol. Cell. Biochem.* 94: 45-52.

Cavallini B, I Faus, H Matthes, JM Chipoulet, B Winsor, JM Egly and P Chambon. 1989. Cloning of the Gene Encoding the Yeast Protein BTF1Y, Which Can Substiture for the Human TATA box-binding Factor. *Proc. Natl . Acad. Sci. USA* 86: 9803-9807.

Chang CT, Wu C-S C and JT Yang. 1978. Circular Dichroic Analysis of Protein Conformation: Inclusion of the β-Turns. *Anal. Biochem.* 91: 13-31.

Chang J-Y, SS Alkan, N Hilschmann and DG Braun. 1985. Thrombin specificity: Selective Cleavage of Antibody Light Chains at the Joints of Variable Joining Regions and Joining with Constant Regions. *Eur. J. Biochem.* 151: 225-230.

Chang J-Y. 1985. Thrombin specificity: Requirement for Apolar Amino Acids Adjacent to the Thrombin Cleavage Site of the Polypeptide Substrate. *Eur. J. Biochem.* 151: 217-224.

Chou PY and GD Fasman. 1978. Prediction of the Secondary Structure of Proteins from their Amino Acid Sequence. *Adv. Enzymol.*47: 45-147.

Cornette JL, KB Cease, H Margalit, JL Spouge, JA Berzofsky and C DeLisi. 1987. Hydrophobicity Scales and Computational Techniques for Detecting Amphipathic Structures in Proteins. *J. Mol. Biol.* 195: 659-685.

Courey AJ and R Tjian. 1988. Analysis of Sp1 *in vivo* reveals Multiple Transcriptional Domains, Including a Novel Glutamine-rich Activation Motif. *Cell* 55: 887-898.

Cousens DJ, R Greaves, CR Goding and P O'Hare. 1989. The C-terminal 79 Amino acids of the Herpes Simplex Virus Regulatory Protein, Vmw65, Efficiently Activate Transcription in Yeast and Mammalian Cells in Chimeric DNA-binding Proteins. *EMBO J.* 8: 2337-2342.

Cress A and SJ Treizenberg. 1991a. Nucleotide and Deduced Amino Acid Sequences of the Gene Encoding Virion Protein 16 of Herpes Simplex Virus Type 2. *Gene* 103: 235-238.

Cress WD and SJ Triezenberg. 1991b. Critical Structural Elements of the VP16 Transcriptional Activation Domain. *Science* 251: 87-90.

Dalrymple MA, DJ McGeoch, AJ Davison and CM Preston. 1985. DNA Sequence of the Herpes Simplex Virus Type 1 Gene Whose Product is Responsible for Transcriptional Activation of Immediate Early Promoters. *Nucl. Acids. Res.* 13: 7865.

Davidson I, S Chaudhary, L Tora, C Brou, J Wu, J White, JJ Hwang, L Shemshedini, H Gronemeyer, JM Egly and P Chambon. 1992. Identification of Transcriptional Intermediary Factors Mediating the Action of Several Trans-activator Proteins. Abstract R-002. *J. Cell. Biol.* Supplement 16E, p. 123.

Demczuk S, M Donovan, G Franklin and R Ohlsson. 1991. Order of Probe and Nuclear Protein Extract Addition can Determine Specificity of Protein-DNA Complexes in Tested Mobility Shift Assays. *Nucl. Acids. Res.* 19: 677-678.

Dewey TG. 1991. *Biophysical and Biochemical Aspects of Fluorescence Spectroscopy*. Plenum Press (New York).

Donaldson L and JP Capone. 1992. Purification and Characterization of the Carboxyl-terminal Transactivation Domain of Vmw65 From Herpes Simplex Virus Type 1. *J. Biol. Chem.* 267: 1411-1414.

Dynlacht BD, T Hoey and R Tjian. 1991. Isolation of Coactivators Associated with the TATA-Binding Protein That Mediate Transcriptional Activation. *Cell* 66:563-576.

Dyson HJ, M Rance, RA Houghten, PE Wright and RA Lerner. 1988. Folding of Immunogenic Peptide Fragments of Proteins in Water Solution II: The Nascent Helix. *J. Mol. Biol.* 201: 201-217.

Edwards AM, SA Darst, WJ Feaver, NE Thompson, RR Burgess and RD Kornberg. 1990. Purification and Lipid-layer Crystallization of Yeast RNA Polymerase II. *Proc. Nat'l. Acad. Sci. USA* 87: 2122-2126.

Eggertsson G and Soll. 1988. Transfer Ribonucleic Acid-Mediated Suppression of Termination Codons in *Escherichia coli. Microbiological Rev.* 52: 354-374.

Eisenberg D, E Schwarz, M Komaromy and R Wall. 1984. Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *J. Mol. Biol.* 179: 125-142.

Eisenberg D, RM Weiss and TC Terwilliger. 1982. The Helical Hydrophobic Moment: A Measure of the Amphiphilicity of a Helix. *Nature* 299: 371-374.

Fasman GD. 1990. The Prediction of the Secondary Structure of Proteins: Fact or fiction. *Current Science* 59: 839-845.

Fields S and SK Jang. 1990. Presence of a Potent Transcription Activating Sequence in the p53 Protein. *Science* 249:1046-1048.

Formosa T, J Barry, BM Alberts and J Greenblatt. 1991. Using Protein Affinity Chromatography to Probe Structures of Protein Machines. *Meth. Enzymol.* 208: 29-45.

Frankel AD and PS Kim. 1991. Modular Structure of Transcription Factors: Implications for Gene Regulation. *Cell* 65: 717-719.

Garnier J, DJ Osquthorpe and B Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular protein. *J. Mol. Biol.* 120: 97-120.

Gasch A, A Hoffmann, M Horikoshi, RG Roeder and N-H Chua. 1990. *Arabidopsis thaliana* contains two genes for TFIID. *Nature* 346: 390.

Gerster T and RG Roeder. 1988. A herpesvirus trans-activating protein interacts with transcription factor OTF-1 and other cellular proteins. *Proc. Natl. Acad. Sci. USA* 85: 6347-6351.

Gill G and R Tjian. 1991. A Highly Conserved Domain of TFIID Displays Species Specificity *in vivo*. *Cell* 65: 333-340.

Gill SC and PH von Hippel. 1989. Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* 182: 319-326.

Giniger E and M Ptashne. 1987. Transcription in Yeast Activated by a Putative Amphipathic α helix Linked to a DNA Binding Unit. Nature 330: 670-672.

Goding CR and P O'Hare. 1989. Herpes simplex virus Vmw65-octamer binding protein interaction: a paradigm for combinatorial control of transcription. *Virology* 173: 363-367.

Goto Y, LJ Calciano and AL Fink. 1990. Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. USA* 87: 573-577.

Greaves RF and P O'Hare. 1989. Separation of Requirements for Protein-DNA Complex Assembly from Those for Functional Activity in the Herpes Simplex Virus Regulatory Protein Vmw65. *J. Virol.* 63: 1641-1650.

Greaves RF and P O'Hare. 1990. Structural Requirements in the Herpes Simplex Virus Type 1 Transactivator Vmw65 for Interaction with the Cellular Octamer-Binding Protein and Target TAATGARAT Sequences. *J. Virol.* 2716-2724.

Green MR, JV Pastewka and AC Peacock. 1973. Differential Staining of Phosphoproteins on Polyacrylamide Gels with a Cationic Carbocyanine Dye. *Anal. Biochem.* 56: 43-51.

Greenblatt J. 1991. Roles of TFIID in Transcriptional Initiation by RNA Polymerase II. *Cell* 66:1067-1070.

Ha I, WS Lane and D Reinberg. 1991. Cloning of a Human Gene Encoding the General Transcription Factor IIB. *Nature* 352: 689-695.

Haigh A, R Greaves and P O'Hare. 1990. Interference with the Assembly of a Virus-host Transcription Complex by Peptide Competition. *Nature* 344: 257-259.

Herr W, RA Sturm, RG Clerc, LM Corcoran, D Baltimore, PA Sharp, HA Ingraham, MG Rosenfeld, MG Finney, M Ruvkun and HR Horvitz. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.* 2: 1513-1516.

Hoffmann A, E Sinn, T Yamamoto, J Wang, A Roy, M Horikoshi and RG Roeder. 1990. Highly Conserved Core Domain and Unique N Terminus with Presumptive Regulatory Motifs in a Human TATA Factor (TFIID). *Nature* 346: 397-390.

Hope IA and K Struhl. 1986. Functional Dissection of a Eukaryotic Transcriptional Activator Protein, GCN4 of Yeast. *Cell* 46: 885-894.

Horikoshi M, K Maguire, A Kralli, E Maldonado, D Reinberg and R Weinmann. 1991. Direct Interaction Between Adenovirus E1A Protein and the TATA Box Binding Transcription Factor IID. *Proc. Natl. Acad. Sci. USA* 88: 5124-5128.

Horikoshi M, C Bertuccioli, R Takada, J Wang, T Yamamoto and RG Roeder. 1992. Transcription Factor TFIID Induces DNA Bending upon Binding to the TATA Element. *Proc. Natl. Acad. Sci. USA* 89: 1060-1064.

Iino T, K Takeuchi, SH Nam, H Siomi, H Sabe, N Kabayashi and M Hatanaka. 1986. Structural analysis of p28 Adult T-cell Leukemia Associated Antigen. *J. Gen. Virol.* 67: 1373-1379.

Ingles CJ, M Shales, WD Cress, SJ Treizenberg and J Greenblatt. 1991. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature* 351: 588-590.

Johnson WC. 1988. Secondary Structure of Proteins Through Circular Dichroism Spectroscopy. *Ann. Rev. Biophys. Biophys. Chem.* 17: 145-166.

Kao CC, PM Lieberman, MC Schmidt, Q Zhou, R Pei and AJ Berk. 1990. Cloning of a Transcriptionally Active Human TATA Binding Factor. *Science* 248: 1646-1649.

Killeen M, B Coulombe and J Greenblatt. 1992. Recombinant TBP, Transcription Factor IIB, and RAP30 Are Sufficient for Promoter Recognition by Mammalian RNA Polymerase II. *J. Biol. Chem.* 267: 9463-9466.

Kim S-H. 1992. β Ribbon: A New DNA Recognition Motif. *Science* 255: 1217-1218.

Kleinschmidt JA, C Dingwall, G Maier and WW Franke. 1986. Molecular Characterization of a Karyophilic, Histone-binding Protein: cDNA Cloning, Amino Acid Sequence and Expression of Nuclear Protein N1/N2 of *Xenopus leavis*. *EMBO J.* 5: 3547-3552.

Knutson JR, DG Walbridge and L Brand. 1982. Decay-Associate Fluorescence Spectra and the Hetrogenous Emission of Alcohol Dehydrogenase. *Biochemistry* 21: 4671-4679.

Kohli J and H Grosjean. 1981. Usage of the Three Termination Codons: Compilation and Analysis of the Known Eukaryotic and Prokaryotic Translational Termination Sequences. *Mol. Gen. Genet.* 182: 430-439.

Kristie TM, JH LeBowitz and PA Sharp. 1989. The Octamer-binding Proteins Form Multi-protein-DNA Complexes with the HSV αTIF Regulatory Protein. *EMBO J.* 8: 4229-4238.

Kristie TM and PA Sharp. 1990. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV α-trans-activator protein. *Genes Devel.* 4: 2383-2396.

Kubota S and JT Yang. 1986. Conformation and Aggregation of Melittin: Effect of pH and Concentration of Sodium Dodecyl Sulfate. *Biopolymers* 25: 1493-1504.

Lacowicz JR. 1991. *Topics in Fluorescence Spectroscopy. Volume 3: Biochemical Applications.* Plenum Press (New York).

Laughton A. 1991. DNA Binding Specificity of Homeodomains. *Biochemistry* 30: 11357-11367.

Laurent TC and J Killander. 1964. A Theory on Gel Filtration and Its Experimental Verification. *J. Chromatography.* 14: 317-330.

Laybourn PJ and ME Dahmus. 1990. Phosphorylation of RNA Polymerase IIA Occurs Subsequent to Interaction with the Promoter and Before Initiation of Transcription. *J. Biol. Chem.* 265: 13165-13173.

Lederberg J and EL Tatum. 1946. Gene recombination in *E. coli. Nature* (London) 158:558.

Lee WS, CC Kao, GO Bryant, X Lin and AJ Berk. 1991. Adenovirus E1A Activation Domain Binds the Basic Repeat in the TATA Box Transcription Factor. *Cell* 67: 365-376.

Lehrman SR, JL Tuls and M Lund. 1990. Peptide α-Helicity in Aqueous Trifluoroethanol: Correlations with Predicted α-Helicity and the Secondary Structure of the Corresponding Regions of Bovine Growth Hormone. *Biochemistry* 29: 5590-5596.

Lemire BD, C Fankhauser, A Baker and G Schatz. 1989. The Mitochondrial Targeting Function of Randomly Generated Peptide Sequences Correlates with Predicted Helical Amphiphilicity. *J. Biol. Chem.* 264: 20206-20216.

Levine M and JL Manley. 1989. Transcriptional Repression of Eukaryotic Promoters. *Cell* 59: 405-408.

Lin Y-S, I Ha, E Maldonado, D Reinberg and MR Green. 1991. Binding of General Transcription Factor TFIIB to an Acidic Activating Region. *Nature* 353: 569-571.

Lin Y-S and MR Green. 1991. Mechanism of Action of an Acidic Transcriptional Activator In Vitro. *Cell* 64: 971-981.

Loret EP, E Vives, PS Ho, H Rochat, J Van Rietschoten and WC Johnson, Jr. 1991. Activating Region of HIV-1 Tat Protein: Vacuum UV Circular Dichroism and Energy Minimization. *Biochemistry* 30: 6013-6023.

Ma J and M Ptashne. 1987. Deletion Analysis of GAL4 Defines Two Transcriptional Activating Segments. *Cell* 48: 847-853.

MacArthur MW and JM Thornton. 1991. Infulence of Proline Residues on Protein Conformation. *J. Mol. Biol.* 218: 397-412.

Marczak JE and MC Brandriss. 1991. Analysis of Constitutive and Noninducible Mutations of the PUT3 Transcriptional Activator. *Mol. Cell. Biol.* 11: 2609-2619.

Martin KJ, JW Lillie and MR Green. 1990. Evidence for Interaction of Different Eukaryotic Transcriptional Activators with Distinct Cellular Targets. *Nature* 346: 147-152.

McKnight JLC, TM Kristie and B Roizman. 1987. Binding of the Virion Protein Mediating a Gene Induction in Herpes Simplex Virus 1-Infected Cell to its *cis* site Requires Cellular Proteins. *Proc. Natl. Acad. Sci. USA* 84: 7061-7065.

McPherson A. 1985. Use of Polyethylene Glycol in the Crystallization of Macromolecules. *Meth. Enzymol.* 114: 120-124.

Mermod N, EA O'Neill, TJ Kelly and R Tjian. 1989. The Proline Rich Transcriptional Activator of CTF/NF-1 is Distinct from the Replication and DNA Binding Domain. *Cell* 58: 741-753.

Mitchell PJ and R Tjian. 1989. Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins. *Science* 245: 371-378.

Nilsson B and L Abrahmsén. 1990. Fusions to Staphylococcal Protein A. *Meth. Enzymol.* 185: 144-160.

O'Hare P and G Williams. 1992. Structural Studies of the Acidic Transactivation Domain of the Vmw65 Protein of Herpes Simplex Virus Using ^{1}H NMR. *Biochem.istry* 31: 4150-4156.

Patel L, C Abate and T Curran. 1990. Altered Protein Conformation on DNA Binding by Fos and Jun. *Nature* 347: 572-575.

Peterson MG, N Tanese, BF Pugh and R Tjian. 1990. Functional Domains and Upstream Activation Properties of Cloned Human TATA Binding Protein. *Science* 249: 1625-1630.

Provencher SW. 1982. CONTIN: A General Purpose Constrained Regularization Program for Inverting Noisy Linear Algebraic and Integral Equations. *Comp. Phys. Commun.* 27: 229-242.

Provencher SW and J Glockner. 1981. Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* 20: 33-37.

Ptashne M and AAF Gaan. 1990. Activators and Targets. *Nature* 346: 329-331.

Pugh BF and R Tjian. 1990. Mechanism of Transcriptional Activation by Sp1: Evidence for Coactivators. *Cell* 61: 1187-1197.

Pugh BF and R Tjian. 1991. Transcription from a TATA-less Promoter Requires a Multiunit TFIID Complex. *Genes. Dev.* 5: 1935-1945.

Ranish JA, WS Lane and S Hahn. 1992. Isolation of Two Genes That Encode Subunits of the Yeast Transcription Factor IIA. *Science* 255: 1127-1129.

Reeves R, TA Langan and MS Nissen. 1991. Phosphorylation of the DNA-binding domain of nonhistone high-mobility group I protein by cdc2 kinase: Reduction of binding affinity. *Proc. Natl. Acad. Sci. USA* 88: 1671-1675.

Reinberg D, M Horikoshi and RG Roeder. 1987. Factors Involved in Specific Transcription in Mammalian RNA Polymerase II. *J. Biol. Chem.* 262: 3322-3330.

Roizmann B and D Spector. 1991. The Induction of a Genes by the α-Trans-inducing Factor in *Herpesvirus Transcription and its Regulation*, EK Wagner, ed. CRC Press (Boston).

Sadowski I, J Ma, S Treizenberg and M Ptashne. 1988. GAL4-VP16 is an Unusually Potent Transcriptional Activator. *Nature* 335: 563-564.

Sambrook J, EF Fritsch and T Maniatis. *Molecular Cloning: a Laboratory Manual, 2nd. ed.* 1989. Cold Spring Harbour Laboratory Press (Cold Spring Harbour).

Saudek V, HS Pasley, T Gibson, H Gausepohl, R Frank and A Pastore. 1991. Solution Structure of the Basic Region from the Transcriptional Activatior GCN4. *Biochemistry* 30: 1310-1317.

Sharp PA. 1991. TFIIB or not TFIIB? *Nature* 351: 16-18.

Sherwood R. 1991. Protein fusions: bioseparation and application. *TIBTECH* 9: 1-3.

Siegel LM and KJ Monty. 1966. Determination of Molecular Weights and Frictional Ratios of Proteins in Impure Systems by Use of Gel Filtration and Density Gradient Centrifugation. Application to Crude Preparations of Sulfite and Hydroxylamine Reductases. *Biochim. Biophys. Acta.* 112: 346-362.

Sigler PB. 1988. Acid Blobs and Negative Noodles. *Nature* 333: 210-214.

Smale ST and D Baltimore. 1989. The "Initiator" as a Transcriptional Control Element. *Cell* 57: 103-113.

Smith DB, KM Davern, PG Board, WU Tiu, EG Garcia and GF Mitchell. 1986. M_r 26000 Antigen of *Schistosoma japonicum* Recognized by Resistant WEHI 129/J Mice is a Pparasite Glutathione S-Transferase. 1986. *Proc. Natl. Acad. Sci. USA* 83: 8703-8707.

Smith DB and KS Johnson. 1988. Single-step Purification of Polypeptides Expressed in *Escherichia coli* as Fusions with Glutathione S-Transferase. *Gene* 67: 31-40.

Spector D, F Purves and B Roizman. 1991. Role of α-Transindusing Factor (VP16) in the Induction of a Genes within the Context of Viral Genomes. *J. Virol.* 65: 3504-3513.

Stern S, M Tanaka and W Herr. 1989. The Oct-1 Homeodomain Directs Formation of a Multiprotein-DNA Complex with the HSV Transactivator VP16. *Nature* 341: 624-630.

Stern S and W Herr. 1991. The Herpes Simplex Virus Transactivator VP16 Recognizes the oct-1 Homeodomain. Evidence for a Homeodomain Recognition Subdomain. *Genes. Dev.* 5: 2555-2566.

Stringer KF, CJ Ingles and J Greenblatt. 1990. Direct and Selective Binding of an Acidic Transcriptional Activation Domain to the TATA-box factor TFIID. *Nature* 345: 783-786.

Swergold GD and CS Rubin. 1983. High-Performance Gel-Permeation Chromatography of Polypeptides in a Volatile Solvent: Rapid Resolution and Molecular Weight Estimations of Proteins and Peptides on a Column of TSK-G3000-PW. *Anal. Biochem.* 131: 295-300.

Takano E, M Maki, H Mori, M Hatanaka, T Marti, K Titani, R Kannagi, T Ooi and T Murachi. 1988. Pig Heart Calpastatin: Identification of Repetitive Domain Structures and Anomalous Behavior in Polyacrylamide Gel Electrophoresis. *Biochemistry* 27: 1964-1972.

Takeda K and Y Moriyama. 1991. Helix Formation of Mellitin on Poly(L-glutamic acid) and Poly(D-glutamic acid). *J. Am. Chem. Soc.* 113: 1040-1041.

Tinoco I, K Sauer and JC Wang. 1978. *Physical Chemistry*. 2nd ed. Prentice-Hall (Englewood, NJ). p.237.

Travers AA. 1992. DNA Conformation and Configuration in Protein-DNA Complexes. *Curr. Opin. Struct. Biol.* 2:71-77.

Uhlén M and T Moks. 1990. Gene Fusions for Purpose of Expression: An Introduction. *Meth. Enzymol.* 185: 129-143.

Van Hoy M, A Hansen and T Kodadek. 1992. Spectroscopic Studies of a Transcriptional Activation Peptide. *J. Am. Chem. Soc.* 114: 362-363.

Weiss MA, T Effenberger, CR Wobbe, JP Lee, SC Harrison and K Struhl. 1990. Folding Transition in the DNA-binding Domain of GCN4 on Specific Binding to DNA. *Nature* 347: 575-578.

Werstuck G and JP Capone. 1989a. Identification of a Domain of the Herpes Simplex Virus trans-Activator Required for Protein -DNA Complex Formation throught the Use of Protein A Fusion Proteins. *J. Virol.* 63: 5509-5513.

Werstuck G and JP Capone. 1989b. Mutational Analysis of the Herpes Simplex Virus Trans-inducing Factor Vmw65. *Gene* 75: 213-224.

White J, C Brou, J Wu, Y Lutz, V Moncollin and P Chambon. 1992. The Acidic Transcriptional Activator GAL-VP16 Acts on Preformed Template-Committed Complexes. *EMBO J.* 11: 2229-2240.

Wishart DS, BD Sykes and FM Richards. 1992. The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secodnary Structure through NMR Spectroscopy. *Biochemistry* 31: 1647-1651.

Xiao P and JP Capone. 1990. A Cellular Factor Binds to the Herpes Simplex Virus Type-1 Transactivator Vmw65 and is Required for Vmw65-Dependent Protein-DNA Complex Assembly with Oct-1. *Mol. Cell. Biol.* 10: 4974-4977.

Yang JT, C-SC Wu and HM Martinez. 1984. Calculation of Protein Conformation from Circular Dichroism. *Meth. Enzymol.* 11: 208-269.

Yamamoto T, M Horikoshi, J Wang, S Hasegawa, PA Weil and RG Roeder. 1992. A Bipartite DNA Binding Domain Composed of Direct Repeats in the TATA box Binding Factor TFIID. *Proc. Natl. Acad. Sci. USA* 89: 2844-2848.

Zakut-Houri R, M Oren, B Bienz, V Lavie, S Hazum and D Givol. 1983. A Single Gene and a Pseudogene for the Cellular Tumour Antigen p53. *Nature* 306: 594-597.

Zamyatnin AA. 1972. *Prog. Biophys. Mol. Biol.* 24: 107-123.

Zawel L and RD Reinberg. 1992. Advances in RNA Polymerase II Transcription. *Curr. Biol.* 4: 488-495.

Zhang H, R Scholl, J Browse and C Somerville. 1988. Double Stranded DNA Sequencing as a Choice for DNA Sequencing. *Nucl. Acids Res.* 16: 1220.

Zhu Q, TF Smith, RH Lathrop and J Figge. 1990. Acid Helix-Turn Activator Motif. *PROTEINS: Structure, Function and Genetics* 8: 156-163.