The Dissertation Committee for Jennifer Beth Hurley Sadow
Certifies that this is the approved version of the following dissertation:

The X-Ray Crystal Structure of Wheat Translation Initiation Factor eIF4E

Committee:

Jon D. Robertus, Supervisor

Marvin Hackert

Karen Browning

Dean Appling

Arlen Johnson

The X-Ray Crystal Structure of Wheat Translation Initiation Factor eIF4E

by

Jennifer Beth Hurley Sadow, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2002

Dedication

To Jonathan – this one counts for both of us…

Acknowledgements

After all this time, it's hard to believe it's nearly over. I would like to acknowledge and thank my advisor, Dr. Jon Robertus, for supporting me and giving me a research home for the last five years. Also the members of my committee: Dr. Marvin Hackert, Dr. Karen Browning (a valued collaborator with much knowledge of wheat translation proteins), Dr. Dean Appling, and Dr. Arlen Johnson.

Many thanks to the Robertus lab. Bold and beautiful are we. To those who have gone before me: Jung-Keun Suh, Maria Svinth, Elisabeth Schelp, John Pascal, Darcie Miller, and Kara Bortone. To those who are left to continue the tradition: Maya Zhang, Wendi Wagner, Huda Suliman (who assisted with some of my crystal screens and protein preparations), and Warren Hoe. All of their help and insights have been invaluable, as well as their listening ears when the graduate student process became stressful. The staff and students of the Hackert lab have been wonderful neighbors: Don Carroll, Andrew Kern, Jeff Almrud, Dan Keller, and Pavel Golubkov. Then there is the Browning group, my "second lab", who have been very generous in sharing resources, equipment, advice, and food from their lab meetings: Sandra Lax, Betty Burks, Kelley Ruud, Anneke Metz, Laura Mayberry, Brandy Gazo, Trish Murphy, Lara Campbell, Mike Henderson, and Michael Hoffman, to name just a few.

Some of the laboratory staff members are of particular importance. Art Monzingo has been a great help on this project, always had a suggestion when I got stuck in the refinement process, and was always there to listen. Steve Ernst has patiently taken care of our computer system and responded immediately whenever something went wrong. Last but certainly not least, Margaret Rodgers has kept the lab running smoothly, kept us in line, and been there every time we needed her.

There are many people elsewhere who deserve a mention, particularly John Hart and Alex Taylor at UTSA, who took one of my crystals to Brookhaven and reduced the synchrotron data. Their colleague, Jennifer Stine, who suggested the cryoprotectant that worked so well for these crystals. Closer to home, my understanding husband, who followed me to Austin; also my parents. My father received the first Ph.D. in our family, inspiring me to go for the second. My mother was a graduate student at the same time I was, and attending her graduation was very inspiring.

Thanks to Rice University and Baylor College of Medicine, my past; Schreiner University, my future. Finally, a very special acknowledgement to my first research advisor, Dr. John L. Margrave at Rice. He gave me my start in the world of research and help at a time when I really needed it, and this support enabled me to get to the point where I am today.

The X-Ray Crystal Structure of Wheat Translation Initiation Factor eIF4E

Publication No. _____

Jennifer Beth Hurley Sadow, Ph.D.

The University of Texas at Austin, 2002

Supervisor: Jon D. Robertus

Proteins are the molecules that serve as the machinery of living organisms. The genetic code of life, stored in DNA, is transcribed to messenger RNA, which is exported from the nucleus to the cytoplasm of the cell. There, in the translation process, proteins and ribosomes assemble to form the active complex that synthesizes proteins from the mRNA template. Eukaryotic translation initiation factor 4E, or eIF4E, is a small protein that recognizes the "cap" structure at the 5' end of mRNA. The binding of the cap by eIF4E and the association of eIF4E with the protein eIF4G allow translation to begin.

Solving the three-dimensional structure of a protein gives valuable insight into its function. The mechanism of its activity can often be deduced from the

structure, which can be used to find possible targets for chemical intervention. Because eIF4E is essential for translation, knowledge of its structure is important and could lead to commercial uses: an inhibitor which affected the cap-binding function of eIF4E would be a powerful herbicide.

Here, the solved X-ray crystal structure of wheat eIF4E at 1.85 Angstroms is presented. As seen in the structures of other eIF4Es, the monomeric unit consists of an eight-stranded beta sheet, three alpha helices, and three large loop regions. Two of these loops contain tryptophans which have been observed, in previous structures, holding the cap in an aromatic stacking interaction. However, crystallized wheat eIF4E does not contain a bound cap molecule, despite having been co-purified with a cap analogue. Instead, two molecules of eIF4E form a dimer, the interface of which is formed by a beta sheet created from the loop regions of the monomer. Each molecule inserts one tryptophan into a hydrophobic pocket on the surface of the other molecule; this pocket is analogous to the cap-binding slot.

The wheat eIF4E structure contains an unusual disulfide bridge, found in each monomer, that may hold the structure in an "open" conformation. This would force an opening between the cap-binding tryptophans, allowing for release of the cap, as well as the dimerization of eIF4E molecules which was observed in this X-ray structure.

Table of Contents

List of Tables xii

List of Figures xiii

CHAPTER 1: INTRODUCTION **1**

Translation 1

Translation Initiation 4

Eukaryotic Translation Initiation Factor eIF4E 5

Previously Solved Structures of eIF4E 9

Associated Proteins 13

CHAPTER 2: MATERIALS AND METHODS **17**

Crystallography Theory 17

X-Ray Diffraction 17

Temperature Factors and Structure Factors 21

Electron Density and the Phase Problem 22

Crystal Symmetry and Spacegroups 23

Hanging Drop Method of Crystallization 26

X-Ray Data Collection 28

Cryo-Crystallography 29

Synchrotron Data Collection 30

Data Reduction 31

Molecular Replacement 32

Electron Density Maps 34

Molecular Model Building 35

Molecular Cloning of eIF4E 35

Molecular Cloning of eIF(iso)4E 37
Expression of eIF4E 39
Purification of eIF4E 39
Expression and Purification of eIF(iso)4E 40
Expression and Purification of eIF(iso)4G Fragment 41
Western Blot 42
Site-Directed Mutagenesis 42
Gel Filtration 44
Ellman Assay for Disulfide Bonds 45
Trypsin Digestion 46

CHAPTER 3: RESULTS **48**

Cloning Results 48
Protein Expression and Purification Results 49
Results for eIF4E 49
Results for eIF(iso)4E 51
Results for eIF(iso)4G Fragment 52
Crystallization Trials 55
Successful Crystal Growth 55
Difficulties in Obtaining Crystals 56
Trypsin Digest Results 57
X-Ray Crystal Structure of Wheat eIF4E 59
Gel Filtration Results 67
Ellman Assay Results 70

CHAPTER 4: DISCUSSION **71**

Structural Features and Comparison 71
Wheat and Mouse Structure Comparison 71
Cap Binding Tryptophans 72
Disulfide Bonds 78

Future Directions 83

Appendices 86

Appendix A: Media Recipes 86

Appendix B: Trypsin Digestion Results 87

Appendix C: Protein Sequence File 89

Appendix D: Sequence of eIF4E 90

Appendix E: Sequence of W62F Mutant 91

References: 93

Vita 98

List of Tables

Table 3.1: Crystal Data Table for Synchrotron Data (Wheat eIF4E) 65
Table 3.2: Table of gel filtration data. 69

List of Figures

Figure 1.1: The 7-methylguanosine "cap". 3
Figure 1.2: The translation initiation complex. 5
Figure 1.3: Alignment of representative eIF4E sequences. 8
Figure 1.4: The structure of murine eIF4E. 12
Figure 1.5: Cap binding by aromatic stacking. 13
Figure 1.6: Schematic alignment of eIF4G sequences. 16
Figure 2.2: Argand diagram. 20
Figure 2.3: The Ewald sphere. 20
Figure 2.4: Lattice planes. 25
Figure 2.5: A unit cell in a hexagonal lattice. 25
Figure 2.6: Sealed capillary tube with crystal mounted inside. 30
Figure 3.1: Purification of wheat eIF4E on a polyacrylamide gel. 50
Figure 3.2: Purification of mutant eIF4E W62F. 51
Figure 3.3: Purification of eIF(iso)4E. 52
Figure 3.4: Purification of eIF(iso)4G fragment. 54
Figure 3.5: Photograph of a Crystal of Wheat eIF4E. 56
Figure 3.6: Trypsin digest of eIF(iso)4G. 59
Figure 3.7: The solved structure of wheat eIF4E. 61
Figure 3.8: The dimer interface. 62
Figure 3.9: Representative electron density at Trp 62. 63
Figure 3.10: Positioning of key residues in wheat eIF4E structure. 64
Figure 3.11: Ramachandran plot. 66
Figure 3.12: Typical chromatogram of a gel filtration run. 68
Figure 3.13: Gel filtration calibration curve. 69
Figure 4.1: Comparison of mouse and wheat eIF4E. 72
Figure 4.2: Electrostatic surface of a monomer of wheat eIF4E. 75
Figure 4.3: Positions of purported cap-binding tryptophans. 76
Figure 4.4: Superposition of the two chains of wheat eIF4E. 77
Figure 4.5: Electron density at Trp 108. 78
Figure 4.6: Electron density at disulfide bond. 82
Figure 4.7: Positioning of cysteines in wheat eIF4E dimer. 83

CHAPTER 1: INTRODUCTION

TRANSLATION

The process of translation, or protein synthesis, bridges the gap between the genetic code stored in nucleic acids and the proteins that perform the essential functions of life. Genomic DNA is composed of four nucleotide bases, which occur in two strands that pair with each other by hydrogen bonding to form a double helix. RNA is synthesized from DNA, and differs in the number of -OH groups on the ribose unit and in the identity of one of its bases.

The first step in "reading" the DNA code, transcription, occurs when a single strand of RNA, known as messenger RNA or mRNA, is synthesized in the nucleus using one of the strands of genomic DNA as the template. The nascent mRNA, in the case of eukaryotes, is "immature" and requires further processing. Higher organisms contain sequences of noncoding DNA called introns; these are spliced out, and other necessary posttranscriptional modifications take place in the nucleus. These include the addition of a poly(A) tail to the 3' end and the addition of a 7-methylguanosine cap to the 5' end (Figure 1.1). The cap is attached to the next base by a triphosphate linkage, as opposed to the usual single phosphate linkage between bases. The fully processed mRNA contains only the codons needed to produce the protein of interest, as well as the 5' cap and 3' poly(A) tail.

The process of protein synthesis is known as translation and occurs in the cytosol. It requires the framework of a ribosome, a structure made of ribosomal RNA and proteins, which must first be assembled from its 40S and 60S subunits. In addition, transfer RNA (tRNA) molecules are required. These small, L-shaped RNA structures, different for each amino acid, recognize each codon by base-pairing to it. After an ATP-dependent activation step, they will carry the proper amino acids with them to add to the growing polypeptide chain that will make up the protein. Since the "start" codon encodes methionine, the first tRNA needed for the chain is a specific Met-tRNA, known as initiator met-tRNA.

The three stages of translation are initiation, elongation, and termination. All require the presence of unique protein "factors". The initiation step involves bringing the mRNA, the activated tRNA, and the ribosomal subunits together to commence translation. The elongation step allows for the continued "reading" of the mRNA transcript by tRNAs, as well as the lengthening of the growing polypeptide chain. Finally, termination occurs when a "stop" codon is encountered in the mRNA. The elongation phase ends and the complex dissociates, leaving a nascent protein to undergo further processing or to begin its work in the cell.

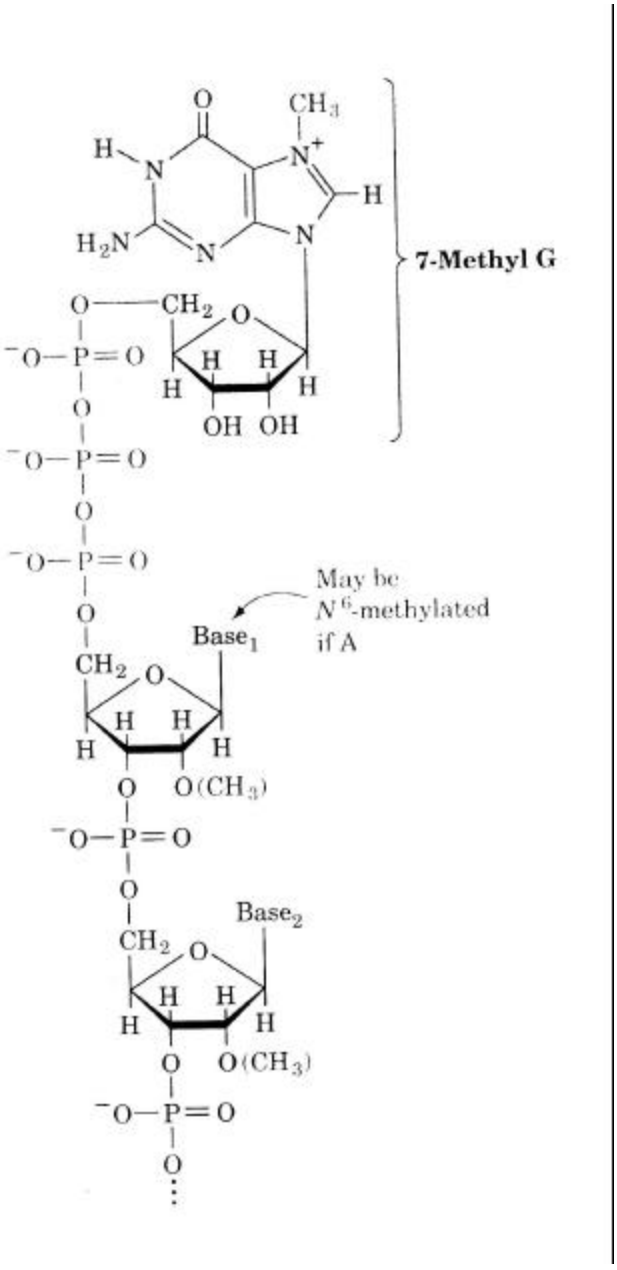


Figure 1.1: The 7-methylguanosine "cap" at the 5' end of a strand of messenger RNA. (From Voet and Voet, 1995).

TRANSLATION INITIATION

The initiation of protein synthesis in eukaryotes is a complex process involving multiple initiation factors, known as eIF proteins. The translation initiation process in higher plants is reviewed in Browning, 1996; the eIF4 factors are reviewed in depth in Gingras *et al.*, 1999. In the case of higher plants and other eukaryotes, eIF3 binds to the 40S subunit of the ribosome, and eIF2 forms a complex with the met-tRNA and GTP. When both of the above complexes associate with each other, possibly aided by another factor called eIF1A, a preliminary translation complex is formed.

At this point, the mRNA has not been recruited to the translational apparatus. Its specially methylated 5' cap structure is recognized by the eIF4F complex; more specifically, by eIF4E. The other subunit of the eIF4F complex, eIF4G, remains bound to eIF4E and also associates with eIF4A, which hydrolyzes ATP. eIF4A presumably works to unwind secondary structural elements that have formed in the mRNA, in order to facilitate translation (Balasta *et al*., 1993). PABP, or poly(A)-binding protein, recognizes the poly-A "tail" sequence at the 3' end of the mRNA. This protein has been shown to interact with eIF4G in yeast and mammals, as well as eIF4B in plants and mammals (Le *et al.*, 2000).

In the presence of mRNA, eIF4F will associate with the eIF3•40S•met-tRNA•eIF2•GTP complex (Carberry *et al*., 1991a). This preliminary translation initiation complex is pictured in Figure 1.2. The cap-binding protein eIF4E is thought to dissociate from the complex after this step, with the mRNA held in place by the other factors. At this stage, the 60S ribosomal subunit will be able to

bind to the complex. The ribosome and translational apparatus will then be fully assembled and elongation can begin.

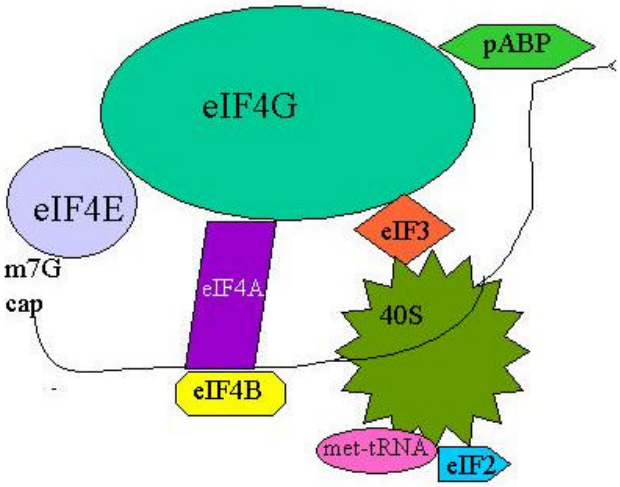


Figure 1.2: Representative figure of the translation initiation complex, with the mRNA shown as a schematic line. In plants, eIF4B can also interact with the poly(A)-binding protein.

EUKARYOTIC TRANSLATION INITIATION FACTOR EIF4E

The cDNA of wheat eIF4E was first isolated from a wheat sprout cDNA library by Anneke Metz in the laboratory of Dr. Joanne Ravel (Metz *et al.*, 1992),

and the amino acid sequence was determined there. The original purification from wheat germ is described in Lax *et al*., 1986. The form of eIF4E originally isolated from wheat germ has a molecular weight of 26 kDa; sometimes the protein is referred to as p26.

The cap-binding factor eIF4E is found in nature complexed with the larger factor eIF4G, and the two proteins together are known as eIF4F. eIF4G serves as a bridge between eIF4E and the rest of the translation initiation complex; while eIF4E only binds to the eIF4G and the cap, eIF4G associates with the unwinding factor eIF4A and other translational factors, such as PABP and eIF3, as well as stabilizing the mRNA in the complex.

The function of eIF4E involves the binding of the 7-methylguanosine cap structure found at the 5' end of mature mRNAs. This is a critical step in translation, essential for activity and thought to be the rate-limiting step (Sonenberg, 1996). An alignment of known eIF4E sequences is shown in Figure 1.3.

N-term of construct used →

```
                1              15 16              30 31              45 46              60 61              75 76              90
Wheat-p26       ---------------- --------MTEDTEM RPASAG--------- ---AEEREEGEIADD ------GDGSSAAAA GRITAHPLENAWTFW
Rice-p26        ---------------- --------MAEEHET RPPSAGRPPSSGRGR ADQADEREEGEIADD ------DSGHAPPQA NPAAPHPLEHAWTFW
Corn-p26        ---------------- --------MAEETDT RPASAG--------- SRGRPAPEDDDREEG ------EITDLACAP SPPATHPLEHSWTFW
Arabidopsis-p26 ---------------- MAVEDTPKSVVTEEA KPNSIENPIDRYHEE GDQAEEGEIAGGEGD ------GNVDESSKS GVPESHPLEHSWTFW

Wheat-p28       ---------------- -MAEVEAALPVAATE TPEVAAEG------- ---------------- ----------DAGAA EAKGPHKLQRQWTFW
Rice-p28        ---------------- --MAEVEAAPIAAAE TPEVAAA-------- ---------------- ----------EGAA AAKAPHKLHRQWAFW
Corn-p28        ---------MAEVEA PATAVEAPAAAVATT TPEVAAPE------- ---------------- ----------AGAAT EAKGPHKLHRQWTFW
Arabidopsis-p28 ---------------- --------MATDDVN EPLPAAAE------- ---------------- ----------LPATE AEKQPHKLERKWSFW

Mouse           ---------------- -----------MATV EPETTPTTNPPPAEE EKTESNQEVANPE-- ---------------- -HYIKHPLQNRWALW
Rat             ---------------- -----------MATV EPETTPTTNPPPAEE EKTESNQEVANPE-- ---------------- -HYIKHPLQNRWALW
Rabbit          ---------------- -----------MATV EPETTPTPNPPPAEE EKTESNQEVANPE-- ---------------- -HYIKHPLQNRWALW
Human           ---------------- -----------MATV EPETTPTPNPPTTEE EKTESNQEVANPE-- ---------------- -HYIKHPLQNRWALW

Xenopus         ---------------- -----------MAAV EPENT---NPQSTEE EKTETGQEIVSPD-- ---------------- -QYIKHPLQNRWALW
Drosophila-1    MQSDFHRMKNFANPK SMFKTSAPSTEQGRP EPPTSAAAPAEAKDV KPKEDPQETGEPAGN TATTTAPAGDDAVRT EHLYKHPLMNVWTLW
Drosophila-2    -----------MVVL ETEKTSAPSTEQGRP EPPTSAAAPAEAKDV KPKEDPQETGEPAGN TATTTAPAGDDAVRT EHLYKHPLMNVWTLW
S.pombe         ------------MQ TEQPPKESQTENTVS EPQEKALR------- ---------------- --------TVFDDKI NFNLKHPLARPWTLW

S.cerevisiae    ---------------- MSVEEVSKKFEENVS VDDTTATP------- ---------------- -------KTVLSDSA HFDVKHPLNTKWTLW
```

W62 **W**108 **C**113

|--Loop 1---------| |--Helix 1----| |–Loop 2-------| |--Helix 2-

```
                91             105 106             120 121             135 136             150 151             165 166             180
Wheat-p26       FDNPQGKSRQVAWGS TIHPIHTFPSTVEDFW GLYNNIHNPSKLNVG ADFHCFKNKIEPKWE DPICANGG--KWTIS CGRG-K---SDTFWL
Rice-p26        FDNPQGKSKQATWGS SIRPIHTFPSTVEDFW SLYNNIHHPSKLVVG ADFHCFKNKIEPKWE DPICANGG--KWTFS CGRG-K---SDTMWL
Corn-p26        FDNPQSKSKQAAWGS SIRPIHTFPSTVEEFW GLYNNIHHPSKLIVG ADFHCFKNKIEPKWE DPICANGG--KWTIS CGRG-K---SDTFWL
Arabidopsis-p26 FDNPAVKSKQTSWGS SLRPVFTPSTVEEFW SLYNNMKHPSKLAHG ADFYCFKHIIEPKWE DPICANGG--KWIMT FPKE-K---SDKSWL

Wheat-p28       YDIQTKPKPGAAWGT SLKKGYTFDTVEEFW CLYDQIFRPSKLVGS ADFHLFKAGVEPKWE DPECANGG--KWIVI SSRKTN---LDTMWL
Rice-p28        YDIQSKPKPGAAWGT SLRKAYTFDTVEEFW GLYDQIFRPSKVTVN ADFHLFKVGVEPKWE DPECANGG--KWIVP CSRKTT---LENMWL
Corn-p28        YDIQTKTKSGAAWGT SLKKAYTFDTVEEFW SMYDQIFRPSKLSGN ADFHLFKAGVEPKWE DPECANGG--KWTVP CNRKAT---FETMWL
Arabidopsis-p28 FDNQSK--KGAAWGA SLRKAYTFDTVEDFW GLHETIFQTSKLTAN AEIHLFKAGVEPKWE DPECANGGKWIWVVT ANRKEA---LDKGWL

Mouse           FFKNDK---SKTWQA NLRLISKFDTVEDFW ALYNHIQLSSNLMPG CDYSLFKDGIEPMWE DEKNKRGG--RWLIT LNKQQRRSDLDRFWL
Rat             FFKNDK---SKTWQA NLRLISKFDTVEDFW ALYNHIQLSSNLMPG CDYSLFKDGIEPMWE DEKNKRGG--RWLIT LNKQQRRSDLDRFWL
Rabbit          FFKNDK---SKTWQA NLRLISKFDTVEDFW ALYNHIQLSSNLMPG CDYSLFKDGIEPMWE DEKNKRGG--RWLIT LNKQQRRSDLDRFWL
Human           FFKNDK---SKTWQA NLRLISKFDTVEDFW ALYNHIQLSSNLMPG CDYSLFKDGIEPMWE DEKNKRGG--RWLIT LNKQQRRSDLDRFWL

Xenopus         FFKNDK---SKTWQA NLRLISKFDTVEDFW ALYNHIQLSSNLMSG CDYSLFKDGIEPMWE DEKNKRGG--RWLIT LNKQQRRNDLDRFWL
Drosophila-1    YLENDR---SKSWED MQNEITSFDTVEDFW SLYNHIKPPSEIKLG SDYSLFKKNIRPMWE DAANKQGG--RWVIT LNKSSK-TDLDNLWL
Drosophila-2    YLENDR---SKSWED MQNEITSFDTVEDFW SLYNHIKPPSEIKLG SDYSLFKKNIRPMWE DAANKQGG--RWVIT LNKSSK-TDLDNLWL
S.pombe         PLMPPTP--GLEWNE LQKNIITFNSVEEFW GIHNNINPASSLPIK SDYSFFREGVRPEWE DVHNKTGG--KWAFQ NKGRGG-NALDEMWL

S.cerevisiae    YTKPAVD-KSESWSD LLRPVTSFQTVEEFW AIIQNIPEPHELPLK SDYHVFRNDVRPEWE DEANAKGG--KWSFQ LRGKGAD--IDELWL
```

Figure 1.3: Alignment of representative eIF4E sequences obtained from GenBank. The isoforms of plant eIF4E are referred to as p26 (eIF4E) and p28 (eIF(iso)4E), respectively. Mammalian eIF4Es are nearly identical. Residues W62, W108, and C113 are noted because of their important roles in the solved structures. The large loops and helices of wheat eIF4E are labeled; the remainder of the structure consists of a beta sheet and turns. The asterisk indicates the position of a serine which may be phosphorylated and dephosphorylated as a means of translational control, although this mechanism is only known in mammals and the serine is not present in p26 forms of plant eIF4E. (Alignment performed with ClustalW at the Baylor College of Medicine Search Launcher: http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html.)

PREVIOUSLY SOLVED STRUCTURES OF EIF4E

Three structures of eIF4Es in different organisms have already been solved. The structure of yeast eIF4E was solved by NMR (Matsuo *et al*., 1997), and the murine eIF4E structure was solved by X-ray crystallography (Marcotrigiano *et al*., 1997). These structures can be found in the RCSB Protein Data Bank (http://www.rcsb.org/pdb/) with ID numbers of 1AP8 and 1EJ1, respectively. The two structures, published around the same time, represented a novel protein fold consisting of a curved, eight-stranded beta sheet with three alpha helices spanning its convex surface. The first two of the alpha helices, indicated here as helices 1 and 2, have been shown to be the binding site for eIF4G and for 4E-binding protein, which inhibits translation by binding in the same slot as eIF4G normally would (Marcotrigiano *et al*., 1999). While the eIF4E-binding proteins are important regulators of mammalian translation, they are not known to exist in plants (K. Browning, personal communication).

The eight-stranded beta sheet forms the structural backbone of the molecule, and three of the loops between beta strand are long and flexible. The cap-binding slot lies between loop 1 and loop 2. A N-terminal region of about 38 residues, depending on the species, has no conserved sequence pattern and has been shown to be disordered in the NMR structure (Matsuo *et al*., 1997). N-terminal truncation mutants have been shown to retain full activity (Vasilescu *et al*., 1996), and the murine structure was solved using a cloned form of eIF4E with the N-terminal 27 residues deleted (Marcotrigiano *et al*., 1997).

The structure of mouse eIF4E is shown in Figure 1.4. It is represented as a crystallographic dimer; however, the monomeric unit is the one believed to function in translation. The dimerization, which occurs on the helical face of the protein around the eIF4G binding site, is an artifact of crystal packing. A cap analogue, 7-methylguanosine diphosphate (m^7GDP), was used in isolation and purification of the protein and remained bound to the structure in its crystal form. The last flexible loop, loop 3, was not implicated in cap binding and was disordered in the structure; four residues had to be omitted from the structure due to lack of electron density. A serine residue located on this loop has been implicated as a phosphorylation site. Phosphorylation of this residue is thought to be a means for regulating translation in mammals, though this has not been observed in plants. Wheat eIF4E does not have a serine in this position, though its isoform does.

A unique feature of eIF4E is the set of eight tryptophans that are conserved among all known species. Two of these tryptophans, located on Loops 1 and 2, bind the cap by providing π–π aromatic stacking interactions with the 7-methylguanosine, forming a three-layer sandwich with the cap in the middle. (Figure 1.5) The eIF4E structures shown thus far have been monomeric and have included the cap. Previous efforts to purify and crystallize various forms of eIF4E without the cap have been unsuccessful.

The cap-binding tryptophans, numbered as Trp 62 and Trp 108 in the wheat eIF4E sequence, are responsible for the aromatic stacking and for interactions with the methyl group of the cap which strengthen the binding of the

cap in preference to any other guanosine group. Nearby residue Glu 109, conserved in all known species, also contacts the methyl group (Matsuo *et al.*, 1997). If this residue is mutated to an alanine, cap-binding function is lost (Morino *et al*., 1996). The rest of the cap is stabilized in the cap-binding slot by hydrogen bonding, van der Waals contacts, and charge-charge interactions; the latter are responsible for stabilizing the negative charges on the triphosphate linkage by lining them up along a positively charged cleft in the molecule.

A very recent publication introduces the structure of human eIF4E (Tomoo *et al*., 2002). When a m^7GTP cap analogue is bound, this structure is virtually identical to that of the mouse. However, when crystals were created with m^7G-PPP-A, loop 3, which had been disordered previously, became ordered and made contact with the adenosine. This suggests that RNA binding by eIF4E may not be limited to the cap alone, and this loop may have a function in holding the second base of the messenger RNA in place.

Figure 1.4: The structure of mouse eIF4E, in a crystallographic dimer as reported. One bound m^7GDP is shown in black. (Figure produced using MolScript (Kraulis, 1991).)

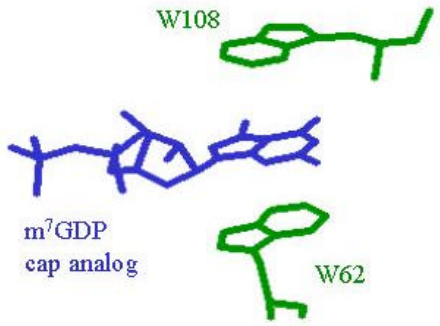


Figure 1.5: From the structure of murine eIF4E, the stacking interaction between the cap analog and two tryptophans. Tryptophans have been renumbered to show correlation with their equivalent residues in wheat eIF4E, as postulated from the sequence alignment. (Figure produced using RasMol, Sayle and Milner-White, 1995.)

ASSOCIATED PROTEINS

Two separate forms of eIF4E have been found in higher plants. The first one to be characterized is known as eIF4E. The second form, or isoform, is known as eIF(iso)4E. This protein, a little larger than eIF4E at 28kDa, is actually more abundant in wheat germ (Browning, 1996). It has 50% sequence homology to eIF4E, and the conserved sequences remain, including the cap-binding tryptophans.

The larger factor, eIF4G, also has an isoform, eIF(iso)4G. *In vivo*, the proteins are found as eIF4F complexes or eIF(iso)4F complexes. eIF4G is a roughly 160 kDa protein, similar to mammalian factor eIF4G; the isoform, eIF(iso)4G, has a molecular weight of only 86 kDa. An alignment of known eIF4G sequences (Figure 1.6) compares the domains from various species.

The two subunits of eIF(iso)4F are distinct from those of eIF4F. eIF4E can associate with eIF(iso)4G, and vice versa, *in vitro*; however, these associations do not appear to occur in nature (K. Browning, personal communication). Both forms are capable of cap binding and translation initiation (Sha *et al*., 1995; Van Heerden and Browning, 1994).

The reason for the presence of two factors is not entirely clear at this time. They may function in recognition of different types of mRNA. eIF4F is more efficient at translation of hairpin mRNA structures (Carberry and Goss, 1991b) as well as uncapped monocistronic or dicistronic mRNAs (Gallie and Browning, 2001), while eIF(iso)4F seems to prefer linear structures (Carberry and Goss, 1991b; Gallie and Browning, 2001) and recognizes hypermethylated caps more readily (Carberry *et al*., 1991). Direct fluorescence experiments with cap analogues indicate that the isoforms have different affinities for different analogues (Carberry and Goss, 1991b).

Just as the crystal structure of wheat eIF4E is of interest, so are the possible structures of its isoform, and of the bound form of eIF4E with eIF4G. The crystal structure at the interface was thought to be accomplished more easily with the isoform, eIF(iso)4F. Large multi-domain proteins such as eIF4G and

eIF(iso)4G are not likely to attain the compact globular shape needed for crystallization. If an eIF(iso)4E-binding domain could be isolated, a truncation mutant could be made. Since the N-terminal end had been determined to bind eIF(iso)4E (Metz and Browning, 1996), a trypsin digest of intact eIF(iso)4G was performed. The intention was to find domain breaks in the complete structure. This would help to isolate an eIF(iso)4G fragment that interacts with eIF(iso)4E and which might form a globular, crystallizable complex.

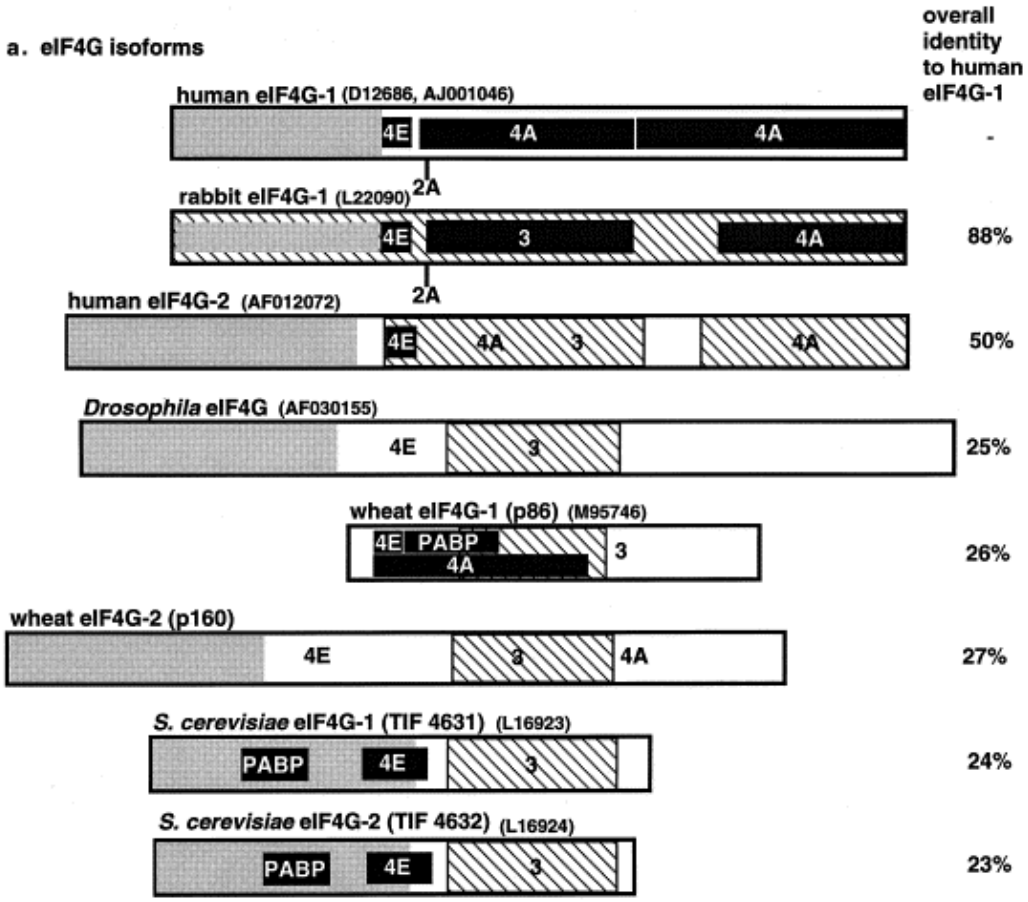


Figure 1.6: Schematic alignment of eIF4G sequences (for wheat proteins: p86=eIF(iso)4G, p160=eIF4G). This figure compares corresponding binding domains from the respective sequences: PABP=Poly-A Binding Protein, 4E=eIF4E, 4A=eIF4A, 3=eIF3. Hatched boxes indicate regions of over 30% sequence homology to human eIF4G-1. Black boxes show regions where binding has been demonstrated, and gray regions have multiple PEST sites. GenBank accession numbers are included. (From Keiper *et al.*, 1999)

CHAPTER 2: MATERIALS AND METHODS

CRYSTALLOGRAPHY THEORY

(The following explanations of crystallographic theory contain general knowledge and information gathered from the texts of Drenth, 1994, Holmes and Blow, 1965, and McRee, 1999, which provide more detail. The Crystallography 101 website (http://www-structure.llnl.gov/Xray/101index.html) has also served as a useful reference.)

X-Ray Diffraction

When an incident beam of X-rays encounters the electrons in a protein crystal, the waves are scattered in all directions by each electron in the crystal. This radiation travels as a cosine function. For each electron, the incident wave vector can be represented as $\mathbf{s_o}$ and diffraction in a particular direction by the vector **s**. (all vectors are denoted here by boldface type) The difference in direction, **S**, is represented as $\mathbf{S}=\mathbf{s}-\mathbf{s_o}$ (Figure 2.1).

In order to use the information provided by the X-rays, it is necessary to be able to add the many waves together which result from X-ray diffraction in a macromolecular crystal. An electromagnetic wave can be described by the equation $A\cos(\omega t + \alpha)$, in which A is equal to the amplitude of the wave, ω represents $2\pi\nu$ (where ν is the frequency, or the inverse of the wavelength), t stands for time, and α denotes the shift in the phase of the wave.

This wave can be divided into its real component, with amplitude $A\cos\alpha$ and phase angle of 0°, and its imaginary component, with amplitude of $A\sin\alpha$ and phase angle of 90°. These waves can be represented as vectors on an Argand diagram (Figure 2.2), with the amplitude of the real component on the real axis, the amplitude of the imaginary component on the imaginary axis, and the angle α. Provided that the wavelength of the radiation is the same, multiple waves can be added by summation of these vectors given in the Argand diagram.

In a crystal with atoms at unit cell positions $\mathbf{r}_j$ and crystal translation vectors **a**, **b**, and **c**, a crystal will only scatter X-rays that interfere constructively if $\mathbf{a}\cdot\mathbf{S}$, $\mathbf{b}\cdot\mathbf{S}$, and $\mathbf{c}\cdot\mathbf{S}$ are equal to integers. These integers are known as h, k, and l, respectively. These are taken as data points when the crystal data is processed, and are referred to as "reflections", because they can be taken as the reflections of the diffracted X-rays against a series of planes perpendicular to **S**. Using d to denote the distance between planes, and θ as the angle between the beam and the plane, Bragg's Law states that $\lambda=2d\sin\theta$. The reflections are considered lattice "points", and the points from the theoretical series of planes scatter X-rays in phase with each other.

The atomic scattering factor, f, describes the scattering of X-rays by the electrons in an atom; it is the Fourier transform of a single atom. Values of f have a Gaussian distribution and are dependent on the angle of the beam as a function of $2\sin\theta/\lambda$. In practice, these values are obtained from tables for each type of atom.

If the axes of the real-space unit cell of the crystal are denoted by **a**, **b**, and **c**, the directional components of **S** are called **a***, **b***, and **c***, where **S** = h·**a*** + k·**b*** + l·**c***. The lattice formed by **a***, **b***, and **c*** is called the reciprocal lattice. This lattice is used in conjunction with another theoretical construct, the Ewald sphere, to predict the direction of the diffraction maximum. This hollow sphere has its center at the point of diffraction, a radius of $1/\lambda$, and intersects with the reciprocal lattice as shown in Figure 2.3. In reciprocal space, the vectors **S**(hkl) have their endpoints on the Ewald sphere. The reciprocal lattice moves with the crystal. When the crystal, and reciprocal lattice, move such that the reciprocal lattice points intersect the Ewald sphere, observable diffraction conditions will occur.

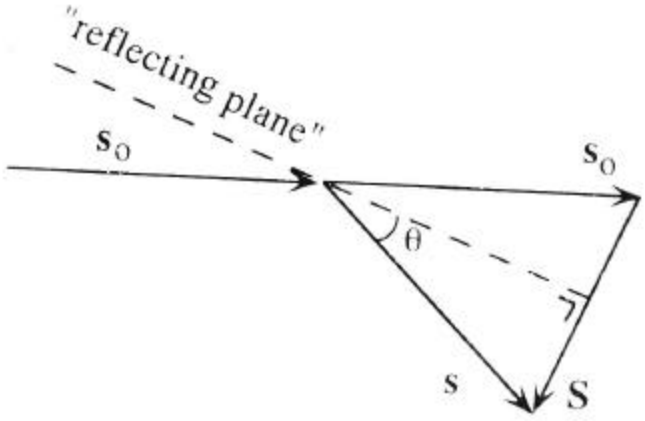


Figure 2.1: Illustration of the incident wave vector s_o and its relationship to **S** and θ. (From Drenth, 1994.)

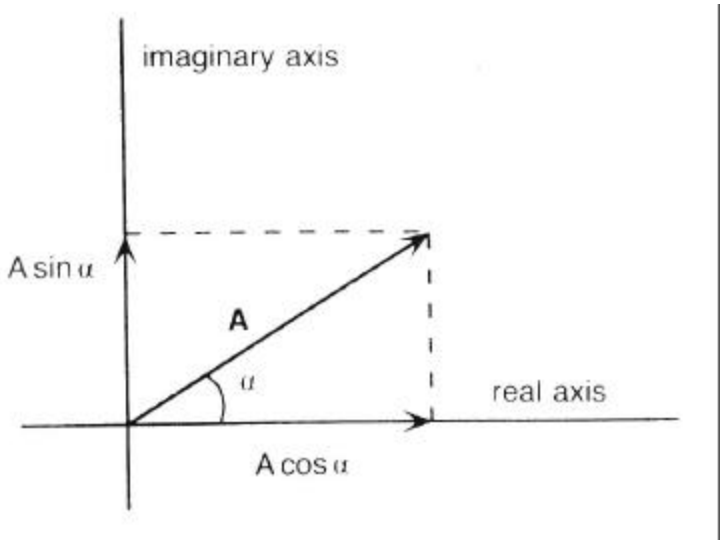


Figure 2.2: Argand diagram showing real and imaginary components of a wave of amplitude A. (From Drenth, 1994.)

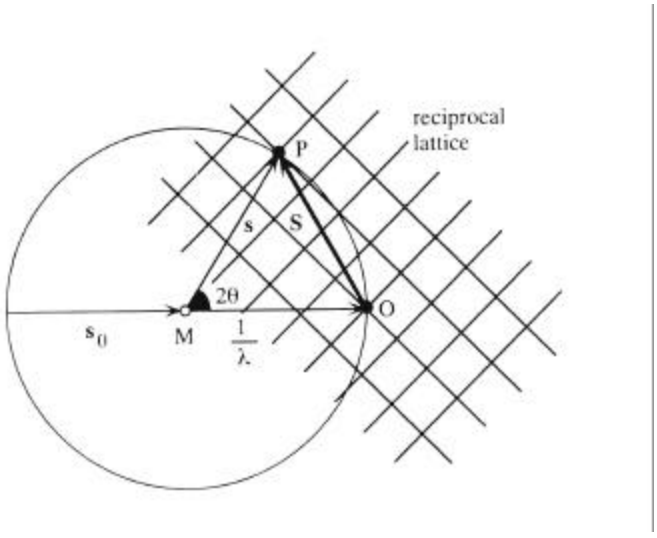


Figure 2.3: The Ewald sphere. (From Drenth, 1994.)

Temperature Factors and Structure Factors

The scattering of X-rays by a crystal is also dependent upon the "temperature", or mobility of the atoms, which in turn is dependent upon the position of the atoms in the structure. Therefore, each atom in the crystal has not only three spatial components, x, y, and z, but a temperature factor, B. It is related to the local atomic vibration by

$$B = 8\pi^2 \times |u^2|$$

where u is the mean square displacement of the atom from its center position. Values of u in fractional coordinates are less than 1, so most temperature factors in ordered structures are between 10 and 30 $Å^2$. The Fourier transform is weighted according to the B factor of that atom; more mobile atoms with higher B factors are downweighted so that their contribution to the scattering is decreased.

The above equation is valid only for the isotropic temperature factor, in which the vibration of the atom is the same in all directions. Anisotropic B factors require an ellipsoidal u^2 with separate components along the **a***, **b***, and **c*** axes.

The overall structure factor, **F**, is given by

$$\mathbf{F(S)} = \sum_j f_j \exp[2\pi i \mathbf{r}_j \cdot \mathbf{S}]$$

where the summation is over all atoms *j*, the vector **S** is that described previously, and **r** is the position of the atom in the unit cell.

Electron Density and the Phase Problem

The structure factor has amplitude F, described in the equation

$$F(hkl) = \Sigma f_j \exp[2\pi \cdot \mathbf{i} (hx + ky + lz)]$$

The intensities of the diffracted beams, I(hkl), can be obtained experimentally and are proportional to the square of F. Corrections for polarization, absorption, and the Lorentz correction are factored in when calculating the structure factors from the intensities. This amplitude does not include phase information.

The overall structure factor **F**(hkl) is related to the electron density by Fourier transform; also, the electron density ρ(xyz) is the Fourier transform of **F**.

As diffraction from a crystal can only be observed as the summation of waves in certain directions, the sum of the electron density can be written as a function of its Fourier transform. Over the volume V of the unit cell:

$$\rho(x\ y\ z) = (1/V)\sum_{h\ k\ l}\sum\sum|F(h\ k\ l)|\exp[-2\pi i(hx + ky + lz) + i\alpha(h\ k\ l)]$$

The intensities of h, k, and l, measured from the diffraction pattern, can be used to determine |F(h k l)|. However, this does not give the complete structure factor **F** because the phases, α(h k l), cannot be calculated. These must be determined in one of three ways: the use of data from heavy-atom derivatives of the crystal, the use of data from one heavy atom derivative at multiple

wavelengths (multiwavelength anomalous dispersion), or through molecular replacement. The latter can be used if the general form of the structure is already known, based on previously solved structures.

CRYSTAL SYMMETRY AND SPACEGROUPS

When a protein molecule is packed into a crystalline array, there are 65 possible arrangements it can take, called spacegroups. These depend upon the "shape" of the unit cell (defined as the smallest repeating unit in a crystalline array) and on the symmetry elements that exist in the crystal. Typically, rotational axes and screw axes are allowed. Other commonly known elements of symmetry, such as mirror planes and centers of inversion, are not seen in protein crystals because they would require inversion of some of the chiral configurations of amino acids.

The crystals of wheat eIF4E were found to belong to the spacegroup $P6_1$. The P signifies a primitive unit cell, with one protein molecule (or group of molecules) per unit cell. The 6 stands for a 6-fold axis of symmetry, with the subscript $_1$ signifying that this is a screw axis. The presence of this particular screw axis means that each repeating unit is also translated upward by 1/6 of a unit cell length.

The asymmetric unit is the repeating unit that defines the crystal; it may be a single protein or include multiple molecules related by non-crystallographic operations. The unit cells stack to form a lattice as pictured in Figure 2.4. This figure also shows the relationship between the Miller indices h, k, and l, which

determine the lattice planes that can be drawn based on the lattice points of the cell. The term d is used to describe the distance between lattice planes.

Each crystal unit cell must belong to one of the seven Bravais lattices (triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, or cubic). The crystals of wheat eIF4E were found to be hexagonal, with the lengths of the **a** and **b** edges of the unit cell equal to one another but not to the **c** vector, which was longer. (Figure 2.5.) The angles were equal to α=90°, β=90°, and γ=120°. The preceding satisfied the criteria for a hexagonal crystal; in this particular case, values of **a** and **b** were 66-67 Å depending on the particular crystal, and values of **c** were between 180 and 182 Å.

Based on the diffraction pattern, which is shown as a spot profile, computer programs such as XDISPLAY and DENZO from the HKL suite (Otwinowski and Minor, 1997) can deduce the most likely spacegroup. For example, 6-fold symmetry in the crystal translates to 6-fold symmetry in the spot profile, as well as a mirror plane in the pattern perpendicular to the axis (h k 0). The hexagonal lattice can also be found in trigonal crystals, but is distinguished by the fact that **c** is not equal to **a** and **b**. Systematic absences, or extinctions, along a major reciprocal axis generally signify a screw axis. When only every sixth spot is visible, this indicates there is a screw axis with translation of 1/6 of a unit cell. This can be $P6_1$ or $P6_5$ depending upon the direction of the screw rotation. The experimenter cannot determine which one of the two it is from the diffraction pattern alone; this can be deduced at a later stage (for example, the search for a translation solution).

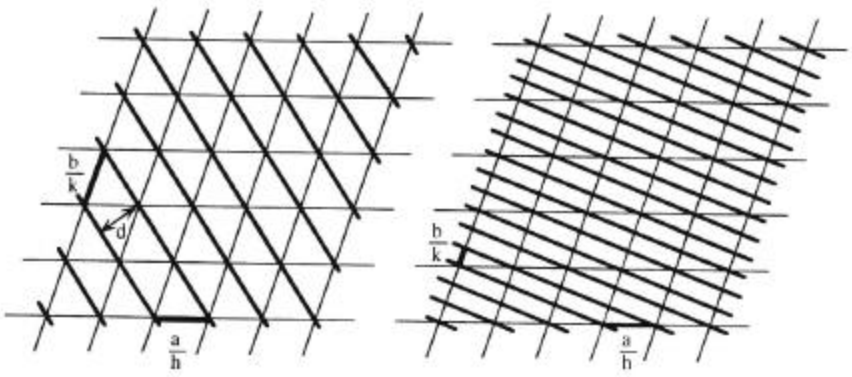


Figure 2.4: Lattice planes. (From Drenth, 1994.)

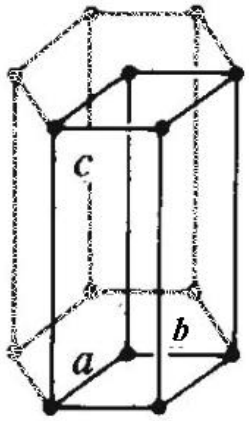


Figure 2.5: A unit cell in a hexagonal lattice. The bold lines indicate a unit cell; along with two other unit cells, it forms a hexagonal prism in the lattice (shown by dashed lines). Vectors **a**, **b**, and **c** are indicated here, and the lattice points are shown as dots. (Adapted from Barrett *et al.*, 1973.)

HANGING DROP METHOD OF CRYSTALLIZATION

Crystals of wheat eIF4E were successfully grown by the hanging drop method. In this method, the concentrated protein solution is equilibrated, usually at a 1:1 ratio, with the chosen crystallization reagent. This mixture is placed on a glass microscope cover slide as a droplet containing a few microliters of liquid. The slide is then turned upside down and sealed over a well containing a much larger volume of the chosen crystallization solution.

As time passes, an equilibrium is reached between the solution in the droplet and the solution in the well. The droplet is usually composed of 50% concentrated protein in its buffer (which is mostly water) and 50% crystallization reagent. The solution in the well is 100% crystallization reagent. This difference in chemical potential drives the water from the protein buffer solution to move in the vapor phase toward the well, causing gradual evaporation in the droplet. As the droplet becomes smaller, and therefore more concentrated with protein and crystallization reagent molecules, the protein molecules gradually move into a crystalline array.

This method, of course, does not work for all possible crystallization conditions. In the case of any protein that has not been previously crystallized, a screening process is used and hanging drops are set up with various concentrations of protein and many different crystallization reagents. While each crystallization reagent usually contains a buffering agent, a precipitant, and sometimes other stabilizing additives, there is no known pattern at this time which might help in predicting which reagent would crystallize which protein. Usually

this is accomplished first through trial and error, using known combinations of buffers and precipitants until one yields small crystals. If necessary, the conditions can then be modified in order to obtain crystals that can be used for diffraction.

Crystals formed by the hanging drop method can appear in a few hours to a few weeks after the crystallization plate is set up. In the case of wheat eIF4E, a few crystals appeared in a matter of days, but most of them took several weeks or longer.

The parameters that affect crystallization of proteins include buffer and precipitant conditions, pH, temperature, and the concentration of the protein. When the hanging drop is set up, the protein should be concentrated to at least 3 mg/mL if possible. Concentration of the wheat eIF4E protein was accomplished with an Amicon concentrator; this protein, like other eIF4Es (Marcotrigiano *et al.*, 1997; Matsuo *et al.*, 1997), was not extremely soluble at high concentrations. Most preparations could only be concentrated to about 1.5 mg/mL before the protein precipitated out of solution.

All crystals of eIF4E were grown at 4°C, with a crystallization reagent of 1.8-2.4 M ammonium sulfate in 0.1 M HEPES or Tris buffer, pH 7.5-8.0. The first crystal used for data collection was grown using the Hampton Research Crystal Screen II kit; however, it took over a year to appear and the exact contents of the well were unknown. The second crystal which proved useful was grown in 0.1 M Tris, pH 8.4, with 1.8 M ammonium sulfate, and the third crystal, which

was taken to the synchrotron, was grown in 0.1 M HEPES, pH 7.5, with 2.4 M ammonium sulfate.

X-RAY DATA COLLECTION

Once the size of a protein crystal has reached 0.1-0.2 mm in all directions, the crystal can be used for X-ray data collection. The crystal is carefully removed from the mother liquor and is mounted on a goniometer head, which is placed on an area detector. The crystal is then centered in the path of the X-ray beam by adjusting the position of the goniometer head.

Inside the X-ray tube used for diffraction experiments, electrons are emitted from a cathode. These reach a metal anode at an accelerated speed and cause electron transitions within the inner orbitals of the metal atoms in the anode. These transitions cause the emission of radiation at a specific wavelength. X-rays for this experiment were emitted from a copper anode, resulting in 1.54 Å radiation. This experiment utilized a rotating anode in order to allow higher intensity radiation and prevent stress on any one part of the anode.

After radiation from X-rays has been scattered by a crystal, it must be collected and recorded on an image plate. Most modern experiments utilize area detectors, which have an electronic image plate and are able to process the signal immediately. The detector used for the wheat eIF4E crystals was a Rigaku RU200 rotating anode (Molecular Science Corporation, The Woodlands, TX) with a Raxis IV image plate, which was used to detect and store the diffraction pattern. The computer program STRATEGY (Raimond Ravelli, Utrecht

University) was used to determine the oscillation range which would allow for the most data to be collected from each individual crystal.

Cryo-Crystallography

Crystal data can be collected either at room temperature or in a cold stream from liquid nitrogen. Room temperature data collection requires mounting the crystal in the center of a thin glass capillary tube (Figure 2.6). The advantage of this method is that the crystal can be mounted as is; the disadvantage is that bombardment with X-rays over a period of about 20 hours will often cause radiation damage to the crystal. Data collection in a cold stream is preferred if possible; in this case, mounting is accomplished easily by freezing the crystal and surrounding solution inside a "cryoloop". The crystal will suffer little decay or radiation damage while protected by the cold stream. To use this method, the crystal must be dipped in a cryoprotectant solution first to prevent ice from forming during freezing. A variety of cryoprotectants have been shown to be useful in the laboratory, but each crystal is different in terms of which cryoprotectants will allow for a good diffraction pattern.

The initial crystals of wheat eIF4E were capillary mounted at room temperature, as none of the cryoprotectants tested at that time allowed for a good diffraction pattern to be observed. Some of the crystals did decay under the X-ray beam, although not the ones used to determine the eIF4E structure.

Crystals mounted at a synchrotron facility require a cold stream and cryoprotectant, because the X-ray beam is extremely intense. In this case, the

crystal was dipped into a saturated solution of sorbitol in artificial mother liquor and frozen in liquid nitrogen. This proved to be a very effective cryoprotectant and improved diffraction on the area detector. The crystal was immediately transported to a synchrotron facility, where higher resolution data was obtained.

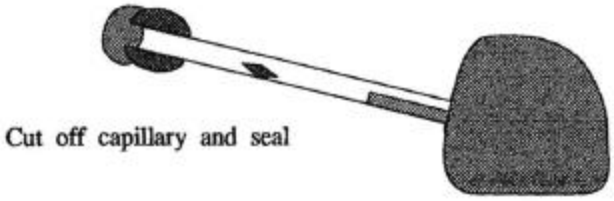


Figure 2.6: Sealed capillary tube with crystal mounted inside. Adapted from McRee, 1999.

SYNCHROTRON DATA COLLECTION

One crystal of wheat eIF4E was taken to the National Synchrotron Light Source at Brookhaven National Laboratories and exposed to synchrotron radiation. These X-rays have higher intensity than radiation from a conventional rotating anode; additionally, the tunability of synchrotron radiation allows a more preferable wavelength to be selected. The data was of a higher quality than that previously obtained. At this point, a molecular replacement model for wheat eIF4E already existed based on the solved murine eIF4E structure and crystal data from the area detector; however, the data were of relatively low resolution and large portions of the structure remained unsolved. When the synchrotron data

was obtained, it was used along with the previous molecular replacement model to find a new molecular replacement solution.

The program Arp_Warp (Lamzin and Wilson, 1993; Perrakis *et al.*, 1997) was used for the first round of refinement using the synchrotron data. This program performs iterations of structure factor refinements and subsequent rebuilding of the molecular replacement model. The output includes an updated model and a figure of merit map at atomic resolution. To ensure correctness of the model, all regions of questionable correctness were left out during Arp_Warp and rebuilt using the figure of merit map.

DATA REDUCTION

Raw X-ray diffraction data must be processed into a usable form. That is, background noise and questionable data points must be eliminated. This is done through programs such as the HKL suite (Otwinowski and Minor, 1997), which includes the data reduction program DENZO and scaling program SCALEPACK. These programs were used for all data sets collected from wheat eIF4E crystals.

DENZO fits the diffraction image to a spot profile, using the strongest reflections as a guide. This program accounts for polarization, Lorentz correction, and general error in the system.

The observed reflections are divided into shells based upon their resolution. Data from successively higher resolution shells is retained in the set until the ratio of intensity to error, I/σ, is less than 3.0 for 50% of the data points. Beyond this resolution, data are not reliable, and it may be too difficult to separate

the useful data from background noise. This sets the effective resolution limit for the structural solution.

Mosaicity, or the measure of defects in the crystal lattice, is one of the parameters estimated from DENZO. While the mosaicities for the crystals used in earlier stages of refinement were around 0.5°, the mosaicity found for the crystal taken to the synchrotron was only 0.2°. Lower values of mosaicity indicate a better ordered crystal.

After DENZO processes data into a useful form, the companion program SCALEPACK is used to scale the data. The resulting "hkl" file contains the crystal data used in the refinement process.

A "free R" flag can be added to 5% of the data points at this stage. These are randomly selected by a computer program and are flagged so as not to be used in refinement. When R_{free} is calculated, these data points are compared with the solution that has been found to the other data points. A free R that is close to the crystallographic R suggests a correct solution.

MOLECULAR REPLACEMENT

When a similar, or homologous, protein has already had its structure solved, molecular replacement is the method of choice for determining the phases of an unknown structure under study. The phases can be determined if the existing structure is used as a model and the proper spatial rotation and translation solutions can be found.

Many computer programs that carry out rotation and translation solution searches are available. The one used to fit the crystal data from wheat eIF4E to the murine model was X-PLOR (Brunger, 1992). Another program, EPMR (Kissinger *et al.*, 1999), was used to find the translation solution for the first crystal of eIF4E.

After initial rotation and translation solutions are found, these can be refined using the rigid-body search in the X-PLOR suite. Because there is only 40% sequence homology between mouse and wheat eIF4E, and some of the predicted loop and helical regions were of different lengths in the two models, a poly-alanine model of the mouse protein was used in the initial rounds of refinement. The rigid-body refinement was followed by the X-PLOR routines PREPSTAGE, SLOWCOOL, and FINALSTAGE, which perform energy minimization and spatial refinement of the model and carry out simulated annealing. These procedures were used in each round of model refinement. In later stages, the BREFI program was added to refine the individual temperature factors and the SLOWCOOL (simulated annealing) step was omitted. Once the structure has settled into a generally correct minimum, continued use of SLOWCOOL undoes more subtle model refinements.

The molecular replacement program CNSsolve (Brunger *et al.*, 1998) was used in place of X-PLOR for the final rounds of refinement. This program improved the energetics of the model, as well as the R-factor.

ELECTRON DENSITY MAPS

The types of electron density maps used to fit and refine the molecular replacement model included $2F_o$-F_c maps, F_o-F_c (difference) maps, and omit maps. Unless noted otherwise, all maps were calculated by X-PLOR and normalized using the RAVE program MAPMAN (Jones and Thirup, 1986).

In the $2F_o$-F_c maps, F_o is the observed diffraction amplitude and F_c is the calculated amplitude. This new, weighted amplitude is combined with the calculated phases from the model. Protein density is easily discernable from this model, but the danger of phase bias exists if the model phases are incorrect.

A difference map, calculated from F_o-F_c amplitudes, emphasizes the difference between calculated phases and what has actually been observed. This is used, at a high σ level, to show regions of density that are incorrect in the current model. If difference density appears around a part of the working structure model when the map is contoured, there is something wrong with that part of the model. Positive density suggests atoms need to be added, and negative density around the model suggests that atoms should be removed from that region.

Omit maps leave out selected parts of the model structure. These are used to eliminate phase bias for regions where the structure is not immediately evident, or where the model is suspect, like some of the loop regions in eIF4E.

MOLECULAR MODEL BUILDING

The computer program O (Jones *et al*., 1991) was used for the "physical" building and manipulation of the structure model. Initially, a poly-alanine model was substituted for the amino acid residues of the murine eIF4E structure. After each round of refinement with the wheat eIF4E crystal data, the model and an electron density map were visualized using O. The residues were moved into unoccupied density where there was an obvious fit, and the correct side chains were added and moved into position when the electron density became visible. The bond lengths and angles of the rebuilt regions were refined in O whenever the structure was manipulated. After every round of X-PLOR refinement, a round of rebuilding using O was performed.

MOLECULAR CLONING OF EIF4E

The cDNA for wheat eIF4E was obtained from a wheat sprout library by Anneke Metz, as described in Metz *et al*., 1992. The plasmid DNA was obtained from the laboratory of Dr. Karen Browning at the University of Texas at Austin.

The N-terminal 38 amino acids of wheat eIF4E do not contain any residues conserved between species. This region is shown to be disordered in the NMR structure of yeast eIF4E (Matsuo *et al*., 1997), and appears to be a "floppy tail" with no known function. The initial crystals of murine eIF4E grew as needles, and more favorable crystallization conditions were only found after the N-terminus was removed genetically (Marcotrigiano *et al*., 1997). The removed N-terminus did not include any conserved sequences.

In the case of wheat eIF4E, the first crystal discovered had come from untruncated protein, but took over a year to appear. After data collection and molecular replacement, the N-terminal residues were not visible in the electron density map; either they had been removed through proteolysis or they were completely disordered. For further crystallization, the N-terminal 38 residues were removed through cloning.

The original DNA had been inserted into a pET15b plasmid vector (Novagen). Truncation of the N-terminus was performed by Dr. Patricia Murphy in the Browning laboratory. Oligonucleotide primers designed to begin and end DNA synthesis at the desired points, allowing for excision and insertion into DNA via *Nco* I and *Bam* HI restriction sites, were custom-synthesized and added to a PCR cocktail reaction. The other components of the reaction were Pfu DNA polymerase (Stratagene), Pfu polymerase buffer, deoxynucleotide triphosphates (dNTPs), and the parent DNA. These were placed in a PCR thermalcycler with an automated program of: 5 minutes at 96°C, 35 amplification cycles (15 seconds denaturation at 94°C, 30 seconds annealing at 65°C, and 4 minutes extension at 72°C), and 4 additional minutes of extension time at 72°C at the end.

The amplified DNA fragments were purified and ligated into the vector pET15b by incubation at 14°C overnight with T4 DNA ligase. (Unless noted otherwise, all enzymes are from Fisher Promega or Life Technologies.) The vector had been digested with *Nco* I and *Bam* HI to accommodate the cohesive ends of the DNA insert. The products of the ligation reaction were transformed into DH5α *E. coli* cells on LB plates containing 100 µg/mL ampicillin. As the

plasmid contained a gene for ampicillin resistance, only those cells containing a plasmid could survive and form colonies on an ampicillin plate. The colonies were checked for the presence of insert in the plasmid. The successful cloning of wheat eIF4E, N-terminal truncation, was confirmed by DNA sequencing at the University of Texas DNA Core Facility.

MOLECULAR CLONING OF EIF(ISO)4E

Like eIF4E, eIF(iso)4E contains an N-terminal sequence that has no conserved residues or apparent function. Before any crystallization was attempted, the N-terminus was removed through cloning.

Custom oligonucleotide primers were designed to clone the isoform by removal of the first 24 residues. The eIF(iso)4E DNA was originally obtained from the Browning laboratory in the pET22b vector. The final PCR experiment contained 1.5 µL each of the two primers, 1.5 µL of 50 mM magnesium chloride, 1 µL of dNTP mix, 1 µL of the parent vector, 0.5 µL Taq polymerase and its associated buffer. The PCR mixture was put into a thermalcycler with a program of: 2 minutes at 94°C, 30 cycles of (1 minute extension at 94°C, 2 minutes annealing at 55°C, 3 minutes extension at 72°C), and 10 additional minutes at 72°C at the end.

The Taq polymerase used in the experiment leaves T-A overhangs on the sticky ends of the PCR product. The TA Cloning Kit (Invitrogen) was used to insert the product into the provided pCR2.1 vector, which contains

complementary overhangs. This was done by ligation with T4 DNA ligase overnight at 14°C at an 8:1 insert-to-vector ratio.

The ligation product was transformed into rubidium chloride competent DH5α cells by incubation for 1 hour on ice, heat shock for 45 seconds at 37°C, incubation on ice for 2 more minutes, the addition of 1 mL 2xYT media (see appendix), and subsequent incubation for 30 minutes in a 37°C water bath. The cells were plated out 200 μL at a time onto 2xYT-ampicillin plates and incubated at 37°C overnight to allow for growth of colonies containing the plasmid, which coded for ampicillin resistance. Colonies were picked and allowed to multiply in 5 mL of 2xYT media overnight. The plasmid DNA from the colonies was isolated using the Qiagen Miniprep Kit.

The oligonucleotide primers, obtained from the Browning laboratory, had been designed to encode restriction sites for *Nco* I and *Bam* HI. The presence of insert in the pCR vector was confirmed by digestion of the plasmid DNA with those enzymes.

Once the insert that coded for truncated eIF(iso)4E was successfully inserted into the pCR vector, the DNA-containing cells were grown in culture and the DNA purified as before. The insert-containing plasmid and the vector pET22b were both digested for 3 hours at 37°C with *Nco* I and *Bam* HI. The products were run out on an agarose gel, and the bands of interest were excised from the gel, crushed, and purified by centrifugation in a punctured tube with filter paper. The DNA was concentrated by ethanol precipitation and resuspended in 10 μL H_2O. These digested DNA fragments were combined at an insert:vector

ratio of 7:1 in a ligation reaction similar to that described previously. The presence of insert in the plasmids from transformation colonies was confirmed by plasmid preparation, digestion with *Nco* I and *Bam* HI, and agarose gel electrophoresis to check for the fragment. The sequence was confirmed by DNA sequence analysis.

EXPRESSION OF EIF4E

The N-terminal 38 amino acids of wheat eIF4E were removed by cloning as described above, with the gene cloned into the vector pET15b (Novagen). The plasmid DNA was introduced into BL21(DE3) *E. coli* cells by transformation. The cells were grown in 2xYT media at 37°C until the A_{595} reached 0.400. At that point, protein expression was induced by the addition of IPTG to a final concentration of 80 mg/L in the culture, and growth was allowed to continue for 2 hours at 30°C. The cells were pelleted and either frozen for future use or resuspended and lysed for protein purification.

PURIFICATION OF EIF4E

The cells from 1L of culture were pelleted and resuspended in 25 mL cap-binding protein buffer (20 mM HEPES, pH 7.0, 0.05 M potassium acetate, 10% glycerol, 1 mM DTT, and 0.1 mM EDTA). An alternative buffer that was used in some preparations contained 20 mM HEPES, pH 7.6, and used 0.1 M potassium chloride instead of potassium acetate; crystals of the same form were obtained from proteins prepared in both buffers. For protease inhibition, PMSF was added

to a total concentration of 24 mM; one mini-protease inhibitor tablet (Boehringer Mannheim) was also used. The cells were lysed by sonication and centrifuged in a Beckman L7 ultracentrifuge. In order to completely fill the ultracentrifuge tubes, the lysate was diluted to a volume of approximately 40 mL.

The supernatant was applied to a long, narrow 5-mL column of m^7GTP-sepharose (Pharmacia Biotech, 7-Methyl-GTP-Sepharose 4B) at a flowthrough rate of approximately 15 mL/hr. The column was flushed with cap-binding protein buffer until the A_{280} of the flowthrough fractions dropped down to a baseline level, usually overnight. Cap analogues, primarily m^7GTP, were used to elute the protein from the column. Approximately one milligram of cap analogue in 5 mL buffer was added directly to the top of the column, and fractions each representing 9 minutes of flowthrough were collected with a Gilson fraction collector. (Cap analogues other than m^7GTP, such as m^7GMP, have proven effective in the laboratory and were used to purify the protein, sometimes resulting in similar crystals.) The protein appeared pure after the initial cap-binding purification step as shown by SDS-polyacrylamide gel electrophoresis.

EXPRESSION AND PURIFICATION OF EIF(ISO)4E

eIF(iso)4E was expressed and purified in the same manner as eIF4E. It bound to the cap analog column and could be eluted from the column using cap analogs. The only difference was that the buffer at pH 7.0 had to be used; the full-length isoform protein had failed to bind the column at pH 7.6 (K. Browning, personal communication).

EXPRESSION AND PURIFICATION OF EIF(ISO)4G FRAGMENT

The DNA for the N-terminal fragment of eIF(iso)4G containing 186 residues (denoted as "C186") was obtained from the Browning laboratory. The clone was inserted into a pET vector, and this construct included a C-terminal sequence of six histidines (His-tag). With the His-tag fused to the protein, it could be purified through immobilized metal affinity chromatography (IMAC) on a nickel column.

The plasmid was transformed into BL21(DE3) E. coli cells and the protein expressed in a similar manner to the others described here. The cells were collected, resuspended in 15 mL nickel column buffer (20 mM Tris-Cl, pH 8.0, 0.5 M NaCl, 10% glycerol) with PMSF added as before, and lysed in a French press.

The nickel column was made from a matrix of His-Bind resin (Novagen). It was prepared for purification by charging with 100 mM $NiSO_4 \cdot 6H_2O$, which imparts a blue color to the matrix. The protein was loaded on the column, causing the blue color to disappear as it saturated the metal-binding sites. The column was run at 10-15 mL/hr, and was washed with buffer until the A_{280} of the eluate dropped to baseline levels.

Elution was accomplished with an imidazole gradient. Each stage of the gradient consisted of applying 5 mL of buffer containing a set amount of imidazole: 20 mM for the first fraction, then 40 mM, 60 mM, 80 mM, 100 mM, 200 mM, and finally 1 M imidazole to remove anything still bound to the column.

The nickel column regained its blue color as the metal-binding protein was stripped away.

WESTERN BLOT

A Western blot was performed to detect the presence of the purified C186 fragment of eIF(iso)4G. The intention was to see whether it would bind to eIF(iso)4E. A SDS-PAGE gel was run containing the proteins of interest, which were blotted onto a PVDF membrane using an EC600 blotter (E-C Apparatus Corporation, St. Petersburg, Florida). The membrane was incubated with a specific mouse monoclonal antibody which recognized His-tags (Novagen). After the binding of this antibody, a secondary antibody (rabbit anti-mouse) was added, which was hybridized to horseradish peroxidase (HRP). This complex was visualized with 4-chloro-1-naphthol (Pierce, Rockford, IL).

SITE-DIRECTED MUTAGENESIS

Studies done on mammalian eIF4E have shown that the first cap-binding tryptophan, denoted here as W62, can be mutated to other aromatic residues without loss of cap-binding activity (Hsu *et al.*, 2000). However, the second cap-binding tryptophan, W108, cannot be altered if the cap structure is to be bound. (Hsu *et al.*, 2000) In order to further study the purported cap-binding site, the mutant wheat eIF4E "W62F" was made, with the first cap-binding tryptophan changed to a phenylalanine. This construct, like the one used in the crystal structure, had the 38 N-terminal residues deleted.

Mutagenesis was accomplished by the method of the Stratagene QuikChange site-directed mutagenesis kit, though materials were obtained from other sources. Custom oligonucleotide primers were ordered from Invitrogen. These were mixed with the parent DNA, wheat eIF4E (N-terminal 38-residue deletion construct) in the vector pET 15b (Novagen), and Pfu Turbo DNA polymerase (Stratagene) with its buffer. The polymerase chain reaction, or PCR, was performed in a Perkin-Elmer GeneAmp 2400 PCR machine with 12 cycles each of: 30 seconds at 95°C (denaturation), 1 minute at 55°C (annealing), and 12 minutes at 68°C (extension). An additional 30 seconds at 95°C was added at the beginning of the cycle to allow for complete strand separation. This PCR reaction was sufficient to anneal primers and complete synthesis of the 5700-bp plasmid encoding mutant primers.

The mutant plasmid DNA produced in the above reaction was nicked circular. The PCR product was treated with the restriction endonuclease *Dpn* I in order to digest any methylated parent plasmids, sparing the unmethylated nicked circular PCR products. At this point, the plasmids were transformed into competent XL1-Blue cells by electroporation and grown in NZY^+ media (see appendix), then plated out on 2xYT-ampicillin plates. The bacterial machinery repaired the nicks in the circular DNA. Bacterial cells from the resulting colonies were grown in media, and the plasmids were extracted using the Qiagen Miniprep kit. The resulting plasmid DNA from each selected colony was sent for sequence analysis.

When the DNA from one of these plasmid preparations was confirmed as having the mutant sequence, the plasmid was transformed into rubidium chloride-competent BL21(DE3) *E. coli* cells. The resulting colonies were ready for expression. The expression and purification of the mutant protein were carried out in the same manner as described previously for wheat eIF4E.

GEL FILTRATION

Gel filtration chromatography is used to separate proteins in a solution by their molecular weights, without denaturing the proteins or separating their subunits. The column matrix consists of porous beads; a larger molecule, assuming it has a roughly globular shape, will migrate through the column at a faster rate than a smaller molecule. This is because the smaller molecule will fit through some of the pores in the beads and will be forced to slowly meander through them, while the larger molecule travels around the beads.

To perform gel filtration chromatography on wheat eIF4E, purified protein samples were concentrated to about 1 mg/mL using an Amicon concentrator with a Millipore ultrafiltration membrane which had a 10,000 Da molecular weight cutoff. The concentrated protein sample was dialyzed from the original elution buffer (pH 7.0) into the alternate buffer (pH 7.6), using a Spectra/Por dialysis membrane (Baxter) with a 6000-8000 Da membrane cutoff. All of the molecules in the solution could pass through this membrane except for the 21,000 Da protein. This process not only changed the buffering solution, but also removed most unbound cap analogue molecules from the protein solution. Glycerol was

added to the samples up to 50% in order to layer the protein on the surface of the column matrix. The protein samples were added one at a time to a 130 mL Sephacryl S-200 column (Pharmacia) that had been calibrated using blue dextran and protein standards with molecular weights ranging from 13.7-158 kDa. Absorbances were measured and recorded by an Isco UA-6 UV/VIS detector. Fractions were collected with an Isco Retriever 500 fraction collector.

ELLMAN ASSAY FOR DISULFIDE BONDS

To test for the presence of free thiols in wheat eIF4E, the Ellman assay was used (Ellman, 1959, Lane and Snell, 1976, and Riddles *et al*., 1983). The reaction, made up to a total volume of 500 µL, contained 4.5 nmol protein and 15 µL of 10 mM DTNB, where DTNB (obtained from Sigma) is 5,5'-dithiobis(2'-nitrobenzoic acid), also known as the Ellman reagent. All solutions were made in 0.1 M potassium phosphate buffer, pH 7.27, with a small amount of EDTA added. The protein was dialyzed exhaustively into this buffer to remove any significant amount of the DTT found in the protein preparation buffer. Dialysis also served to eliminate any unbound cap analogue molecules.

DTNB reacts with free thiols to form the product 2-nitro-5-thiobenzoic acid (2-TNB), which has an appreciable A_{412}. The molar extinction coefficient ε of 2-TNB is 14,150 $M^{-1}cm^{-1}$ at this wavelength in this buffer. Beer's Law states that $A=\varepsilon c l$; therefore, starting with a standardized cuvette length l of 1 cm and reading the absorbance A at 412 nm, the concentration c of 2-TNB can be calculated. When the reaction is performed with an excess of DTNB, the amount

of product will be directly related to the number of free thiols in the protein. (This assay may not work for some proteins whose thiols are inaccessible.)

This experiment was performed to confirm the presence of a disulfide bridge in wheat eIF4E which was observed by X-ray crystallography. The monomeric wheat eIF4E is known to have at least two free thiols; if the disulfide bond does not exist when the protein is free in solution, then four free thiols will be observed per molecule of protein and the amount of TNB formed will be four times the amount of protein in moles. If the disulfide bond does exist in solution, the amount of TNB formed will be twice the molar amount of protein.

TRYPSIN DIGESTION

The whole eIF(iso)4G protein was digested with a small amount of trypsin for a short period of time, to see where digestion would occur first. This would be likely to signal a linkage between domains, which would be more accessible to the protease. In the assay, 250 µL of 0.2 mg/mL protein, 50 µL trypsin digestion buffer (0.1 M Tris-Cl, pH 8.0, 10% acetonitrile, and 0.1 mg/mL calcium chloride), and 5 µL of 10 µg/mL trypsin (Bowman-Birk reagent, Sigma) were incubated at room temperature, and aliquots were taken at specific time intervals. After each aliquot was put into its own tube, further reaction was inhibited with 1 µL of 100 µg/mL soybean trypsin inhibitor (Sigma). The aliquots were electrophoresed on a SDS-PAGE gel.

When trypsin digestion conditions were found that would produce protein fragments of a suitable size, the assay was repeated using those digestion

conditions. Another gel was run and then stained and destained, in order to locate the bands. Afterward, the gel was soaked in Protein Elution Buffer (tris-glycine with 0.1% SDS) to make the proteins mobile again.

The gel was then electroblotted onto a PVDF membrane for 30 minutes using an EC600 blotter (E-C Apparatus Corporation, St. Petersburg, Florida). The transfer buffer used for blotting consisted of 10 mM CAPS in 15% methanol, pH 11.0. All solutions contained a pinch of DTT. The membrane was stained in membrane stain solution (0.1% Coomassie Blue G-250 in 50% methanol and 10% acetic acid) and then destained to make the protein bands visible. The bands were excised from the membrane, dried under a nitrogen stream, and submitted to the University of Texas Protein Facility for sequencing.

CHAPTER 3: RESULTS

CLONING RESULTS

The cloning of eIF4E DNA into pET15b was performed with the bases encoding the first 38 residues removed. The success of this cloning was confirmed by DNA sequencing (see Appendix D). The N-terminal truncation of eIF(iso)4E, removing 24 residues, was also successfully cloned, as proven by DNA sequencing.

After purification of this form of eIF(iso)4E, the protein was used for crystal trials, but these were unsuccessful. Subsequently, analysis of structural data from eIF4E suggested that the N-terminus could be removed at the 32nd residue instead of the 24th residue. This project was undertaken, and cloning attempts have thus far proven unsuccessful because of failure to ligate. In the future, a different combination of vectors and restriction sites could be used to continue this project.

The results of the trypsin digest and purification of the eIF(iso)4G C186 fragment suggested that it would be prudent to clone a truncation mutant with only the N-terminal 137 residues. Attempts to create this construct have been unsuccessful thus far. These have included glutathione S-transferase fusion vectors (Pharmacia Biotech) and the IMPACT system (New England BioLabs).

Site-directed mutagenesis of the N-terminal truncation of eIF4E was accomplished successfully, changing the codon for Trp 62 to one for Phe 62. This mutation was confirmed by sequencing (Appendix E).

PROTEIN EXPRESSION AND PURIFICATION RESULTS

Results for eIF4E

Recombinant eIF4E was purified as previously described. The protein was essentially pure after the initial affinity chromatography step as shown by SDS-PAGE (Figure 3.1). The molecular weight shown from the PAGE gels was about 21 kDa, which is consistent with the N-terminally truncated version of the protein that was being expressed and purified. An average yield from cells grown in 1 L of culture was about 3 mg of protein. This could be concentrated to between 1 and 1.5 mg/mL before it would precipitate out of solution.

Other cap analogues were also used for the successful purification of eIF4E from a m^7GTP-sepharose column: 7-methylguanosine diphosphate, 7-methylguanosine monophosphate, and 7-methylguanosine. These differed only in the length of their phosphate chains. According to the murine crystal structure, the phosphate chain interacts with a positively charged patch on the protein surface. While all of these were successful in eluting the protein from the column, phosphate chains of less than two phosphate units would bind progressively smaller amounts of protein per amount of cap analogue. The m^7G, with no phosphate, particularly bound only small amounts of protein; it was not able to compete sufficiently with the m^7G-sepharose column. Often more protein could be eluted once m^7GTP or m^7GDP was added.

The mutant form of the protein, W62F, was successfully purified in the same manner (Figure 3.2). The yield of protein was usually not as great, which may be due to lower cap affinity of this form, though all yields varied between different preparations. The final protein concentration tended to be around 0.6 mg/mL before the mutant form would precipitate out of solution.

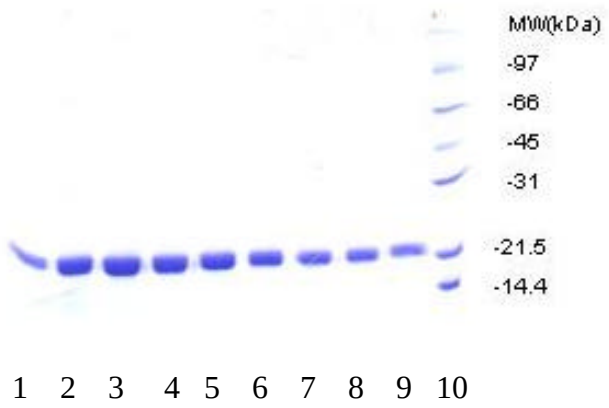


Figure 3.1: SDS-PAGE gel showing purification of wheat eIF4E (as cloned with N-terminal truncation) on m^7GTP-sepharose. Lanes 1-9: fractions of purified eIF4E eluted with m^7GTP. Lane 10: Low Molecular Weight Markers, Bio-Rad (97 kDa, 66 kDa, 45 kDa, 31 kDa, 21.5 kDa, 14.4 kDa).

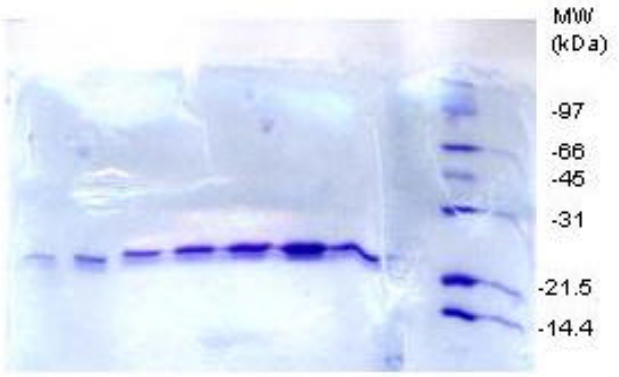


Figure 3.2: SDS-PAGE gel showing purification of the mutant eIF4E (N-terminal truncation plus mutation of Trp 62 to Phe 62) on a m^7GTP-sepharose column. Lanes 1-7: fractions of mutant eIF4E as eluted with m^7GTP. Lane 8: empty. Lanes 9-10: Low Molecular Weight Markers, Bio-Rad (97 kDa, 66 kDa, 45 kDa, 31 kDa, 21.5 kDa, 14.4 kDa).

Results for eIF(iso)4E

The N-terminal truncation of eIF(iso)4E was purified successfully by the same methods used for eIF4E. The protein bound to the m^7GTP-sepharose column and was eluted from it with m^7GTP; however, it was more difficult to purify than eIF4E, with impurities and/or degradation bands in some of the fractions (Figure 3.3). This protein could be concentrated to between 1.5-3 mg/mL. Like the other proteins here, different preparations looked alike but some were more successfully concentrated than others.

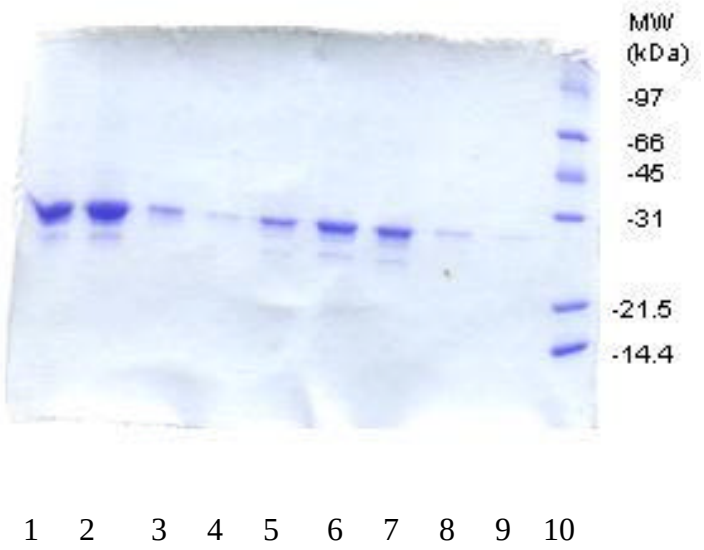


Figure 3.3: SDS-PAGE gel showing purification of the isoform eIF(iso)4E on a m^7GTP-sepharose column. This sample proved more difficult to purify than eIF4E, and showed some degradation. Lanes 1-9: fractions of eIF(iso)4E eluted with m^7GTP. Lane 10: Low Molecular Weight Markers, Bio-Rad (97 kDa, 66 kDa, 45 kDa, 31 kDa, 21.5 kDa, 14.4 kDa).

Results for eIF(iso)4G Fragment

The "C186" fragment of eIF(iso)4G, which had been cloned in the Browning laboratory and included the first 186 residues of the protein, was purified using a histidine tag and a nickel column. While the protein was successfully expressed and was mostly isolated from other proteins through an imidazole gradient, some impurities remained. There was not an extremely high

yield of protein, and it tended to degrade rapidly over time. Attempts were made to purify the protein further through other processes such as ion-exchange columns, but this led to complete loss of the protein.

A Western blot was performed on a sample of this protein, using an anti-His-tag antibody to detect the protein after blotting. A His-tagged construct of the yeast FMO protein, obtained from Dr. Jung-Keun Suh, was used as a control. When the blot was visualized, only the control was visible, showing that the rapid degradation of the C186 construct may have included degradation of the His-tag. It was determined that this construct was unstable and might not be suitable for further studies.

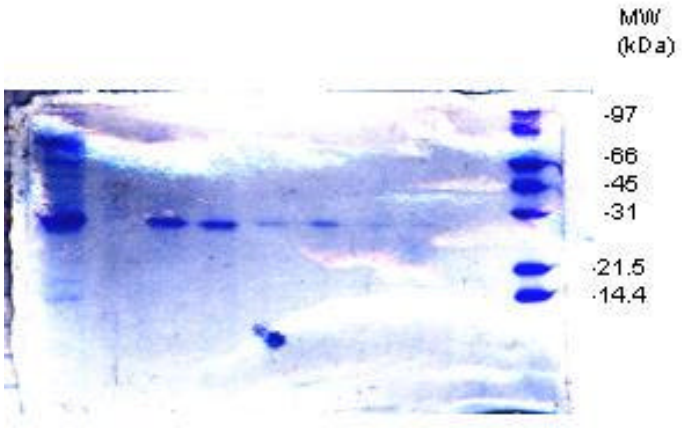


Figure 3.4: SDS-PAGE gel showing purification of a 186-residue N-terminal fragment of wheat eIF(iso)4G on a Ni^{2+}-NTA column. From left, Lane 1: cell lysate containing expressed protein. Lanes 2-9 elution fractions from imidazole gradient (Lane 2, 20 mM imidazole; Lane 3, 40 mM imidazole, Lane 4, 50 mM imidazole, Lane 5, 60 mM imidazole, Lane 6, 70 mM imidazole, Lane 7, 80 mM imidazole, Lane 8, 100 mM imidazole, Lane 9, 200 mM imidazole). Lane 10: Low Molecular Weight Markers, Bio-Rad (97 kDa, 66 kDa, 45 kDa, 31 kDa, 21.5 kDa, 14.4 kDa).

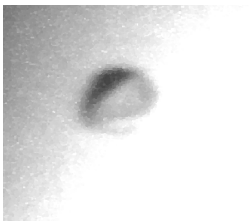
CRYSTALLIZATION TRIALS

Successful Crystal Growth

The crystals of wheat eIF4E used to solve the structure were relatively difficult to obtain. Initial trials by the Browning laboratory using the Crystal Screen kits (Hampton Research, Laguna Niguel, CA) yielded no crystals. After a waiting period of over a year, a single crystal was discovered in a crystallization plate made with Crystal Screen II. By this time, the plates had been in storage and the exact crystallization conditions were not known. Based on X-ray data collected from the crystal and a partial molecular replacement solution that fit the data, the N-terminus of the protein (38 residues) was disordered or had been removed by proteolysis. By this time, the structures of murine and yeast eIF4E had been published, and these showed that the N-terminus was disordered (Matsuo *et al*., 1997) and could be removed by cloning to make the protein more globular and easier to crystallize (Marcotrigiano *et al*., 1997). With the N-terminal 38 residues removed from the encoding DNA, the wheat eIF4E was crystallized in conditions of high ammonium sulfate and pH 7-8. The crystals were oval and had a general "football" shape. (Figure 3.5)

These crystals were grown in hanging drops at a 1:1 ratio of concentrated protein to well solution. As the protein would only concentrate to 1.5 mg/mL, and in some preparations not even that, it usually took months for crystals to form, though in a few cases they appeared in a matter of days or weeks. Most of these crystals were not large enough for data collection until two months or more

had passed, and many stayed small. Increasing the ratio of protein to well solution did not yield crystals. Disruption of the crystal plate may be a factor, as most crystals only appeared when the plate had gone unchecked for several weeks at a time. Often, little growth was observed in future measurements.



|-------| 0.16 mm

Figure 3.5: Photograph of a Crystal of Wheat eIF4E.

Difficulties in Obtaining Crystals

Even the most successful of the crystal conditions attempted yielded crystals less than 50% of the time. Some preparations of the protein never yielded a crystal, although they appeared sufficiently pure on a 12.5% SDS-PAGE gel. Some 15% gels, which give higher resolution for bands associated with small proteins, show an apparent degradation band at about 20 kDa. This was more visible in some preparations than others and is thought to be due to proteolysis, probably at the exposed C-terminal loop. Addition of extra protease inhibitors and running the purification column as rapidly as possible helped somewhat, but

degradation continued to be a problem and may have affected some crystallization trials.

Various other methods were attempted on the N-terminal truncated version of wheat eIF4E, without success. These included further trials with the Hampton Research kits, sitting drop crystallization, crystallization with an excess of cap analogue (50:1 cap to protein molecule ratio), and seeding a new crystallization drop with pieces of a crushed crystal.

Crystal screens were also set up with the N-terminal truncation of eIF(iso)4E, the C186 fragment of eIF(iso)4G, and a combination of both, using the Hampton Research kits. None of these trials led to crystallization.

All successful crystals were grown using eIF4E which had been eluted with one of the following cap analogues: m^7GTP, m^7GMP, and m^7G-PPP-G. These crystals all had the same "football" shape and were successfully grown in conditions of 1.8-2.4 M ammonium sulfate at pH 7-8.

TRYPSIN DIGEST RESULTS

After 15 minutes of digestion of eIF(iso)4G with a trypsin solution, three clear product bands of about 28-30 kDa were visible on an SDS-PAGE gel, with no other major bands. If these represented the three major domains of the protein, their molecular weights would add up to a total molecular weight of about 86 kDa, the molecular weight of the whole protein; of course, the gel bands could also represent or one or two of these domains plus degradation band(s). Protein

was blotted from the degradation bands onto a membrane and submitted for N-terminal sequence analysis.

The sequencing results, shown in Appendix B, indicate that proteolysis occurred at the 137^{th} and 555^{th} amino acids from the N-terminus of the protein. The molecular weights of the fragments were roughly consistent with the positioning of the sequences they corresponded to, and each of these sequences was preceded by an arginine residue, allowing for tryptic cleavage. One of the bands appeared to be a degradation product at the 137^{th} amino acid from the N-terminus, while two of them had sequences similar to that found at the 555^{th} amino acid. One of these two was probably a degradation band of the other. The N-terminal fragment was not found using this assay, and may have undergone considerable proteolysis during the first few minutes of the trypsin digest. This digest indicates two possible domain breaks in the protein.

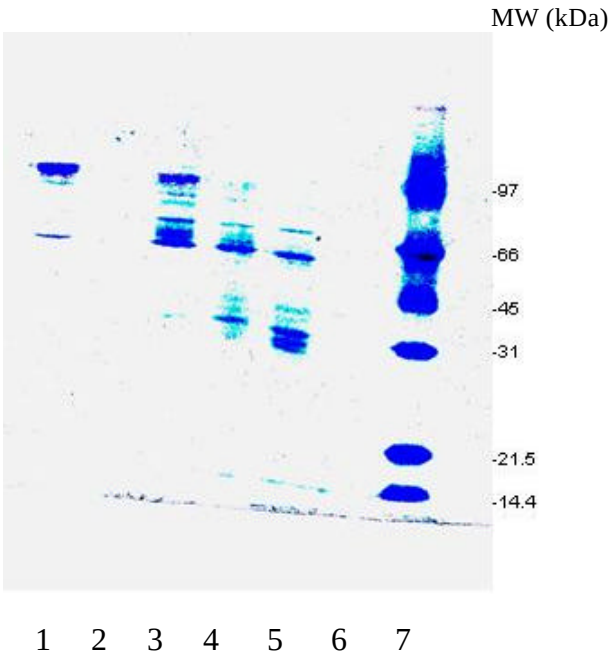


Figure 3.6: A SDS-PAGE gel showing the products of trypsin digestion of eIF(iso)4G. (Lane 1: initial protein sample, MW=86 kDa, showing some slight impurities. Lane 2: empty. Lane 3: digest after five minutes. Lane 4: digest after ten minutes. Lane 5: digest after fifteen minutes. Lane 6: empty. Lane 7: molecular weight ladder (from top: 97 kDa, 66 kDa, 45 kDa, 31 kDa, 21.5 kDa, 14.4 kDa).

X-RAY CRYSTAL STRUCTURE OF WHEAT EIF4E

The X-ray structure of wheat eIF4E was solved using molecular replacement methods. Data is given in the crystal data table (Table 3.1). The structure shows a dimer of two eIF4E units (Figures 3.7 and 3.8), each of which

have the eight-stranded beta sheet cupped hand motif and the three long helices on the convex surface that were described in previous structures of eIF4E.

The loop conformations, however, are different from those observed in previous eIF4E structures. The loop containing residues 53-66, referred to as loop 1, inserts itself into a mostly hydrophobic pocket on the concave surface of the other molecule. As the cap analogue is bound between loops 1 and 2 (residues 104-115) in other structures, this makes cap binding impossible in the observed structure of wheat eIF4E, and indeed no cap was visible. Instead, loop 1 from molecule A and loop 1 from molecule B form a four-stranded beta sheet at the dimer interface. A tryptophan residue on loop 1, Trp 62, is the residue which is inserted most deeply into the "pocket" of the other molecule. This tryptophan is one of the two tryptophans implicated in cap binding by alignment with other eIF4E sequences. Representative electron density at this region is shown in Figure 3.9.

A disulfide bridge, which had not been previously observed in any eIF4E structures, was found between cysteines 113 and 151 of each monomer. The positioning of the disulfide relative to the loops and purported cap-binding tryptophans is shown in Figure 3.10. A Ramachandran plot for the solved structure is shown in Figure 3.11.

The other structural features reported in previous eIF4E structures were also observed in wheat eIF4E, including a disordered loop 3 (four residues of which could not be placed in good electron density and had to be omitted from the

structure) and the positively charged region implicated in binding the triphosphate linkage of the cap.

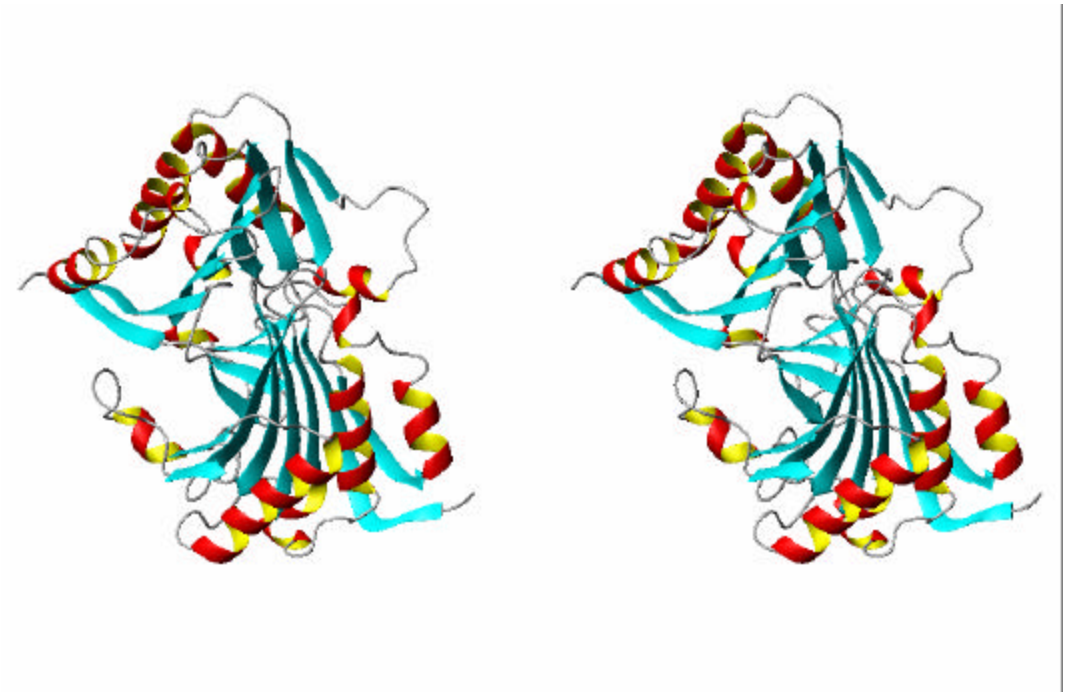


Figure 3.7: The solved structure of wheat eIF4E, shown as a ribbon diagram in cross-eyed stereo. On each monomeric unit, the eight-stranded beta sheet and three long helices are visible, as well as some coil regions with helical character. (Diagrams produced using MOLMOL, Koradi *et al.*, 1996)

Figure 3.8: The dimer interface of wheat eIF4E shown as part of a ribbon diagram. The four-stranded beta sheet stabilizes the interface. (Diagram produced with MOLMOL.)

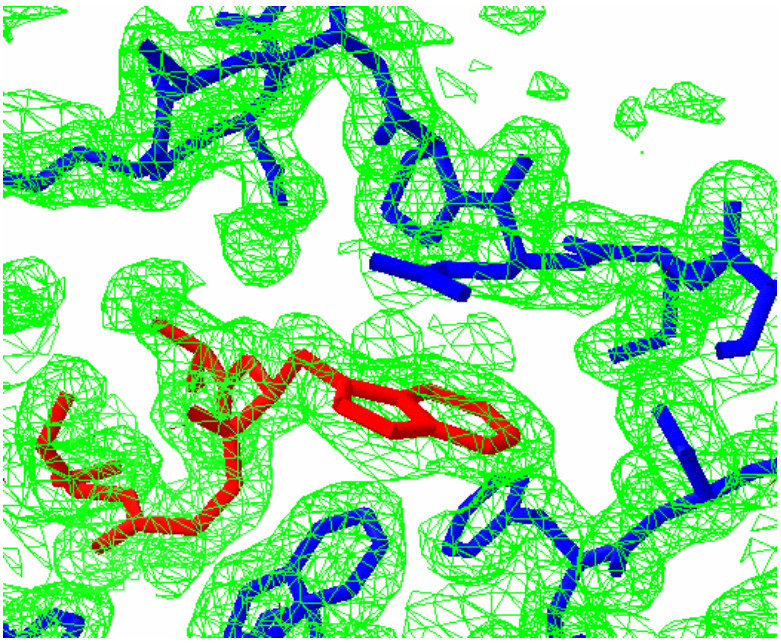


Figure 3.9: Representative electron density from a $2F_o$-F_c map, contoured at 1σ, of the wheat eIF4E structure. Chain A is shown in red, with Trp 62 inserted into a pocket of chain B (blue). Side contacts from a tryptophan and a phenylalanine of chain B help to stabilize the interaction. (Figure produced with Swiss PDB Viewer, Guex and Peitsch, 1997.)

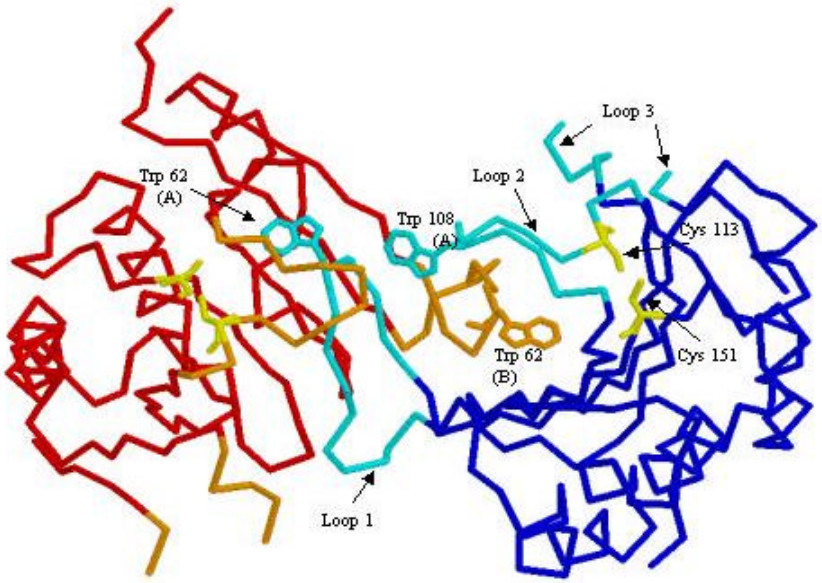


Figure 3.10: The backbone of the dimeric structure of wheat eIF4E, showing the side chains of some key residues. Molecule A is shown in blue, with the three large loops in cyan (labeled); molecule B is shown in red, with loops in orange. The tryptophans analogous to those previously implicated in cap binding, Trp 62 and Trp 108, are shown, as well as the disulfide cysteines 113 and 151. Loop 1 of each monomer extends into the neighboring molecule with Trp 62 in the center of the binding pocket. The end of loop 2 is anchored to the β sheet by a disulfide bridge. The flexible loop 3 is shown with a gap in the middle due to poor electron density – a similar situation to that reported in other eIF4E structures. (Figure produced with RasMol.)

Low Resolution (Angstroms)	20.0
High Resolution (Angstroms)	1.85
Measured Reflections (#)	586391
Unique Reflections (#)	38362
Completeness	100%
Crystallographic R Factor	0.243 (highest shell 0.331)
Free R Factor	0.265 (highest shell 0.288)
Ramachandran:	(Figure 3.10)
Percentage in Most Favored Regions	90.7%
Percentage in Disallowed Regions	0.0%
Number of Ordered Waters	95

Table 3.1: Crystal Data Table for Synchrotron Data (Wheat eIF4E).

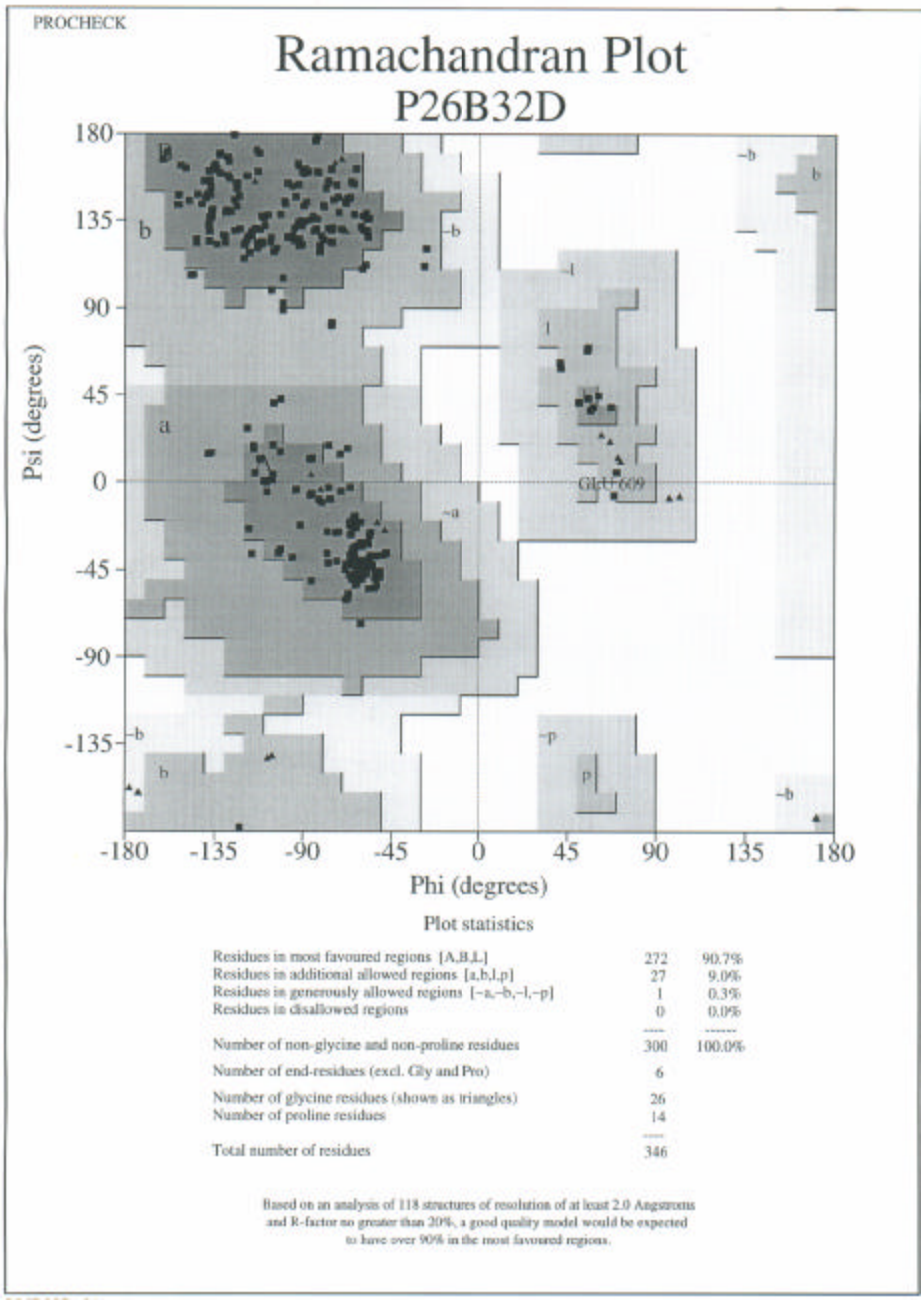


Figure 3.11: Ramachandran plot for the wheat eIF4E structure, made with PROCHECK (Laskowski *et al*., 1993).

GEL FILTRATION RESULTS

A gel filtration experiment was performed using purified wheat eIF4E, in order to see whether the protein behaved as a monomer or a dimer in solution. The gel filtration column was calibrated with standards as summarized in Table 3.2. The calibration is obtained by calculating the log(MW) of the standards and then the K_{av} parameter for each run.

$K_{av}=(V_e-V_o)/(V_t-V_o)$

where V_e=the elution volume, V_o=the void volume (time taken for elution of blue dextran, which runs through the column at the maximum speed), and V_t=the bed volume of the column. A plot was created with K_{av} on the y axis and log(MW) on the x axis, to give a linear curve over six calibration standards of varying molecular weights.

Gel filtration peaks tend to be broad, so it is difficult to obtain precision in the resultant data points. However, when deciding whether a protein runs as a monomer or dimer, there should be enough difference in molecular weights that the answer is obvious. The calibration curve is shown in Figure 3.13; the equation ultimately used was determined by a line between two of the smaller molecular weight standards (42 and 13.5 kDa, respectively), since the curve was not quite linear between those molecular weights and higher ones. The protein in question, truncated eIF4E, would be either 21 kDa (monomer) or 42 kDa (dimer).

Fractions were collected in 0.5-mL increments, and the UV absorbance was recorded by a chart recorder. Each protein showed a distinct peak; for most

runs with eIF4E, this peak came around 78 mL of elution. (Figure 3.12) Following the formula, this translates to a molecular weight around 18 kDa, closest to that of the monomer. This was between the calibration standards chymotrypsinogen A (25 kDa, 70.5 mL) and ribonuclease A (13.7 kDa, 81 mL).

The W62F mutant form of eIF4E was also run on the gel filtration column and found to be a monomer, running at the same rate as the original. A version of eIF4E without the N-terminal truncation was also run on the column; based on the gel filtration run, its molecular weight was calculated to be 24.5 kDa, close to its actual monomeric weight of 26 kDa.

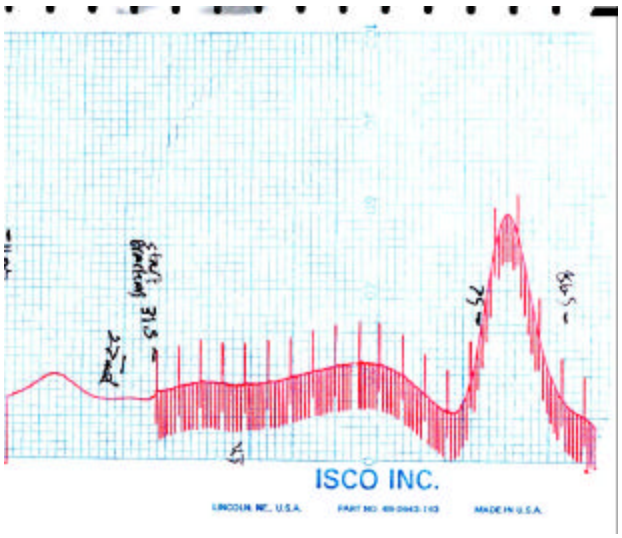


Figure 3.12: Typical chromatogram of a gel filtration run.

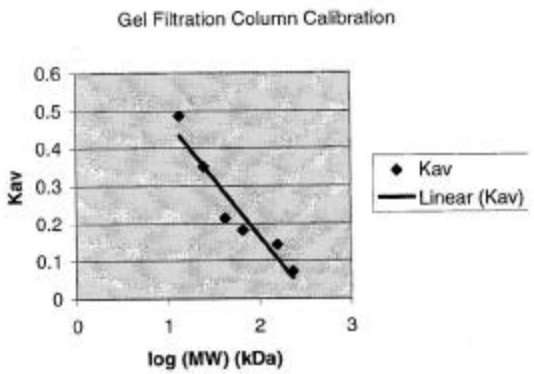


Figure 3.13: Gel filtration calibration curve.

standard	Kav	logMW	MW (kD)
aldolase	0.142	2.2	158
ovalbumin	0.214	1.63	43
RNase a	0.486	1.14	13.7
catalase	0.0713	2.37	232
BSA	0.182	1.83	67
chym. A	0.35	1.34	25

peak	Kav	logMW	MW
eIF4E	0.448	1.25	18
native	0.357	1.4	25

Table 3.2: Table of gel filtration data.

ELLMAN ASSAY RESULTS

The Ellman assay was performed in order to determine the number of free thiols in a solution of eIF4E, and therefore to ascertain whether the disulfide bond existed outside of crystal packing conditions. When 4.5 nmol eIF4E was incubated with an excess of DTNB in phosphate buffer, the absorbance at 412 nm was 0.295 units greater than that of DTNB in buffer alone. In order to remove the DTT found in the preparation buffer, the protein sample was exhaustively dialyzed against phosphate buffer so that the DTT content would be negligible. The buffer used for the control reading was taken from the container in which the last round of dialysis was performed, so that any background reaction would be detected.

According to Beer's Law, $A=\varepsilon cL$, the absorbance A for this reaction is 0.295 units, $\varepsilon=14{,}150\ M^{-1}cm^{-1}$, and L=1 cm. Therefore, the concentration of TNB in the cuvette is 20.8 μM, meaning that 10.4 nmol were present in the 0.5-mL reaction volume. This is roughly twice the amount of protein in the cuvette. Since each free sulfhydryl on a protein molecule can react to form one molecule of TNB, this means there were two free sulfhydryls on each protein molecule.

It is significant that two thiols were free to react in eIF4E, which has four cysteines. Assuming all thiols on the protein were in a position to react with Ellman's reagent, a disulfide bond between two of the cysteine residues does exist in solution.

CHAPTER 4: DISCUSSION

STRUCTURAL FEATURES AND COMPARISON

Wheat and Mouse Structure Comparison

The monomeric structures of wheat and mouse eIF4E are superimposed in Figure 4.1. The cupped-hand beta sheet motif is retained, as are the three alpha helices. The eIF4G-binding region on the first helix of murine eIF4E corresponds well to the first helix of wheat eIF4E; in fact, all helices are in essentially the same positions in both structures, though the second helix has one extra turn in the wheat structure.

The loop regions also correspond, though they may be ordered in different ways. The loop closest to the C-terminus, loop 3, was disordered in both structures. In both cases, about four residues could not be clearly positioned in the electron density maps even after refinement. These are left out of the murine structure; based on weak electron density, this loop is included in this picture of the wheat structure, but has been left out of the pdb file used in the final refinement and Ramachandran plot. The C-terminus itself corresponded to that found in the murine structure and was easily located through molecular replacement.

One of the most obvious differences between the solved structures is in the method of dimerization. The murine structure dimer is purely crystallographic, with the two helix-containing convex surfaces facing each other. However, the

wheat eIF4E has a dimeric asymmetric unit. Two non-crystallographically related molecules form a dimeric unit with a shared four-stranded beta sheet formed from some of the loop regions.

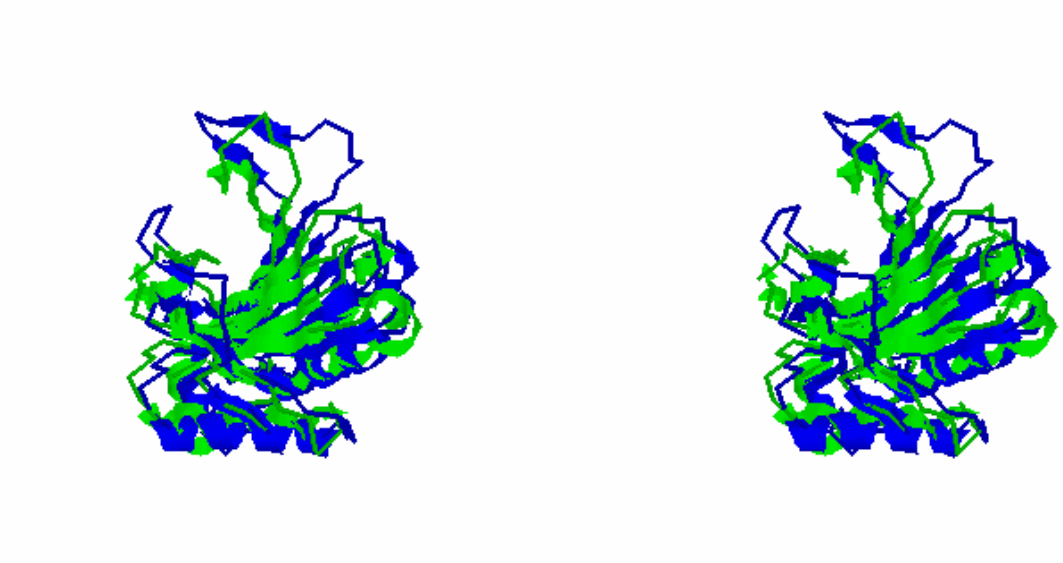


Figure 4.1: Cross-eyed stereo view of the monomeric structure of wheat eIF4E (dark) superimposed upon that of mouse eIF4E (light). Helical and sheet regions are represented as ribbons, and generally superimpose. Loop 1 is at the top of the figure; the two loops on the left side are loop 3 (foreground) and loop 2 (background). (Figure produced using RasMol.)

Cap Binding Tryptophans

The dimerization effect leads to another important structural difference between the wheat eIF4E and the others which have been previously solved: the

absence of bound cap analogue from the wheat structure. The murine structure contained the 7-methylguanosine diphosphate molecule that it had been co-purified with; in fact, the researchers were not able to obtain a crystal without a cap analogue in it (N. Sonenberg, personal communication). However, the wheat structure did not include a cap, and no crystals of wheat eIF4E with a cap analogue have been produced. The cap-binding motif of two stacked tryptophan sidechains, found in the mouse and yeast structures, does not exist in the wheat structure.

The general "hydrophobic pocket" is still present in the wheat eIF4E structure (Figure 4.2), though it is filled not by a cap analogue, but by one of the tryptophans (Trp 62) that corresponds to a cap-binding tryptophan in the previously solved eIF4E structures. The tryptophan belongs to the other molecule in the dimer; if there is a hydrophobic (formerly cap-binding) pocket on molecule A, it will be filled by the tryptophan from Molecule B, and vice versa. The tryptophan is part of an extended loop that reaches into the other molecule and forms part of the dimeric beta sheet. With the tryptophan itself taking up the cap-binding region on its neighboring molecule, the "tryptophan sandwich" disappears and there is no longer a way to bind the cap. It is assumed that some sort of cap binding must have taken place in order for the protein to be purified on the m^7GTP-sepharose column. This could have been accomplished with the previously reported cap-binding mechanism of aromatic stacking, and the cap release and dimerization could have taken place later. However, it is theoretically

possible that the cap could have been bound loosely through retention in the hydrophobic pocket alone, without actual stacking of the tryptophans.

The other cap-binding tryptophan, Trp 108, is not a part of this non-crystallographic symmetry interaction. Trp 108 was difficult to place, especially for the second molecule (B), and multiple conformations may exist for this residue. The Trp 108 residues of molecules A and B are located close together and may not both be able to exist in a normal configuration because of packing constraints. It appears that one tryptophan was forced to swing "out" away from the dimer interface, and Trp 108 of molecule B appears disordered in the structure and could not be located despite repeated attempts. The positions of the cap-binding tryptophans in this wheat eIF4E structure are shown in Figure 4.3.

While molecules A and B have a very similar backbone structure, they are each given separate coordinates because of minor differences in loop conformations and a few of the sidechains. (Figure 4.4) However, the only major conformational difference is in the positioning of Trp 108. If both units of the dimer were identical, their Trp 108 residues would overlap; this may be one reason that they were forced to assume different configurations. Minor differences in conformation may in fact be present from one molecule to another in the crystal; the density for this residue was not as strong as that for other tryptophans, and the side chain density for Glu 109 does not have the definite shape typically seen in atomic resolution maps (Figure 4.5).

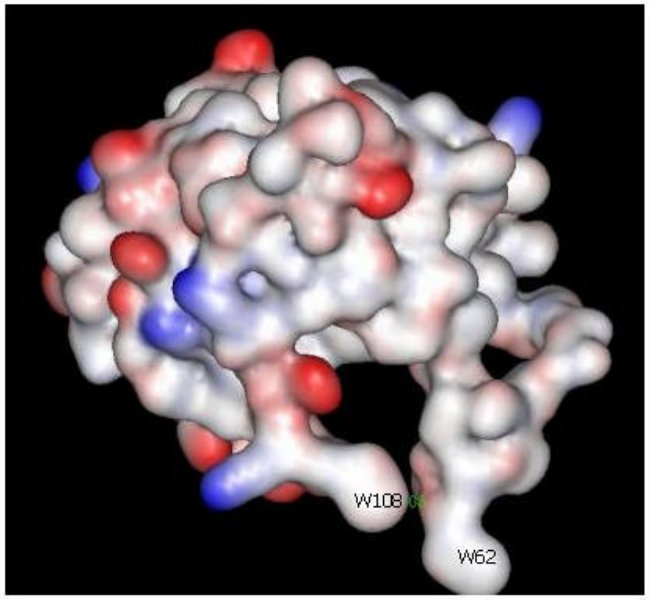


Figure 4.2: The electrostatic surface of one of the two monomers of wheat eIF4E. The hydrophobic areas are lighter in color. The "hole"seen in this monomeric structure is filled, in the dimer, by W62 of the neighboring molecule. Its surface is mostly hydrophobic, except for one glutamate side chain. (Figure produced using WebLab ViewerLite, Molecular Simulations, Inc.)

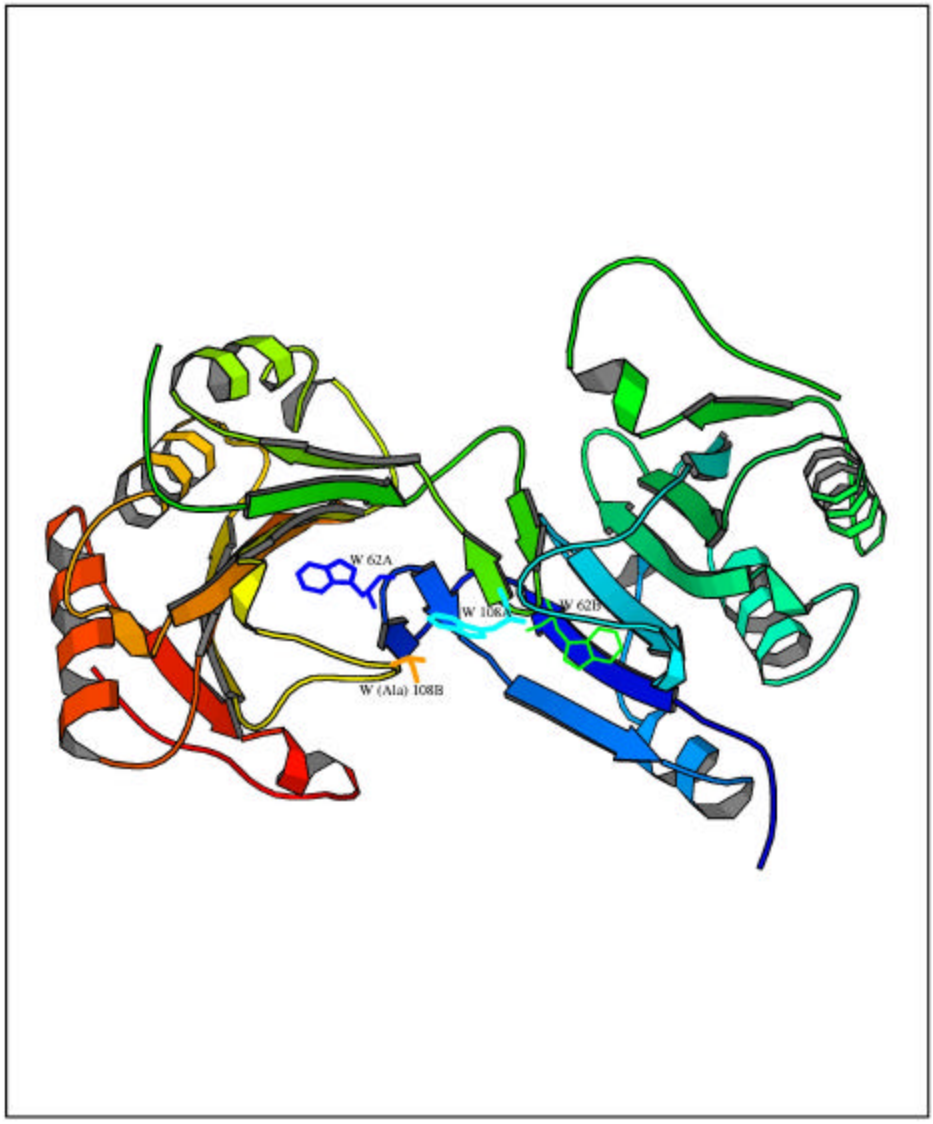


Figure 4.3: Ribbon diagram of wheat eIF4E, showing all of the tryptophans that are homologous to cap-binding tryptophans in murine eIF4E. Trp 108 of the B chain is disordered and is represented by an alanine. (Figure produced using MolScript.)

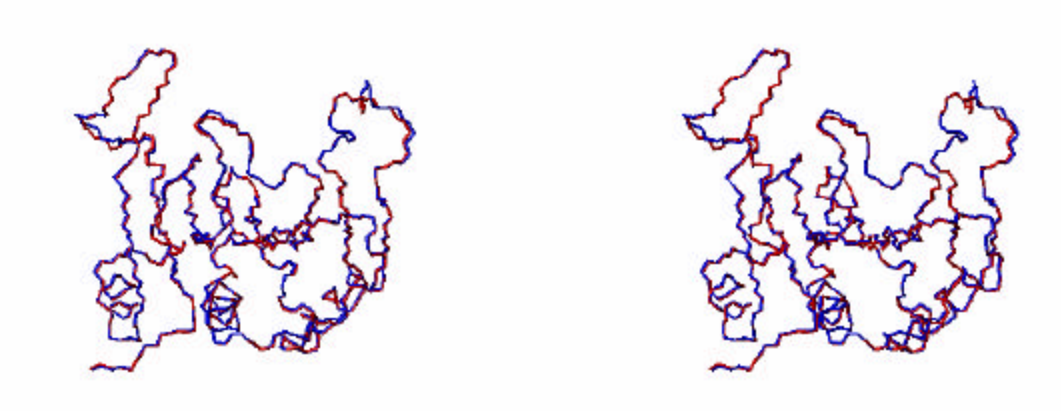


Figure 4.4: Stereo diagram superimposing the two chains of wheat eIF4E. (Figure produced using MOLMOL.)

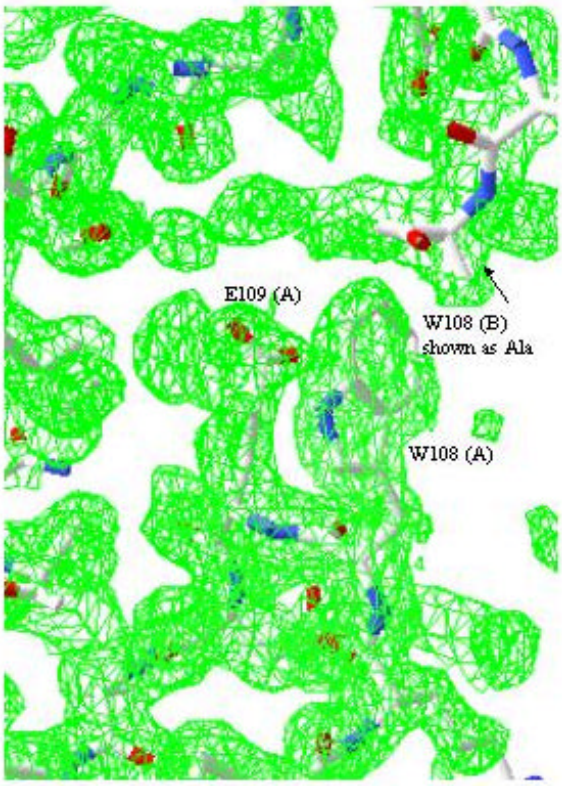


Figure 4.5: Electron density at the Trp 108 region of wheat eIF4E. Trp 108 of molecule B is shown as an alanine, due to lack of electron density for the rest of the sidechain. (Figure produced with Swiss PDB Viewer, Guex and Peitsch, 1997.)

DISULFIDE BONDS

Another feature in the wheat eIF4E structure that had not been observed in other eIF4E structures was a disulfide bridge between Cys 113 and Cys 151. The mouse and yeast eIF4E sequences do not have cysteines in both of those positions, making this cystine impossible. There are two other cysteine residues

in the wheat protein, but they are known to exist as free thiols from observing the electron density map.

The presence of the disulfide bond is not due to phase bias. A round of Arp_Warp was performed using the synchrotron data and a molecular replacement solution which left out all of Loops 1, 2, and 3, which included the cap-binding tryptophans and Cys 113. A Figure of Merit map of atomic resolution was produced, from which the backbone and side chains were readily visible. Starting with the already solved regions, which differed little from those of mouse eIF4E, and tracing the chain, all of the sidechains in the region of residues 100-115 were visible in their proper places (except for Trp 108) and the bridge was clearly evident between the cysteines. This was the case for both molecules A and B of the dimer. When the cysteine residues were added to the structure and refined, the initial S-S distances were 2.05 and 2.16 Angstroms, respectively, for each monomer. This is close to the standard disulfide bond length of 2.0 Angstroms. The disulfide bridge and surrounding electron density are shown in Figure 4.6.

Cystines are not normally found in eukaryotic cytosolic proteins, raising the question of whether the oxidation to a disulfide bond occurred naturally or only after time in the crystallization drop. The disulfide bond probably has a profound impact on structural conformation; Cys 113 is only five residues away from the cap-binding tryptophan Trp 108, and both exist as part of the same loop. The cysteine is located near the edge of the loop region. If it makes contact with Cys 151 and forms a bridge, the conformation of the loop could be changed so

that the cap-binding tryptophans are too far apart to hold the cap tightly. This could occur due to oxidation in the crystal, or possibly due to an oxidative "switch" mechanism that would allow for cap binding and release. Several sequences of eIF4E have been obtained from higher plants (see alignment), and all of them contain both cysteines, as does the sequence of eIF(iso)4E. However, no other organisms have both cysteines in their eIF4E sequence.

The existence of the disulfide bond was substantiated by the results of the Ellman assay. By quantitation of the amount of TNB formed from DTNB, the assay indicated the presence of two free thiols in each monomer of wheat eIF4E. Since there are four cysteines in the protein, this signifies that the other two are involved in a disulfide bond. All of the cysteines are solvent-exposed (Figure 4.7), so it is likely that they would react with DTNB.

The Ellman reaction took place in solution and did not involve previously crystallized protein, thus indicating that the crystal packing conditions are not the sole cause of the disulfide bond. However, usually the protein is kept under somewhat reducing conditions with a small amount of DTT; this was removed through dialysis before the Ellman assay because DTT has its own free thiol groups. Therefore, if the existence of the disulfide is dependent upon environmental conditions, this environment would have been sufficiently "oxidizing". While the environment inside the cell is usually not conducive to the presence of a disulfide bond, these experiments have shown that this structure can at least form a cystine given the right conditions.

It is worth noting that dialysis removed the excess cap analogue molecules from solution, and possibly the "bound" caps from the protein as well. Structurally, the existence of the disulfide appears to lead to lower cap affinity as it pulls the cap-binding loops further apart. This could also theoretically lead to higher cap affinity, since the bridge keeps Loop 2 in position; however, the flexibility of this loop may be important in maintaining strong cap interactions. If this is a permanent feature of eIF4Es in higher plants, it may lead to a lower cap affinity overall. A mechanism that switches cap binding on or off depending upon oxidation may be useful as a means of providing translational control.

The existence of two isoforms of eIF4E should also be considered, as the two may perform different functions. While previous studies have shown that the isoforms of eIF4F prefer different mRNAs, no structural interactions have been found yet between the mRNA (apart from the cap) and eIF4E, with the exception of loop 3 interacting with the second nucleotide (Tomoo *et al*., 2002). The other subunit, eIF4G (with isoform eIF(iso)4G), may be responsible for many of the differences between isoforms. Attempts to co-crystallize mRNA with eIF4E have been unsuccessful thus far; obviously the current conditions would not work, since the cap must be retained in the crystal.

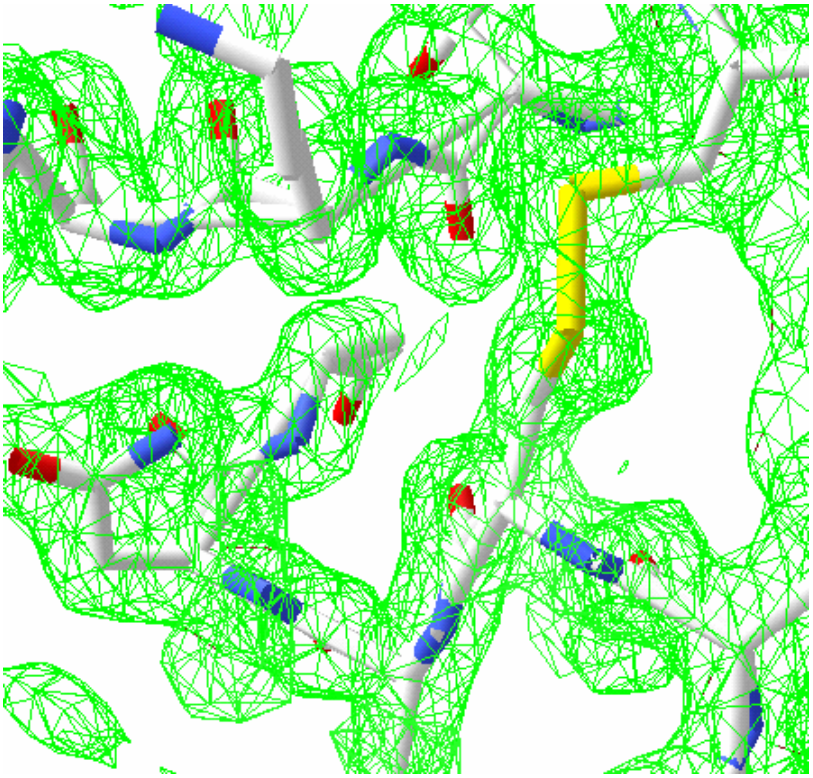


Figure 4.6: The disulfide bridge between Cys 113 and Cys 151, along with electron density from a $2F_o$-F_c map contoured at 1σ. (Figure produced with Swiss PDB Viewer.)

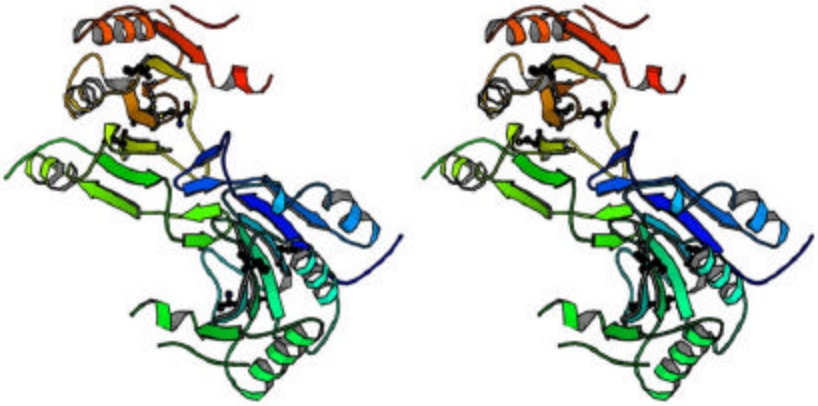


Figure 4.7: Cross-eyed stereo view of the wheat eIF4E dimer, shown as a ribbon with ball-and-stick cysteines. Particularly in the case of the top molecule, this indicates the solvent accessibility of the cysteine groups. Two of these cysteines are within disulfide bonding distance. The C-termini of the molecules appear to be cut off from the rest of the chain because some of the coordinates of loop 3 had to be omitted from the structure. (Figure produced with MolScript.)

FUTURE DIRECTIONS

The crystal structure of eIF4E has been solved at 1.85 Å resolution and shows a dimer held together by a beta sheet. While this protein was forced to perform cap binding in order to be purified, the cap was not retained in the crystal, even when no dialysis or other cap-removing processes had been performed on the protein. The dimerization and loss of the cap may have been made possible by a disulfide bridge that was clearly visible in the structure, but is

not found in other eIF4Es and would be an unusual finding for a eukaryotic cytosolic protein.

One future direction would be to determine whether the cystine occurs in solution; based on gel filtration results, the dimerization does not show up in solution and appears to be either a function of high concentration or crystal packing. The disulfide bridge may occur due to oxidation in the crystal; when the concentration of DTT in the usual crystallization conditions was increased fivefold to discourage oxidation, no crystals appeared. However, further attempts could be made, as crystallization did not occur in all plates of eIF4E even under optimum conditions. If the "control" half of a plate, with a lower concentration of DTT, clearly grew crystals more efficiently than the half with added DTT, this would support the idea that the disulfide bridge was a function of oxidation in the crystal. This does not determine whether the oxidation can occur in solution or in nature, though, which is another problem to be investigated. The Ellman assay can also be repeated with eIF4E in a less oxidizing environment, in order to see if that changes the number of free thiols available for reaction.

The isoform, eIF(iso)4E, has yet to be crystallized. The successful cloning of a N-terminal truncation mutant, removing more of the N-terminus than was previously done, would be a first step, followed by purification and crystal screening. It would be interesting to determine whether this protein preferred to crystallize with or without the cap, and to see if the disulfide bridge appeared in its structure.

The only crystal structures which currently exist for eIF4G, or fragments thereof, are those of oligopeptides bound to eIF4E and that of the eIF4A-binding domain (Marcotrigiano *et al*., 2001). If the N-terminal 137 residues of eIF(iso)4G, or their equivalent in eIF4G, could be expressed separately, these could be crystallized alone or as cocrystals with eIF4E to give more insights into the function of eIF4G and the structure of the bound form.

Appendices

APPENDIX A: MEDIA RECIPES

2xYT

16 g tryptone peptone

10 g yeast extract

5 g sodium chloride

1 L water

(for solid media, 1.5% agar w/v)

NZY+

10 g NZ amine (casein hydrolysate)

5 g yeast extract

5 g sodium chloride

1 L water

12.5 mL of 1 M magnesium chloride

12.5 mL of 1 M magnesium sulfate

10 mL of 2 M glucose

Materials: tryptone peptone, yeast extract, and agar from Difco; all other chemicals from Mallinckrodt, Sigma, or Fisher

APPENDIX B: TRYPSIN DIGESTION RESULTS

N-terminal Protein Sequencing Results for eIF(iso)4G Trypsin Digestion Fragments:

Band 1:													
results	d (e,s,a)	q	P	P	e (g,s,a)								
corresponding sequence	*E*	*K*	*P*	*P*	*A*								
Band 2:													
results	E,d,l	L	P	Q	g,s	t,k	v,g,l	a	a	i?	v	k	a
corresponding sequence	*L*	*L*	*P*	*Q*	*G*	*T*	*G*	*A*	*L*	*I*	*G*	*K*	*S*
Band 3:													
results	V	?	P	Q	v,g	v,t	s,g,v	a	t				
corresponding sequence	*L*	*L*	*P*	*Q*	*G*	*T*	*G*	*A*	*L*				

Results were obtained from the University of Texas Protein Microanalysis Facility. "Results" above indicate the amino acids shown by the sequence analysis (lower-case letters, particularly those in parentheses, indicate uncertain assignments). Corresponding sequences in the actual eIF(iso)4G sequence, which has been determined by the Browning laboratory, are in italics.

Band 1 corresponded to a fragment beginning at the 137th amino acid of eIF(iso)4G; and Band 2 indicated a fragment beginning at the 555th amino acid. Band 3 appears to have the same N-terminal sequence as Band 2, and is probably a C-terminal degradation product of the other band. The molecular weights of each fragment are consistent with their position in the sequence. The N-terminal

fragment was not found, and may have undergone considerable proteolysis at this point in the trypsin digest.

APPENDIX C: PROTEIN SEQUENCE FILE

Protein sequence of wheat eIF4E, as entered in the Entrez protein database (http://www.ncbi.nlm.nih.gov/). Special notes below in regard to residues left out of the solved structure.

LOCUS IFE1_WHEAT 215 aa linear PLN 15-JUL-1999
DEFINITION EUKARYOTIC TRANSLATION INITIATION FACTOR 4E (EIF-4E) (EIF4E) (MRNA
CAP-BINDING PROTEIN) (EIF-4F 25 KD SUBUNIT) (EIF-4F P26 SUBUNIT).
ACCESSION P29557
PID g1352441
VERSION P29557 GI:1352441
DBSOURCE swissprot: locus IFE1_WHEAT, accession P29557;
class: standard.
created: Apr 1, 1993.
sequence updated: Feb 1, 1996.
annotation updated: Jul 15, 1999.
xrefs: gi: gi: 450367, gi: gi: 1208929, gi: gi: 283021
xrefs (non-sequence databases): HSSP P07260, PROSITE PS00813
KEYWORDS Initiation factor; Protein biosynthesis; RNA-binding; Multigene family.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (residues 1 to 215)
AUTHORS METZ,A.M.
TITLE Direct Submission
JOURNAL Submitted (??-JAN-1994)
REMARK SEQUENCE FROM N.A.
REFERENCE 2 (residues 1 to 215)
AUTHORS Metz,A.M., Timmer,R.T. and Browning,K.S.
TITLE Isolation and sequence of a cDNA encoding the cap binding protein of wheat eukaryotic protein synthesis initiation factor 4F
JOURNAL Nucleic Acids Res. 20 (15), 4096 (1992)
MEDLINE 92375704
REMARK SEQUENCE OF 19-215 FROM N.A., AND PARTIAL SEQUENCE.
COMMENT On Jun 4, 1996 this sequence version replaced gi:266337.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute.
The original entry is available from http://www.expasy.ch/sprot and http://www.ebi.ac.uk/sprot

---------- .
[FUNCTION] RECOGNIZES AND BINDS THE 7-METHYLGUANOSINE-CONTAINING
MRNA 'CAP' DURING AN EARLY STEP IN THE INITIATION OF PROTEIN SYNTHESIS AND FACILITATES RIBOSOME BINDING BY INDUCING THE UNWINDING OF THE MRNAS SECONDARY STRUCTURES.
[SUBUNIT] EIF4F IS A TRIMER COMPOSED OF EIF4E, EIF4G AND EIF4A (WHICH CAN CYCLE IN AND OUT OF THE COMPLEX). IN HIGHER PLANTS TWO
ISOFORMS OF EIF4F HAVE BEEN IDENTIFIED, NAMED EIF4F AND EIF(ISO)4F. EIF4F HAS SUBUNITS P220 AND P28, WHEREAS EIF-(ISO)4F HAS SUBUNITS P82 AND P26.
[SIMILARITY] BELONGS TO THE EUKARYOTIC INITIATION FACTOR 4E FAMILY.
FEATURES Location/Qualifiers
source 1..215
/organism="Triticum aestivum"
/db_xref="taxon:4565"
Protein 1..215
/product="EUKARYOTIC TRANSLATION INITIATION FACTOR 4E"
ORIGIN

1 mtedtemrpa sagaeereeg eiaddgdgss aaaagrit*ah p**l**enawtfwf d**n**pqg**k**s**r**qv
61 awgstihpih tfstvedfwg lynnihnpsk lnvgadfhcf knkiepk*w*ed picanggkwt
121 iscgrgksdt fwlhtllami geqfdfgdei cgavvsvrq*k* qervaiwtkn aaneaaqisi
181 gkqwk**e**fldy **k**dsigfivhe da*k***r**sdkgpk n*r*ytv
//

Notes: the sequence of the truncated version used in these experiments begins after the asterisk, with the alanine residue. The underlined residues did not have clear electron density and were left out of the model entirely for refinement purposes (particularly those used to obtain the R factors and Ramachandran plot). Residues in **bold** were found to have disordered sidechains and have been replaced by alanines in the model. *Italicized* residues were also found to have disordered sidechains and replaced by alanines, but only on one of the two monomers (chain B for Trp 108, chain A for the others).

APPENDIX D: SEQUENCE OF EIF4E

DNA sequence of N-terminally truncated wheat eIF4E, as confirmed by the University of Texas DNA Core Facility.

```
+3  M  A  H   P  L  E  N   A  W  T   F  W  F   D  N  P  Q   G  K  S
 1 ATGGCCCACC CGCTCGAGAA CGCCTGGACC TTCTGGTTCG ACAACCCGCA GGGCAAGTCC
```

```
 +3  R  Q  V   A  W  G  S   T  I  H   P  I  H   T  F  S  T   V  E  D
 61 AGGCAGGTGG CCTGGGGGAG CACCATCCAC CCCATCCACA CCTTCTCCAC CGTCGAGGAC

 +3  F  W  G   L  Y  N  N   I  H  N   P  S  K   L  N  V  G   A  D  F
121 TTCTGGGGCC TTTACAACAA TATCCATAAC CCTAGCAAGT TGAATGTGGG AGCCGACTTC

 +3  H  C  F   K  N  K  I   E  P  K   W  E  D   P  I  C  A   N  G  G
181 CATTGCTTCA AGAACAAGAT TGAGCCAAAA TGGGAAGACC CCATTTGTGC CAATGGCGGT

 +3  K  W  T   I  S  C  G   R  G  K   S  D  T   F  W  L  H   T  L  L
241 AAATGGACCA TCAGCTGTGG CAGAGGGAAA TCTGACACAT TCTGGTTGCA TACTTTGCTG

 +3  A  M  I   G  E  Q  F   D  F  G   D  E  I   C  G  A  V   V  S  V
301 GCAATGATTG GTGAACAATT TGACTTTGGT GATGAAATTT GTGGAGCAGT CGTTAGCGTG

 +3  R  Q  K   Q  E  R  V   A  I  W   T  K  N   A  A  N  E   A  A  Q
361 CGTCAGAAAC AGGAAAGAGT AGCTATCTGG ACCAAAAATG CTGCCAATGA AGCTGCTCAG

 +3  I  S  I   G  K  Q  W   K  E  F   L  D  Y   K  D  S  I   G  F  I
421 ATAAGCATTG GCAAGCAGTG GAAGGAGTTT CTGGACTACA AGGACTCCAT TGGGTTCATT

 +3  V  H  E   D  A  K  R   S  D  K   G  P  K   N  R  Y  T   V  &
481 GTTCATGAGG ATGCAAAGAG GTCTGACAAA GGCCCCAAGA ACCGCTACAC CGTTTGA
```

This sequence analysis was performed on the plasmid used to express protein, some of which led to the crystals used in solving this structure. (Figure produced with Gene Runner, Hastings Software, Inc..)

APPENDIX E: SEQUENCE OF W62F MUTANT

DNA sequence of the W62F mutant of N-terminally truncated wheat eIF4E, as confirmed by the University of Texas DNA Core Facility.

```
 +3  M  A  H   P  L  E  N   A  W  T   F  W  F   D  N  P  Q   G  K  S
  1 ATGGCCCACC CGCTCGAGAA CGCCTGGACC TTCTGGTTCG ACAACCCGCA GGGCAAGTCC

 +3  R  Q  V   A  F  G  S   T  I  H   P  I  H   T  F  S  T   V  E  D
 61 AGGCAGGTGG CCTTCGGGAG CACCATCCAC CCCATCCACA CCTTCTCCAC CGTCGAGGAC

 +3  F  W  G   L  Y  N  N   I  H  N   P  S  K   L  N  V  G   A  D  F
121 TTCTGGGGCC TTTACAACAA TATCCATAAC CCTAGCAAGT TGAATGTGGG AGCCGACTTC

 +3  H  C  F   K  N  K  I   E  P  K   W  E  D   P  I  C  A   N  G  G
181 CATTGCTTCA AGAACAAGAT TGAGCCAAAA TGGGAAGACC CCATTTGTGC CAATGGCGGT

 +3  K  W  T   I  S  C  G   R  G  K   S  D  T   F  W  L  H   T  L  L
241 AAATGGACCA TCAGCTGTGG CAGAGGGAAA TCTGACACAT TCTGGTTGCA TACTTTGCTG

 +3  A  M  I   G  E  Q  F   D  F  G   D  E  I   C  G  A  V   V  S  V
301 GCAATGATTG GTGAACAATT TGACTTTGGT GATGAAATTT GTGGAGCAGT CGTTAGCGTG
```

```
 +3  R  Q  K   Q  E  R  V   A  I  W   T  K  N   A  A  N  E   A  A  Q
361 CGTCAGAAAC AGGAAAGAGT AGCTATCTGG ACCAAAAATG CTGCCAATGA AGCTGCTCAG

 +3  I  S  I   G  K  Q  W   K  E  F   L  D  Y   K  D  S  I   G  F  I
421 ATAAGCATTG GCAAGCAGTG GAAGGAGTTT CTGGACTACA AGGACTCCAT TGGGTTCATT

 +3  V  H  E   D  A  K  R   S  D  K   G  P  K   N  R  Y  T   V  &
481 GTTCATGAGG ATGCAAAGAG GTCTGACAAA GGCCCCAAGA ACCGCTACAC CGTTTGA
```

This sequence analysis confirms the site-directed mutagenesis of the codon for Tryptophan 62 to a codon for Phenylalanine 62 (shown in bold). (Figure produced using Gene Runner, Hastings Software, Inc..)

References:

Balasta, M.L., Carberry, S.E., Friedland, D.E., Perez, R.A., Goss, D.J. (1993) *J Biol Chem* **268**, 18599-18603.

Barrett, C.R., Nix, W.D., Tetelman, A.S. (1973) The Principles of Engineering Materials. (Prentice-Hall, Englewood Cliffs, NJ)

Browning, K.S. (1996) *Plant Molecular Biology* **32**, 107-144.

Brunger, A.T. (1992) XPLOR Version 3.1. (Yale University Press, New Haven, Ct.)

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilgest, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. (1998) *Acta Cryst* **D54**, 905-921.

Carberry, S.E., Darzynkiewicz, E., Goss, D.J. (1991) *Biochemistry* **30**, 1624-1627.

Carberry, S.E. and Goss, D.J. (1991a) *Biochemistry* **30**, 6977-6982

Carberry, S.E. and Goss, D.J. (1991b) *Biochemistry* **30**, 4542-4545.

Drenth, J. (1994) Principles of Protein X-Ray Crystallography. (Springer-Verlag, New York, NY)

Ellman, G. (1959) *Arch Biochem Biophys* **82**, 70-77.

Gallie, D.R. and Browning, K.S. (2001) *J Biol Chem* **276**, 36951-36960.

Gingras, A-C., Raught, B., Sonenberg, N. (1999) *Annu. Rev. Biochem* **68**, 913-963.

Guex, N. and Peitsch, M.C. (1997) *Electrophoresis* **18**, 2714-2723.

Holmes, K.C. and Blow, D.M. (1965) The Use of X-ray Diffraction in the Study of Protein and Nucleic Acid Structure. (Robert E. Kriegel Publishing Company, Huntington, NY)

Hsu, P., Hodel, M.R., Thomas, J.W., Taylor, L.J., Hagedorn, C.H., Hodel, A.E. (2000) *Biochemistry* **39**, 13730-13736.

Jones, T.A. and Thirup, S. (1986) *EMBO J* **5**, 819-822.

Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M. (1991) *Acta Crystallogr A* **47**, 110-119.

Keiper, B.D., Gan, W., Rhoads, R.E. (1999) *International Journal of Biochemistry and Cell Biology* **31**, 37-41.

Kissinger, C.R., Gehlhaar, D.K., Fogel, D.B. (1999) *Acta Crystallographica* **D55**, 484-496.

Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J Mol Graphics* **14**, 51-55.

Kraulis, P.R. (1991) *J Appl Crystallogr* **24**, 946-950.

Lamzin, V.S. and Wilson, K.S. (1993) *Acta Cryst* **D49**, 129-149.

Lane, R. and Snell, E. (1976) *Biochemistry* **15**, 4175-4179.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993) *J App Cryst* **26**, 283-291.

Lax, S.R., Lauer, S.J., Browning, K.S., Ravel, J.M. (1986) *Meth Enzymol* **118**, 109-128.

Le, H., Browning, K.S., Gallie, D.R. (2000) *J Biol Chem* **275**, 17452-17462.

Marcotrigiano, J., Gingras, A., Sonenberg, N., Burley, S.K. (1997) *Cell* **89**, 951-961.

Marcotrigiano, J., Gingras, A.C., Sonenberg, N., Burley, S.K. (1999) *Mol. Cell* **3**, 707-716.

Marcotrigiano, J., Lomakin, I.B., Sonenberg, N., Pestova, T.V., Hellen, C.U.T., Burley, S.K. (2001) *Mol. Cell* **7**, 193-203.

Matsuo, H., Li, H., McGuire, A.M., Fletcher, M.A., Gingras, A.C., Sonenberg, N., Wagner, G. (1997) *Nat Struct Biol* **4**, 717-724.

McRee, D.E. (1999) Practical Protein Crystallography, Second Edition. (Academic Press, San Diego, Ca.)

Metz., A.M. and Browning, K.S. (1996). *J Biol Chem* **271**, 31033-31036.

Metz, A.M., Timmer, R.T., Browning, K.S. (1992) *Nucl Acids Res* **20**, 4096.

Morino, S., Hazama, H., Ozaki, M., Teraoka, Y., Shibata, S., Doi, M., Ueda, H., Ishida, T., Uesugi, S. (1996) *Eur J Biochem* **239**, 597-601.

Otwinowski, Z. and Minor, W. (1997) *Meth Enzymol* **276**, 307-326.

Perrakis, A., Sixma, T.K., Wilson, K.S., Lamzin, V.S. (1997) *Acta Cryst* **D53**, 448-455.

Riddles, P., Blakeley, R., Zerner, B. (1983) *Meth Enzymol* **91**, 49-60.

Sayle, R. and Milner-White, E. J. (1995) *Trends in Biochemical Sciences* **20**, 374.

Sha, M., Wang, Y., Xiang, T., van Heerden, A., Browning, K.S., Goss, D.J. (1995) *J Biol Chem* **270**, 29904-29909.

Sonenberg, N. (1996) In Translational Control, J.W.B. Hershey, M.B. Mathews, and N. Sonenberg, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.)

Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S., Ishida, T., Taniguchi, T., Hasegawa, H., Terashima, A., Sasaki, M., Katsuya, Y., Kitamura, K., Miyoshi, H., Ishikawa, M., Miura, K. (2002) *Biochem J.* **362**, 539-544.

Van Heerden, A., Browning, K.S. (1994) *J Biol Chem* **269**, 17454-17457.

Vasilescu, S., Linz, M.P.B., Muller, P.P., McCarthy, J.E.G. (1996) *J Biol Chem* **271**, 7030-7037.

Voet, D. and Voet, J.G. (1995) Biochemistry, Second Edition. (John Wiley & Sons, New York.)

Vita

Jennifer Beth Hurley (Sadow) was born on March 20, 1975 in Gainesville, Florida to Jamie W. Hurley and Judith M. Hurley. She graduated from White Station High School in Memphis, Tennessee in 1992 and then entered Rice University in Houston, Texas, where she received a bachelor's degree in 1996, majoring in chemistry. After a brief experience in the chemical industry, she worked as a research technician at Baylor College of Medicine in Houston, studying the siloxane byproducts of silicone breast implants. In 1997 she entered the graduate school at The University of Texas at Austin. She currently has one publication (Lykissa, E.D., Kala, S.V., Hurley, J.B., Lebovitz, R.M. (1997) *Analytical Chemistry* **69**, 4912-4916) and another in preparation. In summer 2002, she will begin a new career as an assistant professor of chemistry at Schreiner University.

Permanent address: 2100 Memorial Boulevard, Kerrville, Texas 78028.

This dissertation was typed by the author.